U.PORTO
FLUP FACULDADE DE LETRAS
UNIVERSIDADE DO PORTO

3º CICLO DE ESTUDOS
CIÊNCIAS DA LINGUAGEM:
TECNOLOGIAS DA LINGUAGEM HUMANA

POST-EDITING: A THEORETICAL AND PRACTICAL CHALLENGE FOR TRANSLATION STUDIES AND MACHINE LEARNING

Félix Emanuel Martins do Carmo

D

2017

Félix Emanuel Martins do Carmo

POST-EDITING: A THEORETICAL AND PRACTICAL CHALLENGE FOR TRANSLATION STUDIES AND MACHINE LEARNING

Tese realizada no âmbito do Doutoramento em Ciências da Linguagem: Tecnologias da Linguagem Humana, orientada pelo Professor Doutor Thomas Juan Carlos Hüsgen.
e coorientada pela Professora Doutora Belinda Mary Harper Sousa Maia

Faculdade de Letras da Universidade do Porto

Setembro de 2017

DOUTORAMENTO EM CIÊNCIAS DA LINGUAGEM

PÓS-EDIÇÃO: UM DESAFIO TEÓRICO E PRÁTICO PARA OS ESTUDOS DE TRADUÇÃO E A APRENDIZAGEM AUTOMÁTICA

Tema:	Pós-Edição e Aprendizagem Automática
Área científica principal:	Estudos de Tradução
Curso de doutoramento:	Ciências da Linguagem
Domínio científico:	Tradução Automática

Autor: **Félix do Carmo**

Orientadores: **Prof. Dr. Thomas Hüsgen e Prof. Dra. Belinda Maia**

PhD in LANGUAGE SCIENCES

POST-EDITING: A THEORETICAL AND PRACTICAL CHALLENGE FOR TRANSLATION STUDIES AND MACHINE LEARNING

Theme:	Post-editing and Machine Learning
Main scientific area:	Translation Studies
PhD Program:	Language Sciences
Scientific domain:	Machine Translation

Author: **Félix do Carmo**

Supervisors: **Prof. Dr. Thomas Hüsgen and Prof. Dra. Belinda Maia**

Table of Contents

AGRADECIMENTOS

À Profa. Dra. Belinda Maia, por ter tido a coragem de criar e investir no curso de Doutoramento em Tecnologias da Linguagem Natural, abrindo uma porta inexistente no cruzamento entre a investigação em Línguas, Tradução, Engenharia e Matemática; por me ter aberto essa porta a um mundo fascinante de investigação e por se ter mantido ao meu lado, dando apoio e orientando um projeto do qual eu próprio desconfiei, durante um longo percurso, até atingir uma forma de que ambos talvez possamos orgulhar-nos. E pela forte amizade que criámos, com base na sua infindável capacidade de compreensão e apoio.

Ao Prof. Dr. Thomas Hüsgen pelo apoio e por ter acreditado no meu trabalho.

Ao Prof. Dr. Pavel Brazdil, por me ter lançado a grande velocidade num mundo de complexas tecnologias e ter confiado na minha capacidade de a usar de uma forma útil.

À Profa. Dorothy Kenny e ao Prof. Joss Moorkens pela participação no júri de provas de doutoramento e por todos os contributos para a discussão e melhoria da tese.

A todos os membros do júri de defesa da tese, pelo seu trabalho e contributo para a avaliação e aprovação da presente tese.

Ao meu irmão Helder, que apoiou inabalavelmente este projeto, prolongando esse apoio para lá da última entrega e dando asas a longas e frutuosas discussões que se abriram em múltiplas direções.

Ao Luís Trigo, colega e amigo que fez comigo este percurso em cima de uma ponte periclitante e que a atravessou, sozinho e comigo, num e noutro sentido, em busca de formas eficazes e eficientes de fazermos um trabalho útil.

Ao Chris Hokamp, sem a colaboração de quem este projeto simplesmente não teria passado das especulações.

E ainda…

Aos heróis não reconhecidos desta área de investigação, pessoas que fizeram muito mais do que o que lhes era pedido pela investigação em Língua Portuguesa e em Tradução, com destaque para a Profa. Belinda Maia, o Prof. Eckhard Bick, a Profa. Diana Santos e tantos outros que já deviam ter recebido medalhas do nosso país.

A todos os investigadores que, além de me inspirarem com trabalhos extraordinários, me deram o prazer de partilhar pessoalmente ideias e impressões, pessoas como Alon Lavie, Andy Way, Chris Hokamp, David Farwell, Dorothy Kenny, Ignacio García, Joss Moorkens, Jorg Dietemann, Lúcia Specia, Mikel Forcada, Olga Torres-Hostench, Pilar Sanchéz-Gijón, Sharon O'Brien. Até Alan Melby, um dos meus heróis de há muito tempo, tive o privilégio de conhecer...

Aos colegas da FLUP, que sempre me apoiaram e mostraram interesse pelo meu trabalho, com um especial reconhecimento à Elena Galvão.

À minha grande família, que, mesmo geograficamente afastada, esteve sempre presente e me fez sentir o seu apoio e orgulho no que eu estava a fazer.

À Gisela, à Eunice e ao Eduardo, pelo privilégio que é crescermos, aprendermos e sonharmos juntos. *(Always wear sunscreen…)*

Ao Dr. Abel e à Mizinha, esteios constantes para as diferentes gerações da família.

Aos amigos e companheiros da TIPS, uns heróis de todos os dias. Sem a força da vossa equipa, não era possível acreditar em muito do que aqui defendo.

Aos participantes do workshop sobre Tradução Automática Interativa e Pós-edição, que disponibilizaram o seu tempo e que contribuíram com o seu trabalho para este projeto.

A todos os meus amigos que, de vez em quando, tinham coragem para perguntar se eu já tinha acabado.

ACKNOWLEDGEMENTS

To Prof. Belinda Maia, for having had the courage to create and invest in the PhD course in Natural Language Technologies, opening a door that simply did not exist, in the intersection between research in Languages, Translation, Engineering and Mathematics; for letting me through that door and enter a fascinating world of research, and for having remained by my side, supporting and guiding a project which even I thought of abandoning, in a long road that eventually led to a form of which maybe both of us can be proud. Finally, for the strong friendship that we developed, based on her limitless capacity to understand and support.

To Prof. Thomas Hüsgen for the support and for having believed in my work.

To Prof. Pavel Brazdil, for having given me the first push into a world of complex technology, and for having trusted that I could learn how to work with it in a useful way.

To Prof. Dorothy Kenny and Prof. Joss Moorkens for having accepted to participate in the jury of evaluation of the present thesis and for the contribution to the discussion and improvement of the current thesis.

To all members of the jury of the thesis, for their work and contribution to the evaluation and approval of the present thesis.

To my brother Helder, who firmly supported this project, well beyond its final delivery, offering help and opening up multiple dimensions for fruitful discussion.

To Luís Trigo, a colleague and a friend, who crossed the same bridge I did, with me and on his own, in one direction and the other, looking for effective ways to build a useful project.

To Chris Hokamp, without whose collaboration this project would not have grown out of its speculative nature.

And…

To the unsung heroes that do research in this domain, people who so often did much more than what was expected of them, namely in research on Portuguese Language and Translation, like Prof. Belinda Maia, Prof. Eckhard Bick, and Prof. Diana Santos, among many others to whom Portugal has been indebted for a very long time.

To all researchers who not only inspired me with their extraordinary research, but also gave me the pleasure to discuss ideas and impressions, people like Alon Lavie, Andy Way, Chris Hokamp, David Farwell, Dorothy Kenny, Ignacio García, Joss Moorkens, Jorg Dietemann, Lúcia Specia, Mikel Forcada, Olga Torres-Hostench, Pilar Sanchéz-Gijón, Sharon O'Brien. I even met Alan Melby, someone I consider a hero for a very long time...

To my colleagues from FLUP, who have always supported me and showed an interest in my work, with a special mention to Elena Galvão.

To my big family, which, though far apart, always made me feel their presence, support and pride.

To Gisela, Eunice and Eduardo, for the privilege which is to grow, learn and dream with you. *(Always wear sunscreen…)*

To Dr. Abel and Mizinha, two solid rocks on which different generations of the family rely.

To my colleagues and friends at TIPS, who every day behave like real heroes. Without your strength as a team, many of the beliefs that sustain this work would simply not exist.

To everyone who participated in the workshop on Interactive Machine Translation and Post-editing, and who offered their time, feedback and activity data for this project.

To all my friends who dared to ask when I was going to finish this.

Resumo

O projeto aqui apresentado no âmbito do doutoramento em Ciências da Linguagem aborda a pós-edição como uma área privilegiada para o estudo dos detalhes técnicos que determinam a utilidade das ferramentas de apoio à tradução.

Uma vez que a literatura dos Estudos de Tradução e a literatura da Tradução Automática normalmente deixam de fora os conhecimentos da outra disciplina, foi feito um estudo cuidadoso aos trabalhos destas duas áreas. Desse estudo, resultou a noção da centralidade de quatro ações (apagar, inserir, substituir e mover) como as dimensões técnicas da tarefa de edição, tarefa que, sendo parte fundamental da pós-edição, não deve ser confundida com esta. Para clarificar os dois conceitos, é proposta a definição de um limiar que separe o esforço de edição do esforço de tradução. Esta apresentação culmina no conceito de Tradução Assistida por Conhecimento, que é um ambiente tecnológico global, no qual a informação processual recolhida durante o trabalho de tradução e edição é integrado com dados de texto, servindo de base a um sistema que aprende e sugere ações de edição e conteúdos textuais.

Esta abordagem foi testada num workshop no qual participaram cerca de 50 tradutores. Os resultados foram recolhidos e analisados a partir de inquéritos qualitativos e quantitativos, para além dos registos de atividade do software. Apesar de a aplicação que foi usada no workshop não dispor dos elementos de interatividade fundamentais para a demonstração da utilidade das funções de suporte à edição, as respostas ao workshop foram positivas, não só no que toca à perspetiva da edição que foi apresentada, mas também à possibilidade de existência de uma ferramenta de edição interativa.

Este projeto, que adotou uma abordagem especulativa, apresenta uma visão nova da pós-edição e das tarefas com ela relacionadas, abrindo linhas de investigação em redor de dimensões fundamentais dos processos de tradução, revisão e pós-edição. A compreensão destes processos pode permitir o desenvolvimento de melhores ferramentas e o reconhecimento da pós-edição enquanto tarefa especializada.

Palavras-chave:

Tecnologias de tradução; Ferramentas de apoio à tradução; Tradução automática; Pós-edição; Sistemas interativos de tradução; Tradução assistida por conhecimento.

ABSTRACT

My project for the PhD course in Language Sciences looks at post-editing as a privileged area for the analysis of the technical details that determine the usefulness of translation tools.

Since the views from Translation Studies and Machine Translation often miss important concepts from the other discipline, a thorough research in the literature was performed. This research highlighted the centrality of four editing actions (deletion, insertion, replacement and movement) as the technical dimensions of editing, a task that plays an important role in post-editing, but which should not be confused with it. To clarify the distinction, a threshold that separates editing from translating effort is proposed. The presentation of this research culminates in the concept of Knowledge-Assisted Translation, a global technological environment in which the processual information collected during translation and editing work is integrated with textual data, and serves as the basis for a system that learns and suggests editing actions and textual contents.

This theoretical framework was tested in a workshop with the participation of about 50 translators, and results were collected and analysed from qualitative and quantitative questionnaires, besides software activity logs. Although the tool that was used in the workshop did not feature the fundamental interactivity elements that demonstrate the usefulness of the editing aids, feedback from the workshop participants was very positive, not only in terms of the new view on post-editing but also on the perspective of a complete interactive editing tool.

This research project, which started as an open and speculative approach, presents a novel view on post-editing and related tasks, and it opens up several lines of research that involve fundamental dimensions of the translation, revision and post-editing processes. A better understanding of these processes may enable the development of better tools, and may lead to the recognition of post-editing as a specialised task.

Keywords:

Translation technologies; Computer-Assisted Translation; Machine Translation; Post-editing; Interactive translation; Knowledge-Assisted Translation.

List of Tables

LIST OF FIGURES

LIST OF ACRONYMS

AI – Artificial Intelligence
AC mode – Auto-complete mode (HandyCAT)
AL – Active Learning
APE – Automatic Post-Editing
AU – Alignment Unit
ATS – Applied Translation Studies
BLEU – BiLingual Evaluation Understudy
CAT – Computer-Aided Translation
CE – Confidence Estimation
CRITT – Center for Research and Innovation in Translation and Translation Technology
CRITT TPR-DB – CRITT Translation Process Research Database
CS – Computer Science
DARPA – Defense Advanced Research Projects Agency
DCU – Dublin City University
DFKI – Deutsches Forschungszentrum für Künstliche Intelligenz (The German Research Center for Artificial Intelligence)
DQF – Dynamic Quality Framework
DTS – Descriptive Translation Studies
EBMT – Example-Based Machine Translation
EM – Expectation-Maximisation (algorithm)
GALE – Global Autonomous Language Exploitation
HCI – Human-Computer Interaction
HT – Human Translation
HTER – Human-Targeted Edit Rate
IMT – Interactive Machine Translation
IT – Information Technology
ITP – Interactive Translation Prediction
KAT – Knowledge-Assisted Translation
KBMT – Knowledge-based Machine Translation
L1 – Mother tongue
L2 – Second language
LFPE – Learning from Post-edits
LgM – Language Model
MERT – Minimum Error Rate Training
METEOR – Metric for Evaluation of Translation with Explicit word Ordering
ML – Machine Learning
MQM – Multidimensional Quality Metrics
MT – Machine Translation
MTM – Machine Translation/Translation Memory
MT text/content – Machine-translated text/content
NIST – National Institute of Standards and Technology
NLP – Natural Language Processing
NMT – Neural Machine Translation
NN – Neural Network

NP – Noun Phrase

OL – Online Learning

PAHO – Pan-American Health Organization

PBSMT – Phrase-based Statistical Machine Translation

PE – Post-editing

PE mode – Post-editing mode (HandyCAT)

PEA – Post-Editing Action

PEC – Post-Edit Compare (SDL Studio add-in)

PEM% – Post-Edit Modifications percentage

PEr – Post-editing relative effort

PoS – Part of speech

QA – Quality Assurance

QE – Quality Estimation of Machine Translation

RBMT – Rule-Based Machine Translation

RNN – Recurrent Neural Network

RT – Revision Process

SL – Source language

SMT – Statistical Machine Translation

SPE – Statistical Post-Editing

ST – Source text

TAP – Think-Aloud Protocols

TCI – Translation-Computer Interaction

TEnT – Translation Environment Tools

TEP – Translation, Editing and Proofreading

TER – Translation Edit Rate

TERp – Translation Edit Rate plus

TL – Target language

TM – Translation Memory

TP – Translation Process

TAUS – Translation Automation User Society

TPR – Translation Process Research

TrM – Translation Model

TS – Translation Studies

TT – Target text

TU – Translation unit

UAD – User Activity Data

VP – Verb phrase

WER – Word Error Rate

LIST OF SOFTWARE

The list below presents in alphabetical order all the software that was used in producing this dissertation. Only legally licensed software was used, some of which in limited demo versions. The author of the dissertation is not the owner of any of this software, and he used the software as a regular end user, without applying any software modification or reverse engineering techniques, or without the purpose of producing a commercial review of any product. The sole purpose of the use of this software was academic research. Besides the software listed, the author tested other software packages that did not present results that made them appropriate for his work. Besides this, he created a few macros in Microsoft Word and Excel to increase the productivity in processing large amounts of data in these programs.

Name	Version	License	Use	Website
Adobe Acrobat	8.00	Personal	Creation of publishable versions of documents	https://helpx.adobe.com/uk/acrobat/kb/acrobat-8-9-product-downloads.html
Google Docs Forms	Online	Academic	Questionnaires for workshop	https://docs.google.com/forms
HandyCAT	Porto_v0.1	Open source	Experimental tests	https://github.com/chrishokamp/handycat
JSON to CSV Desktop Edition	1.0.6228.19245	Demo	Conversion of JSON logs to CSV	https://json-csv.com/
Kutools for Excel	16.00	Demo	Improved data processing in Excel	https://www.extendoffice.com/product/kutools-for-excel.html
Mendeley Desktop	1.17.8	Demo	Bibliography management	https://www.mendeley.com/downloads
Microsoft	2016	Personal	Data processing	https://products.office.com

Excel				/pt-pt/excel
Microsoft OneDrive	2016	Personal	Data and documentation storage	https://onedrive.live.com/
Microsoft PowerPoint	2016	Personal	Presentations	https://products.office.com/pt-pt/powerpoint
Microsoft Word	2016	Personal	Word processing and data processing	https://products.office.com/pt-pt/word
Notepad++	V7.3.3	Open source	Data processing	https://notepad-plus-plus.org/download/v7.3.3.html
Post-Edit Compare	1.7.8.0	Personal	Add-in to SDL Studio to measure edit distance	https://github.com/sdl/Sdl-Community/tree/master/Post%20Edit%20Compare and http://appstore.sdl.com/app/post-edit-compare/610/
R	3.2.2	Free	Data analysis	http://www.R-project.org/
SDL Trados Studio	2015 SR3	Personal	Data processing	http://www.sdl.com/solution/language/translation-productivity/trados-studio/
Text Inspector	Web app	Personal	Text statistics	http://textinspector.com/

INSPIRATIONS

One of the most important assignments for linguists in the future is the formulation of satisfactory theories of the nature of the translation bridges between languages. Do we really know how we translate and what we translate? […]
Translators know they cross over but do not know by what sort of bridge. They often re-cross by a different bridge to check up again. Sometimes they fall over the parapet into limbo. There is a good deal of smuggling and surreptitious evasion, and deliberate jettisoning of embarrassing difficulties.
(Firth, 1957, p. 27)

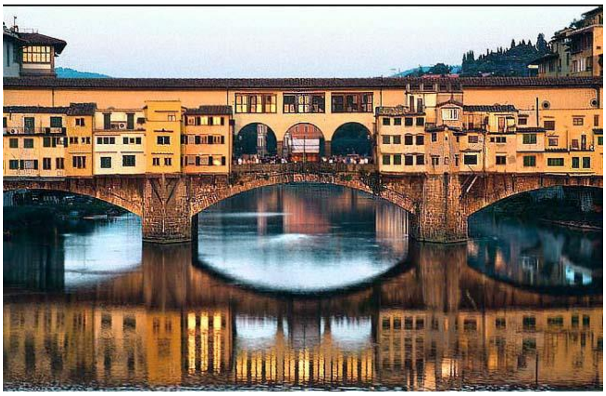


Figure 1 – The Vecchio Bridge, Florence, Italy.

Author unknown

Source: http://www.essential-architecture.com/ITALY/FL/FL-007.htm

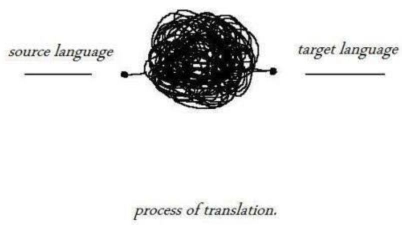


Figure 2 – The Translation Project.

Author unknown

Source: Shared in social networks (2016)

1. INTRODUCTION

1.1. Overview

This section provides a brief presentation of this thesis and its context.

1.1.1. Motivations

In 2010, the Faculty of Arts and Humanities of the University of Porto, together with the Faculty of Engineering and the Faculty of Economics, created a new PhD course, under the name "Doutoramento em Tecnologias da Linguagem Humana" (PhD in Human Language Technologies). An inherent feature of this course was giving access for students to the most advanced approaches and methodologies for the study of languages. This course, more than proposing the addition of skills, proposed a change of paradigms, an opening up of intellectual and technical horizons.

I decided to embark on this challenging project supported by my core area of expertise, translation, and I was soon studying one of the most complex activities in Natural Language Processing (NLP), as Machine Translation (MT) is so often classified. The purpose, from the outset of this project, was to apply most of the knowledge acquired to studying the translation process (TP), with a view to improving it, or at least simplifying it, or maybe just clarifying it, for the translators that every day produce millions of translated texts.

This research is guided by views on the TP that are illustrated in the "Inspirations" section of this dissertation. The citation from Firth links to the first illustration, the "Vecchio Bridge" in Florence, one of the few, if not the only one, on which people may live or make a living. This is perhaps the ideal form of a bridge – one that not only allows people to cross from one margin to the other, but one that allows people to stand on it, to communicate without taking sides, to make profitable exchanges for both parties: an ideal metaphor for the purpose and meaning of not only translation but communication itself. The second illustration describes the TP as something complex, a ball of yarn that is hard to untangle, but where, somehow, one may find the thread that links a source text (ST) to a target text (TT). If the text in each language can be described in a clear way for speakers of each language, the same does not happen in the space between those two texts. Nevertheless, each attempt to research a translation product, or a translation process, is an attempt to separate each thread in that ball of yarn and understand the way through it.

1.1.2. Breadth

The space covered by this PhD course (Human Language Technologies) is very broad, and its borders are not well defined. If we approach it from the side of Translation Studies (TS), it becomes clear that it will be hard, or even impossible, to encompass it all, and present a research project that guarantees that it applies the best techniques, among the many existing alternatives, to describe and build the best solutions for problems that have roots in different dimensions of a very broad space. So, the only reasonable approach is to define a specific area of work, identify a relevant issue in that field of work, identify approaches that are adjusted to the study of that issue, and then apply those approaches to a hopefully relevant and useful research project.

Over the course of this project, I have dealt with different approaches to scientific work, and crossed bridges from the typical Humanities and Social Sciences frame of mind to the mathematical and engineering way of doing research. However, I never actually decided to cross this bridge to the "other" margin. As translators do, I chose the less stable ground of the bridge itself, the space between two types of research, in an attempt to clarify the potential of both approaches to describe a process which is in itself essentially a communication process, but which lives with the permanent risk of miscommunication.

Nevertheless, this dissertation is essentially a TS work, which attempts to incorporate the understanding of theoretical and methodological features of disciplines on the side of Computer Sciences (CS), in the context of real situations, and by doing it, hopefully, to contribute to their development. I think this fulfils the requirements for approval of the dissertation in such a broad and ambitious PhD course. Furthermore, I hope that this may serve as an incentive for more TS researchers to explore this multidisciplinary area, even at the risk of feeling the distance from other research paths, because this domain does not correspond to the expectations of each of the stable, well-defined, but opposing margins.

A final word on this question of breadth. This project has been developed since 2011. In this period, there has been considerable progress in this field of research, which I tried to accompany not just by keeping up to date with the information published on paper and online, but also by attending and participating in international conferences and workshops. Instead of dragging me far from the initial purpose, this process drove me closer to the ideas that I started with, since research seems to be moving in the same direction I had traced for my research: an approximation of two worlds, the one of translation and the one of technologies. While this provided some reassurance about the focus of this work, several technological developments highlighted the transitory nature of research in such fast-paced domains.

1.1.3. Structure

This dissertation is composed of seven chapters, including this introductory Chapter 1, and is complemented by the Bibliography and the Annexes. Each chapter contains several sections and subsections.

This chapter, besides these initial considerations, provides a brief discussion of the problems that exist in this area of study and how contextual frameworks determine the solutions or approaches to them. These contexts include not just technological advances enabled by applied research, but also the workplace and business conditions of the localisation industry. Chapter 1 closes with the objectives and the questions that are addressed in this dissertation.

Chapter 2 “Translation Studies and Translation Technology” analyses different approaches in TS research to study and describe the Translation Process (TP). A particular focus is placed on the study of the influence of the computer tools used by professional translators.

Chapter 3, “Computer Science and Translation”, consists of a view on how the TP is approached in different disciplines that contribute to Machine Translation (MT). The different types of MT will be presented, but special attention will be given to Statistical Machine Translation (SMT). The role of Machine Learning (ML) as a provider of instruments for the incorporation of new knowledge into MT is also analysed, zooming in on uses of this technology in tasks such as predicting the needs for editing MT output, a notion related to the estimation of the quality produced by such systems.

In Chapter 4, the many dimensions of Post-editing (PE) are described as “The Post-editing Intersection”. This chapter presents the two challenges that are mentioned in the title of the dissertation: the theoretical challenge of defining PE, and the practical challenge of describing it in a way that enables the creation of tools that adequately support it as a process.

Chapter 5, “Interactive Post-editing” contains the main lines of the proposals and claims that this dissertation puts forward. Theoretical claims are presented, together with a model for an editing tool that gives support to the actions performed by translators. The framework in which these proposals fit is a form of translation that is sustained by the management of the knowledge transported in a translation process.

Chapter 6 “Testing interactive post-editing” describes the tests and results that were obtained in an experiment that involved professional translators. It will be explained how two working modes were tested, based on an experimental tool developed by a researcher from Dublin City University. An extensive process of data collection is described and the results are discussed.

The final chapter, “Chapter 7 – Conclusions”, documents the road taken from the beginning of the dissertation to its end. It brings to light those which I believe are the achievements of this project, in view not only of how it fulfils its defined objectives, but also how it may contribute to future research in this area.

1.2. Problem and approaches

Translation is a multidisciplinary field of research. The same multidisciplinarity should be seen in Machine Translation (MT). However, as will be seen in Chapter 3, for many years, MT has developed solely in the labs and departments of Computer Science and Applied Maths. During those years, MT has not been the object of study by TS scholars, and MT developers did not think much about how humans translate.

The distance between these two separate groups of researchers studying the same activity gave rise to many differences in how common terms used on both sides were understood. Even grounding terms such as “translator” have different meanings for scholars and researchers, depending on which area of study they work in. For TS scholars, a “translator” is obviously a person who translates, but this concept may confound MT scholars reading TS papers because in MT literature a “translator” is a software application that produces machine-translated text (henceforth, MT text). Other terms, such as “Linguistics”, “Translation”, “Revision” and “Editing” also developed different meanings that hinder communication between these two sides. These misunderstandings may also explain why some lines of research are so difficult to follow by researchers in the other domain.

This section will describe how these scientific fields dealt with and worked on the main problems of translation.

1.2.1. Language barriers and the inefficiencies of translation

Although since the fifties there have been clear statements about the impossibility of building systems that could produce high-quality translation with no human intervention, most of the MT projects that currently deserve major attention and funding still promise, if not really "to eliminate the language barriers", at least to "solve the problems related to them". A current example of this view is the initiative that gathers several research projects under the motto "Cracking the Language Barrier" (http://www.cracking-the-language-barrier.eu/about/).

As is probably clear by now, language barriers are not seen as a problem, just as much as rivers and canyons are not problems – there are bridges to overcome physical distances and translation to overcome the abysses between human languages. Moreover, translation is not an artificial process imposed on languages, it is part and parcel of languages, since everything in language and translation has to do with communication.

Nevertheless, translation has its problems, and one that is important for the growing localisation industry is the fact that translation is slow and skilled translators are a hard-to-find resource. In this context (which is presented in section 1.3.2 below), translation is seen as a highly inefficient process. This is the root cause of the major pressure placed on productivity and on the development of technologies that aim at increasing productivity. This is one of the focus points for this study: how to make the translation process (TP) more productive, by resorting not just to technology but also to a sound knowledge of the TP.

1.2.2. Translation Studies and the Translation Process

Attention given to the Translation Process (TP) by TS literature has been irregular. The reason for this may be found in James Holmes' papers, written in the 1960's and 1970's, but published several years later (Holmes, 1988). This book contains the first attempt to organise and give form to a discipline that still had no identity, but which became "Translation Studies" (TS). In the map of the discipline, Holmes reserves a place for the studies dedicated to the TP. The author includes these in an area of study called "Descriptive Translation Studies" (DTS), and in this we may find "Process-oriented Descriptive Translation Services". He describes the problem addressed by these studies as:

> *The problem of what exactly takes place in the 'little black box' of the translator's 'mind'* […] *Admittedly, the process is an unusually*

> *complex one, one which, if I. A. Richards is correct, 'may very probably be the most complex type of event yet produced in the evolution of the cosmos'*. (Holmes, 1988, pp. 72–73).

This statement, at the outset of a discipline, is perhaps a word of caution that made researchers shy away from analysing the TP. As will be seen in section 2.2 below, only recently have researchers developed methods to study the TP that have brought important advances to the understating of the translation processes.

The focus of this dissertation is not on the mental processes, or on the methods that try to tap into what happens inside the translators' minds. The focus is on another level of Holmes' TS map: one that is dedicated to Applied Translation Studies (ATS), namely the area dedicated to the development of "translation aids". At this point, the author presents a statement that sums up quite well the objective of this research:

> *There would seem to be a need for scholars in applied translation studies to clarify and define the specific requirements that aids of these kinds should fulfil if they are to meet the needs of practising and prospective translators…* (Holmes, 1988, p. 77).

Despite this clear identification of one of the areas in which the involvement of TS scholars is required, this was another area that did not gain much interest in TS.

One of the reasons why the development of translation aids has never been central to TS may have been the fact that the tool manufacturers' main objective has been to increase translators' output, leaving language quality to the responsibility of translator trainers, which happens to be in line with the main concern of TS scholars. A good example of this is the scarcity of papers on MT coming from TS researchers, at least until the moment when interdisciplinary teams realised that computer technologies and MT could be useful for translators. So, just as language barriers were not a problem for TS, neither did inefficiency seem to be an issue that TS teachers needed to solve: their central concern rarely shifted away from language quality.

I take a different view on this, which is built not only on the study of the everyday use of translation tools but also because my perspective on translation includes "efficiency" as a defining feature. Professional translators see translation as part of an industrial process, as described by Juan Sager, in which time and cost factors require that it should be efficient, at the risk of not fulfilling its purpose:

> *Any modification of a text is carried out in the perceived interest of an improvement of the text, as part of a specific message. It involves making texts more effective for and in a particular communicative situation.* (Sager, 1993, p. 113)
>
> […] *we must assume a strong motivation to communicate or to receive information which cannot be fulfilled in the intended manner because of a language barrier. The need for communication must be considered important enough to wait for the translation to be carried out and to engage other parties, i.e. mediators, to assist. Time and cost factors introduce a certain level of formality into the proceedings and increase the complexity of the task.* (Sager, 1993, p. 139)

It is clear that efficiency and quality must go together, since investing solely in one of them is not an excuse for producing translations that fail in the other. So, every analysis of the TP should aim at understanding it and improving it, in these two dimensions: efficiency and quality.

1.2.3. Computer Science and Translation

The history of MT research has been one of alternating euphoria and dysphoria. Even after different generations have failed in their original objective of "solving the language barrier problem", it seems all too tempting not to aim at fulfilling that objective, when a new technology comes around and shows some efficacy in dealing with the complexities of this task. Chapter 3 presents a brief description of the main lines of investigation in MT, from Rule-Based Machine Translation (RBMT) to Neural Machine Translation (NMT), but the focus will be on the currently most developed method, Statistical Machine Translation (SMT).

Many authors have studied and published accounts of the history of the different generations of MT. A few examples of such works are: "Machine Translation: An Introductory Guide" (Arnold, Balkan, Meijer, Humphreys, & Sadler, 1994), "Machine translation: past, present, future" (Hutchins, 1986), "Machine translation: a concise history" (Hutchins, 2010), "Machine translation: its scope and limits" (Wilks, 2009). "Milestones in the history of MT" (Hutchins, 2012) is a good chronology of the evolution from the early steps to 2012. In "Translation engines: techniques for machine translation", Arturo Trujillo gives an overview of several approaches to MT, from different technological solutions to different linguistic problems (Trujillo, 1999). Besides, reference books on NLP often include chapters on MT, such as "Foundations of Statistical Natural Language Processing" (Manning & Schütze, 1999) or "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition" (Jurafsky & Martin, 2009). Besides these references, a great timeline of MT history is available online, at the site of the localisation industry think-tank TAUS (Translation Automation User Society): https://www.taus.net/academy/timelines/translation-automation-timeline. And in 2015, a team from the Universities of Stanford and Washington published a 13-page article that summarises the development of interactive MT systems, which begins very early in the history of MT (Green, Heer, & Manning, 2015).

For a broad and detailed description of SMT, the primary reference is "Statistical Machine Translation" (Koehn, 2008). For a summary of the fundamentals of SMT, and how it relates to other MT paradigms, two papers are recommended: "Statistical Machine Translation: A Guide for Linguists and Translators" (Hearne & Way, 2011) and the chapter "Machine Translation" in "The Handbook of Computational Linguistics and Natural Language Processing" (Way, 2010a).

One of the main issues discussed in these works is the model of the TP used by MT systems. Wilks (2009) quotes authors that take a radical stance, arguing that no linguistic models are needed to build an autonomous MT system, a position which reminds one of a discussion on the roles of models in Science in general and Artificial Intelligence (AI) in particular (Norvig, 2011). However, there are also radical statements from those with an opposing view, with authors claiming that fully automatic MT systems will never see the light of day – one such author is Martin Kay, in a paper that will be reviewed later in this dissertation (Kay, 1997). The debate over translation models in MT, although a most necessary one, especially for one looking for connections between disciplines that seem far apart, is also a very deep and long running one. As such, if this matter were to be studied adequately, one would have to take a different road and deviate from the intended destination. For now, I shall just present the roads that the research follows, and then comment on some of the reasons for the divide between Computer Science (CS) on the one side, and Linguistics and TS on the other.

It is broadly recognised that MT research has frequently turned its back on Linguistics – which immediately brings to mind the particularly well-known, but difficult to confirm, statement by Jelinek: *"Anytime a linguist leaves the group the recognition rate goes up."* (Jurafsky & Martin, 2009, p. 214). Andy Way (2009) gives a vivid account of some of the reasons for this detachment in SMT. He says that when SMT became the mainstream model, the one that was virtually undeniably better, the researchers in that field felt that they no longer needed to communicate in a way that was easy to understand for people outside their domain. When one compares the two texts by Brown et al. (1988; 1993) that Way comments on, one cannot fail to see that the language used denotes not only an increased lack of detailed explanations of the technicalities but also the disappearance of some modesty in how scientific proposals were presented to peers. In this paper, Way makes a blunt statement on the relation between Linguistics and SMT:

> […] *we believe there to be no linguistic or cognitive plausibility in the statistical model of translation. What's more, in our view, a statistical approach is almost the least appropriate way to go about understanding and explaining human translation data.* (Way, 2009, p. 16).

However, even for authors that recognise the need for contact with Linguistics, the calls for participation of researchers from this domain are still too limited and somewhat baffling for linguists and TS researchers. As may be seen in "Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?" (Manning, 2012), the author's appeal for "more Linguistics" is not an invitation for theoretical and practical participation in the design and implementation of NLP projects, but only a call for more and better-annotated data. Moreover, in Way and Hearne (2011) although there is a clear appeal for the input of disciplines such as Contrastive Linguistics and Corpus Linguistics, there is no reference to TS research or papers from TS scholars in the paper's reference list. In fact, after mentioning that Contrastive Linguistics convey information about systematic differences between languages, which, if not accounted for, may make the output incorrect, the authors state: "*Translation studies, on the other hand, might be said to inform us about stylistic differences; one could produce adequate translations without recourse to these divergences, but not fluent high-quality translations.* (Way & Hearne, 2011, p. 10). Even for people outside TS, it should be clear that no discipline studying such a complex activity could be dedicated solely to "stylistic differences", whose absence only affects fluency. This discussion could be developed further, but that would force the discussion of the uses and misuses of the concept of "style", and that is surely beyond the scope of this work. Nevertheless, it is safe to say that MT has evolved solely on its own devices, owing little to Linguistics, or even to TS, even though that meant that researchers were faced with a challenge:

> *Those without a linguistic background, then, appear to have two choices: (i) to attempt to include the linguists, so that they may be of help; or (ii) to continue to exclude linguists, while at the same time trying to make sense out of their writings.* (Way, 2009, p. 19).

Before moving on, it is important to mention a dual notion that is pervasive in both TS and MT literature, regarding the uses of MT. That is the notion of "assimilation" and "dissemination" (Hutchins, 2010). This duality refers to how MT results may be used, depending on different levels of quality: if MT systems produce texts that can only be used as indicators of the "gist" of the source text, then they are said to be appropriate for "assimilation". "Dissemination" is only possible when MT texts are of good enough quality to be read and understood.

1.2.4. Post-editing

There is one overlapping area between the research carried out in TS and CS, and that is the area which describes what happens when MT results are used not as finalised products, but as intermediate, somehow incomplete products that need to be improved before being ready for "dissemination". This is the main link between MT and TS: MT creates translation suggestions, better known as "hypotheses", with different levels of quality, and human translators intervene to guarantee that the final versions present "human-like" quality. The activity performed by translators when they receive these MT texts that need to be improved is known as "Post-editing" (PE).

One of the most comprehensive analyses of the history of PE is "A brief history of post-editing and of research on post-editing" (García, 2012). This paper shows how PE has been around since the beginning of MT. This somehow goes against the general impression that MT has always looked at being autonomous, but it may be explained by another fact that defies the general knowledge of the earlier history of MT. It is a bit surprising to realise that some of the most controversial texts in the history of MT, like Bar-Hillel's "The present state of Mechanical Translation" (Bar-Hillel, 1951), the infamous ALPAC report (Pierce et al., 1966) and Martin Kay's paper mentioned above (Kay, 1997), which are briefly quoted as pure and even cruel criticisms of MT, are fundamentally apologies for cooperation between MT systems and human translators, as a means to promote the use of these technologies. These texts will be further analysed below, in Chapter 4, and PE will also be further discussed in this dissertation.

1.3. Academic and industrial context

As many authors point out, MT is going through a favourable period and it lives in a fostering context, with several research projects and a booming industry pushing it forward. This is an important factor that must be taken into account by any research project that wants to adjust to reality and to be ultimately useful.

1.3.1. Research in Translation

It is hard to tell whether MT research is still living a moment of euphoria – which may have begun with Google Translate's launch in 2006 (Google, 2016) – or if its promised success story is coming to a halt, having reached a limit, defined, for example, by the "data glass ceiling" (Wiggins, 2011). Some consider that the current state of MT is at the level of mainstream use, so much so that its degree of adoption may help determine the level of maturity of translation companies (Vashee, 2016a). The recent developments in Neural Machine Translation (NMT) gave rise to a new wave of enthusiasm. However, no matter how far the extent of application has improved, it is still early to determine whether this enthusiasm will last for long. The effect of this technology on the perspective of this thesis will be discussed in section 3.3.2 below.

Nevertheless, even if one cannot determine the current place of MT in a curve of euphoria, it is far from dysphoria. In fact, the publication of a report that may call off funding and deviate commercial attention away from MT seems to be unlikely in the near future. MT is still heavily funded, and there are plenty of projects that receive a lot of commercial and public attention, such as the aforementioned "Cracking the Language Barrier Initiative". Apart from European funding, which covers many countries and projects, countries like Spain (Faes, 2015) and Ireland (Diño, 2016) are still investing strongly and positioning themselves at the forefront of this research movement.

If this was not strong enough evidence of the exciting momentum for MT, one just needs to search for regular and specific events organized by institutions such as AMTA–Association for Machine Translation in the Americas (https://amtaweb.org/), EAMT–European Association for Machine Translation (http://www.eamt.org/), ACL–Association for Computational Linguistics (https://www.aclweb.org/) or ICCL–International Committee on Computational Linguistics (http://nlp.shef.ac.uk/iccl/), which, every two years, organises COLING–International Conference on Computational Linguistics, and many others, to realize the extent of research currently being produced in this domain. Interestingly enough, one of the longest running conferences in the world of translation is devoted to translation technology, and it is attended by many MT researchers: the "Translating and the Computer" conferences, held since 1978 (http://www.asling.org/tc38/).

A deeper search on those events shows that most of these papers still come from CS labs and that there are only a few outstanding research projects that are run by or include the input of TS specialists. Dublin City University (DCU) (https://www.dcu.ie/) is an excellent example of a multi-disciplinary institution, where you may find computer scientists working in the same projects with TS scholars, in dynamic and highly productive teams. Another great example is the Center for Research and Innovation in Translation and Translation Technology (CRITT), of the Copenhagen Business School, which is one of the most prolific institutions specialised in TPR – Translation Process Research: (https://sites.google.com/site/centretranslationinnovation/home).

This profusion of research makes it difficult to keep up with the advances: there are literally thousands of pages presented and published every year just on MT. To accompany the pace of research in this domain, one needs not only to look for the proceedings of these numerous conferences but also to check regularly publications such as the "Machine Translation" journal published by Springer (http://link.springer.com/journal/10590).

Another feature worth noting of research in MT is that most of it follows the "open-source" model. Since the Moses toolkit, developed by Philipp Koehn's team at the University of Edinburgh (Koehn et al., 2007), that was made available even for commercial use, virtually every piece of research in this domain has been published and distributed openly. This enables researchers to test and try different technical approaches, but it has the downside that it assumes that all researchers possess deep computer skills and have been able to accompany the technical evolution.

1.3.2. Translation Industry

No studies on MT, or on translation technologies at large, should be done without some knowledge of its industrial surrounding. The presence and influence of the translation industry may be seen in several research projects, but it also constitutes a testing field for the validity of the theories and applicability of the methods and tools proposed by academics. Companies that play a major role in this dynamic industry are interested in applying knowledge coming from academia, but at the same time invest in developing their own solutions. And, as if this was not enough to justify the reference to this side of the translation world, institutions that fund research projects also foster the increase of proximity between academia and industry.

Most current successful research projects involve industrial partners, so as to make sure their results are applicable, they are tested in the "real world", and they have proved their worth to the global community. MateCAT (https://www.matecat.com/) and CasMaCat (http://casmacat.eu/) are two examples of products that were created by projects that involved universities and companies.

But commercial companies also have their own research and development teams, with an opposite attitude towards publishing their research. Most Computer-Aided Translation (CAT) software is sold with a license for use and contains proprietary code, developed by the owner company, which is seldom published or discussed in translation forums, except as a means to advertise their products. This makes it difficult to study and analyse the development and integration of new features in CAT tools. The commercial success of this software has, nevertheless, been a sound basis for the continuous evolution of the field, since the industry depends very much on having tools that receive, transform and return content to the most advanced applications and platforms. CAT tools accompanied the evolution of web tools, multimedia platforms, the "cloud", and every other technological environment where linguistic data needs to be transferred across languages.

Working for this industry, or within this industrial context, are independent consultants such as Kirti Vashee, company owners like Dion Wiggins (both quoted in the previous section) but also freelance translators who maintain regular publications on which they present profound and knowledgeable views about technology and the industry, like Jost Zetzsche (http://www.internationalwriters.com/aboutus/jost.html), Kevin Lossner (http://www.translationtribulations.com/), or consulting companies like Common Sense Advisory (https://www.commonsenseadvisory.com/). A great source of news on this industry is also the website Slator (https://slator.com/).

The need for consensus around complex technical and professional issues, the need to publish research carried out at universities, and the global pressure around this industry, has led to the creation of associations and groups of interest that discuss the common problems of this sector, and search for the best solutions to move it forward.

Associations like FIT–Fédération Internationale des Traducteurs/International Federation of Translators (http://www.fit-ift.org/), GALA–Globalization & Localization Association (https://www.gala-global.org/) or especially TAUS–Translation Automation User Society (https://www.taus.net/) are involved in discussing the future of not only technologies but the whole setting of professional translation.

Complex industries, at the centre of constant methodological changes and content exchanges, demand the standardisation of procedures, data formats and tools. The localisation and translation industry has created a few industry standards, most of which focused on quality, like the European norm for Translation Services EN15038:2006, which has now been superseded by ISO17100:2015 (ISO, 2015). For a discussion of the impact of these standards, one should resort, for example, to (Koby & Melby, 2013). But this industry has also produced standards regarding technical requirements for the interchange of data. Examples of these are XLIFF (https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xliff), a file format that allows the interchange of bilingual data, independently of the tool it has been created in, or TMX (http://xml.coverpages.org/tmx.html) that standardised the way Translation Memories (TMs) are published and shared.

A specific resource that was missing in the translation industry standards were metrics for evaluation of the quality of translation. The two most important metrics proposed by joint industrial and academic projects (MQM–Multidimensional Quality Metrics and DQF–Dynamic Quality Framework) were recently merged, under the auspices of TAUS, into a single one: the "Quality dashboard" (https://www.taus.net/evaluate/about-quality-dashboard). This metric became a standard for the industry, and currently, CAT tools are incorporating it as a specific stage in the TP. However, as this phase is external to the TP itself, this evaluation stage is not included in this study.

This brief description of the business universe that surrounds translators and translation technologies is included in this section of the dissertation not just for contextualisation, but also because all these institutions, companies and people have a voice that is globally accepted as representing the "real world" of translation. The economic power behind this voice should not be underestimated. It is not uncommon to find these institutions and their reports and standards cited as the most reliable sources for academic publications. This means that economic thinking is pervasive, making it sometimes difficult to separate academic research from simple commercial and marketing communication. One effect of this is the appearance of terms that, although not consensual for TS academics, make their way into mainstream communication about translation. One such term is “localisation”, a term which is still being discussed as meaning something that is, at the same time, “more” and “less” than translation, but which everyone still uses. This is a short list of papers that include some of the best discussions on this term: (Bravo & Enríquez, 2006; Melby, Fields, Hague, Koby, & Lommel, 2014; Pym, 2002, 2010, 2011a).

1.4. Objectives and research questions

The objectives that were formulated for this dissertation have been adjusted over a long period of time. As will become apparent in the coming chapters, some of the questions and ideas that sustain this dissertation have only arisen after a detailed exploration of different theoretical and technical areas, and they have evolved with the continuation of the project. The objectives and the questions that guide this dissertation are presented next.

1.4.1. Objectives

The main objective of this project is to obtain a clear view of the PE process, in comparison to the translation and the revision processes. This will be approached through a deep analysis of how these processes are described in TS literature. Then, with a good description of the PE process, one can check whether some of the ideas and techniques applied by MT to produce a translated version of a text may be used instead to support the work of translators. To achieve this, the technologies used to create a translation hypothesis will be analysed, and it will be checked whether these are suitable to support a PE workflow. A thorough analysis of current translation tools will also help contextualise the proposal of a tool that aims at managing the content and resources that are produced in technological contexts in which MT is combined with Computer-Aided Translation (CAT) tools.

1.4.2. Research questions

The first proposition that is presented in this dissertation is that PE is a mixed process, in which translators not only edit but also revise and translate sub-units of a sentence that was previously translated by an MT process.

However, this view collides with the simplified views of the external manifestations of the TP conveyed by the literature of this area. The initial chapters will describe how such a view can be found both in the TS and in the MT literature. So, the search began for a simplified view of PE, focused on an external and technical view of the process, based on simple editing actions performed by translators.

The second, very simplified, proposition was that PE could be described, defined and analysed based on four simple editing actions: deleting, inserting, moving and replacing words and multi-word units in a MT sentence.

The second proposition has the advantage that it may be simpler to implement by technology, so it may be a better connecting element with the MT technology that is going to be studied. This simplified view of PE became the guidance for the first stages of the work, and the first proposition played the role of a fall-back plan, if it proved to be the most correct description of PE.

The best description of the PE process, either as a combination of translation and revision, or as a simplified process, should allow for an approach to MT processes that revealed which technology might be used to support it. The role of Machine Learning (ML) as a technology that enables the extraction of information that may be reused in complex processes soon became apparent and the focus concentrated on that.

So, the first research question was posed: could Machine Learning (ML) techniques applied in MT be used to present suggestions to post-editors, for each of these four editing actions? Considering that useful computer tools must always be adapted to the intended uses, but that this is often not sufficient, another question, focused on users, was also brought up. If translators could have these PE aids incorporated into the computer tools that they use every day, would they be more efficient during PE?

Two propositions and two research questions guide this dissertation, from the theoretical questioning to the technical tests that are planned. However, these propositions and questions will only be discussed again at the final section, when the balance of the work is done. In view of the complexity and multidimensionality of the object of study, I assume that this open and speculative approach may enable rich conclusions and fruitful results, which may be applicable in the current technological context, and may be useful for translators.

2. TRANSLATION STUDIES AND TRANSLATION TECHNOLOGY

This Chapter begins by presenting the main concepts and approaches to the TP in TS research. Next, it focuses on how the development of assisted-translation tools affected these views, and finally on the processes of "translation", "revision", and "editing".

2.1. Building blocks of a theory of the translation process

As was explained in Chapter 1, the theoretical framework for the TP fits within Descriptive Translation Studies (DTS). This section will present the main guidelines in this branch of TS and identify the main building blocks of a view of the TP that allows the study in some depth of the practice of translation with the aid of computer tools.

2.1.1. Starting from Descriptive Translation Studies

To begin the search for a theory of the TP that encompasses the tools that translators use, we must go back to two of the most productive decades in TS (the '80s and '90s of the last century) and to the field of DTS. This stage will focus mainly on two authors: Juan Sager and Gideon Toury.

In his book "Language engineering and translation: Consequences of automation" (Sager, 1993), Juan Sager presents a model of the TP in the context of translation as a type of text modification, motivated by the needs of a communication industry. He mentions that the electronic tools that are used determine the strategies of translation, and he approaches the translation activity from the point of view of the "process of understanding", divided into three dimensions: cognitive, pragmatic and linguistic (see section 2.2.1 below). His description then moves on to highlighting "dictionary look-up" as one of the tasks that follow the whole translation process, from the reading comprehension phase (oriented towards the source language – SL) to the translation phase (bi-directional) and the revision phase (oriented towards the target language – TL). Although very systematic, his work suffered from the passing of time, as can be seen in one of his concluding remarks: *In the future, the most important new tools for translators will be collections of existing translations.* (Sager, 1993, p. 209) This inescapable effect of any research into technology (to refer to a future which will soon be past) is compensated for by his vision: the collections of existing translations are, in effect, at the basis of not only the omnipresent CAT tools, but they also sustain the most current and popular forms of MT. Sager's extensive work embraces other concepts that are central to the study of the TP.

Central concepts in Translation Theory

Sager explains the main lines of a theory of translation:

> *A theory which says something about how one gets from a source language text to a target language text must justify the stages in this process and must describe these stages in terms of reception and production, cognitive units, psychological processes, matching of equivalents, etc.* (Sager, 1993, p. xix).

He focuses on the notions of "translation equivalence" and "translation unit", as he describes translation work as an iterative process of looking for equivalents (for units of varied sizes) and compensation techniques to resolve the lack of those equivalents. His discussion of "translation techniques" starts with Vinay and Darbelnet's presentation of categories of production of translation and is complemented with his own and other authors' views on the subject. He finishes his description with short references to the stages of "text production/encoding", evaluation and revision, and finally presentation (Sager, 1993, pp. 211–242). In the next sections of his book, he shows how his vision of the TP encompasses not only human translation (HT) but also MT, although he only analyses RBMT systems.

Gideon Toury is a very respected author because of his sagacious descriptions of regularities in translation activities, which he terms "translational norms". Although it is possible to follow his discussion on "process-oriented empirical studies", in which he explores several dimensions of the decision model that might adequately describe the complexity of translation, the main focus of his work is on what he calls the "matricial norms". He presents a good reason to move in this direction: *"What one is after is (more or less cogent) explanatory hypotheses, not necessarily 'true-to-life' accounts, which one can never be sure of anyway".* (Toury, 1995, p. 95) Thus, if we cannot be sure of what "really happens" during translation, we must look at the target text (TT) and describe its relation to the source text (ST) by trying to identify the substitutes of corresponding source language material. In such a task, we are looking for the degree of "fullness" of translation, which is closely linked to procedures of *"omissions, additions,* *changes of location and manipulations of segmentation”*. (Toury, 1995, p. 59) Let us retain this list of actions as the ones that Toury highlights when he comments on how to approach what “really happens” during translation.

These two brief summaries of the works of these authors provide some of the central concepts of any theory of translation with a special focus on the TP. In the Translation Theory literature, one may find references to dimensions that are specific to translation, known as the “universals of translation”. These are *“linguistic features that typically occur in translated rather than original texts and are thought to be independent of the specific language pairs involved in the process of translation*” (Baker, 1993, p. 243). These features may include: “lexical simplification”, “explicitation”, “adaptation” to the norms of the target language, “equalizing” – adapting oral text to written conventions, and absence of “unique items” (words, concepts, or collocations that only exist in the target language) (Pym, 2010), but they may also include phenomena known as “normalisation” and “interference” (Baker, 1999). Nevertheless, these are seen as the traces that the TP leaves behind: when these effects are identified in a translated text, the differences between source and target language cannot account for their presence. If these phenomena were easy to explain, TS research would be totally focused on these features, as the ones that fully justify their specificity as part of a scientific domain. However, even these features are somewhat elusive, and it is difficult to build a whole theory of translation and of how it is done based on these concepts alone.

In the next sections of this Chapter, the focus will close on the notions of “translation shifts”, “equivalence”, “translation unit”, and “translation techniques”.

2.1.2. Equivalence

“Equivalence” relates to the stable relationship between elements in the source text (ST) and its translation, or target text (TT). This concept has been a source of argument throughout the history of TS but has remained, in one sense or another, in all the work that looks for similarities and tries to establish relations between units in both texts.

Snell-Hornby is critical of this term, but a closer reading of one of her papers shows that she admits that some texts may be analysed in terms of “equivalence”: technical texts. The reason for this is that, unlike other types of texts, technical texts allow us to look for symmetries, not between language systems, but between “isolatable lexical items". (Snell-Hornby, 1988, pp. 32–34) So, we are in the domain of "contrastive lexicology", visible in a specific type of text, which is associated with "restricted equivalence" at the word or phrase level only.

Lederer takes a different view and underlines that only a restricted understanding of equivalence sustains common statements like those concerning the untranslatability of words from one language to another. As the author points out, these statements often misunderstand what translation is, since all words are translatable, even if a single word needs to be translated into a multi-word unit, a descriptive phrase or even a sentence (Lederer, 1986, pp. 27–30). Sager also discusses the concept of translatability, relating it not only to the concept of equivalence but also to the translation unit:

> *This involves a discussion of the existence or absence of units of equivalence between linguistic items of two languages, possibly established according to text types, sublanguage, text segments or whatever other criteria may be considered to be relevant. The detailed examination of such units could lead to ranges of equivalence from 'one-to-one' matches of units, signifying total semantic and pragmatic identity, to 'no match' which would mean that a unit of text is untranslatable in the conventional sense of the word…* (Sager, 1993, p. 131).

Sager takes a dynamic view of equivalence, highlighting its usefulness as an operational term: *"While appropriateness is a valid criterion for measuring the success of a speech act, a concept of equivalence or correspondence is still required in order to deal with the evaluation of the result of the process at the micro-levels."* (Sager, 1993, p. 145) Here, we are clearly not in the domain of the description of what goes on during the TP, but on how to describe it from its results, the translation products.

Sager goes on to say that *"There must, however, be a sufficient number of culture-independent or culture-neutral units within the source text to establish synonymy between pairs of languages so that the possibility of constant content can be realised."* (Sager, 1993, p. 135). Sager uses different verbs to refer to equivalence: "search for equivalents", or "determining the units of equivalence", or even "establishing" or "matching" (1993, pp. 222–233). When he refers to translation techniques, he refers to them as *"means of creating equivalences"* (1993, p. 144). This shows not only that the term is quite useful to describe different aspects of the practice of translation, but also that it is dynamic, in the sense that it may be adapted to different approaches on the TP.

2.1.3. Translation shifts

After having presented briefly concepts that describe regularities, let us look at how TS deal with irregularities. The term "shifts" has been adopted by TS to describe *"all that appears as new with respect to the original, or fails to appear where it might have been expected"* (Shuttleworth & Cowie, 1997, p. 153). This concept was introduced by Catford (1965) but extended by Popoviĉ. The concept of shift is very productive, in the sense that it shows that similarities are just as justifiable in a translated text as dissimilarities, and that both are connected to translation decisions. Gentzler comments on Popoviĉ's contribution in the following terms:

> *Shifts have invariably been attributed to deliberate distortions, incompetence on the part of the translator, or linguistic incompatibility between the two languages. Popoviĉ extends the theoretical horizon by analysing shifts in terms of the differing cultural values and literary norms. Instead of accusing translators of ignorance or unfaithfulness, Popoviĉ argues that they resort to shifts precisely because they are attempting to render faithfully the content of the original despite the differences between the languages.* (Gentzler, 1993, p. 86).

Bassnett-McGuire (1980, pp. 138–139) presents the full list of Popoviĉ's shifts:

- Constitutive shift – shifts based on the differences between languages;
- Generic shift – changes at the level of the constitutive features of text genre;
- Individual shift – deviations due to the translator's style;
- Negative shift –information incorrectly translated, due to unfamiliarity with the language or structure of the original;
- Topical shift – changes in facts of the original.

As one can see, "translation errors" are part of this list, but they are side-by-side with other reasons for deviations from a literal rendering of the source. One can divide the other types of shifts into two categories: those that are caused by differences between two languages, and those that arise from decisions made by the translator.

Toury (1995) comments that both "so-called obligatory shifts" and "non-obligatory shifts" are types of translation norms, since they may be considered regularities of the TP.

2.1.4. Translation techniques

After having presented the units and relations (parallelisms and dissimilarities) between source and target texts, let us now focus on the techniques, methods, procedures, or other descriptions of how translators transform a text from one language into another.

The most significant work classifying the techniques employed by translators is a book first published in 1958 by Vinay and Darbelnet (1977). This book, a manual of "compared stylistics", attempts to identify procedures to compensate for all the differences between the two languages and to enable the transfer of the message of the ST to the TT. It was so influential that most recent lists of techniques still incorporate most of its terms and concepts, despite all the criticisms it received. The purpose of the book was to identify the differences between French and English and to list the techniques that translators might employ to overcome these differences.

The list is introduced with the basic concepts of "direct translation" and "indirect translation", in a description that has a parallel with the notions of "literal" or "equivalent", and "shifts" or "compensation". The full list of procedures includes:

- borrowing words or phrases from the source;
- calquing, or importing and slightly modifying phrases from the source;
- literal translation, which does not require explanation;
- transposition, which is a grammatical change, from a part of speech to another;
- modulation, which is a substitution of an expression by another one which has the same effect in the target culture;
- *équivalence*, which is the replacement of idioms by their equivalents;
- adaptation, which is the full replacement of a whole sentence by a cultural equivalent.

There are several subsequent lists of techniques (or a synonymous term), such as the ones by Newmark (1984, 1988), Leuven-Zwart (1989, 1990) or Malone (1988). A short paper by Molina and Hurtado Albir (2002) aims at revisiting this concept. In this paper, the authors present a clear position concerning translation techniques: these are, in essence, tools for translation analysis. Besides, they also clearly separate what they consider to be "strategies", which are concerned with the TP, and "techniques", which are related and visible in an analysis of translation products.

The main point to take from these works is that none of them has been able to achieve consensus in the TS community. This lack of consensus is due to different reasons, some of which will be discussed in the next section when these concepts are approached from the point of view of today's TS literature. However, if we move back to an author from the pre-TS stage, we will find a very simplified view of what he calls "techniques of adjustment": *"Here, we are concerned, therefore, not with why the translator does one thing or another, but with what he does, in terms of additions, subtractions, and alterations."* (Nida, 1964, p. 226) After having seen so many complex lists of procedures, one can only be amazed at the elegance of such a description of what a translator does.

2.1.5. Current views on these concepts

On moving forward in time, and searching in more recent TS literature, one finds that these terms are present in all books and papers that attempt to organise the terminology of TS, such as the "Dictionary of Translation Studies" (Shuttleworth & Cowie, 1997), the "Routledge Encyclopedia of Translation Studies" (Baker, 1999), "Key Terms in Translation Studies" (Palumbo, 2009), the "Handbook of Translation Studies" (Gambier & Doorslaer, 2010), Anthony Pym's "Translation research terms: a tentative glossary for moments of perplexity and dispute" (Pym, 2011a), or "The Routledge Handbook of Translation Studies" (Millán & Bartrina, 2013). This section analyses works by Anthony Pym, an author who is known for his prolific publication and his broad and evolutive views, thus allowing an up-to-date presentation of these central terms to Translation Theory.

A good example of a work that deals with concepts such as the ones this thesis presents is "The Moving Text: localisation, translation and distribution" (Pym, 2004). In a chapter dedicated to issues of quantity, Pym starts from the concept of "NANS – No-Addition-No-Subtraction" as an ideal for a translation that is perfectly balanced regarding quantity. However, he admits that effective translation rarely obeys this principle since it depends on a reasonable relationship between the ST and TT elements. There is another principle that connects to this "NANS" principle. Shuttleworth &

Cowie call this the *no-leftover principle* (1997, pp. 112–113), a technique termed by Toury to be used in Descriptive Translation Analysis. This concerns the need to define the limits of identical units so that there is nothing in a specific ST segment that is outside the TT segment that it aligns with, or missing from it. This is a methodical way to enable translation analysis, and it is related to the *manipulation of segmentation* that the same author proposes as an operational norm. (Toury, 1995, pp. 78–79)

Throughout the rest of the chapter, Pym presents the "hidden logics" that justify different gradations on the cline between ideal sameness and radical discrepancy. He relates these with "quantitatively based equivalence", and presents four modes of presentation:

- transliteration, where a foreign language string appears in the TT,
- double presentation, where the translator presents the source expression side-by-side with its translation;
- single presentation, where only the translated elements are presented;
- multiple presentation, when several alternatives are presented.

In the discussion about "single presentation", he comments on four things that may affect quantity:

- expansion – the same semantic material is expressed by a greater textual quantity;
- abbreviation – the same material is expressed by smaller textual quantity;
- addition – new semantic material is added, increasing textual quantity;
- deletion – semantic material is removed, reducing textual quantity.

Even if the author does not use all of the terms mentioned in the previous sections, it seems clear that this discussion involves "equivalence", "translation units", "shifts" and "techniques".

Pym uses the term "translation solutions" as a better alternative to "techniques" (Fawcett, 1997), "procedures" (Newmark, 1984, 1988), "strategies" (Chesterman, 2016) or even "strategies and tactics" (Gambier & Doorslaer, 2010). The term "solutions" is associated with *"what translators produce as potential or final end-points of the problem-solving actions".* "Translation actions" are *"what we actually observe translators doing (e.g. typing, correcting typographical mistakes, looking up terms in* *glossaries, etc."* (Pym, 2011a). Later, he proposes his own typology of translation solutions that stem from three basic concepts: copying, changing expression and changing content. The author considers that the main purpose of this typology is pedagogical (Pym, 2016).

Pym has a pessimistic view of the role occupied by TS in the current scientific scenario. In a talk with the title "Where Translation Studies Lost the Plot" (2015), Pym comments on how TS were removed from the central stage and became a minor discipline. Bibliographical research may confirm that since 2000 the literature production in TS seems not to have been able to introduce new terms and views to replace the tradition created in the golden decades of the 1980's and 1990's. This is obvious when we look at the bibliography of any recently published TS book, and we realise that most of the references are still from those years. The author relates the loss of attraction of TS with the detachment of this discipline from language teaching.

The next section studies one area of TS that challenges Pym's pessimistic view on TS, and in which there is a growing interest, with visible and promising results from research.

2.2. Process-oriented Descriptive Translation Studies

This section will present descriptive approaches that try to tap into the translator's brain, or the "black box", as described by Holmes (section 1.2.2. above), in an attempt to understand or describe the TP. According to Muñoz Martín (2014), between 2006 and 2013, at least 11 books compiled more than 100 chapters devoted to Translation Process Research (TPR). Sharon O'Brien says that this interest may be due to *"a thirst for a greater understanding of translation as an expert task."* (O'Brien, 2013, p. 5) Let us, again, see how the TP was described in the 1990's before moving ahead to more recent views.

2.2.1. Psycholinguistic models

According to Alves & Hurtado Albir (2010), the first studies of translation as a cognitive activity go back to the "Interpretive Theory of Translation" in the 1960's and the works of Danica Seleskovitch, which were devoted specifically to interpreting. Several models of the TP were proposed in the 1990's, but the authors comment that most of these were not based on empirical data. For this introductory section, let us look at a model of the TP from an author who was referred to in the previous section.

Sager explains the reason why we need to look at human factors to get a more reliable picture of the TP: only human factors can explain the reasons why different translators produce different translations of the same text (Sager, 1993, p. xix). This author establishes a connection between the search for equivalents and the “minimax principle”: translators look for solutions that achieve the maximum effect using the minimum effort (Shuttleworth & Cowie, 1997). On a first reading, translators identify the pragmatic equivalence, at the level of the message and the type of text, then they look for the cognitive units for which it is necessary to find correspondences, and finally they try to find the corresponding smaller units at the nearest grammatical level (Sager, 1993, p. 145).

Sager’s psycholinguistic model, part of which is built on previous work by Lörscher, includes the following characterisation principles:

- experienced translators use mentally less complex techniques before more complex ones;
- translation is a forward and backwards looking mental operation: moving forward in the text, but going back to recover previously processed information. The translation phase is iterative since it implies dealing with relationships that the different sized working units establish not only with the units of the source text but also between the units that are used in the translation;
- the speed of translation increases in direct proportion to the production of target sentences, i.e. the translator works faster the more he advances in the translation.

Sager explains that some of the backwards movements performed by the translator imply rewriting, rephrasing, or editing parts of the translation. As he moves along the text, the translator reuses elements that appeared before, thus simplifying his decision process and speeding up his work. According to the author, the way translators deal with ambiguity reveals some of these forward and backwards movements, and the strategies to compensate and clarify these issues (Sager, 1993, pp. 213–235).

No matter how interesting and insightful this description may seem, one cannot help wondering at the number of tests that would be needed to confirm that this a true account of what actually happens in real-life translation assignments. Testing the adherence to reality is one of the main concerns of current TPR.

2.2.2. Methodological grounds of current process-based studies

As we will see in the rest of this Chapter, current TPR is heavily dependent on the methods of data collection and analysis that are used. This section briefly describes most of these methods, by summarising an article by Birgitta Englund Dimitrova (2010).

The TP has been defined in a very simple formulation: *"The cognitive activity of producing a target text in one language, based upon a source text in another language"* (Dimitrova, 2010, p. 406). But it has also been presented in a longer fashion:

> *The translation process is defined as everything that happens from the moment the translator starts working on the source text until he finishes the target text. It is all encompassing, from every pencil movement and keystroke, to dictionary use, the use of the internet and the entire thought process that is involved in solving a problem or making a correction – in short everything a translator must do to transform the source text to the target text.* (Hansen, 2013, p. 88)

However, since the *"cognitive processes involved in performing a translation task are not available to direct observation"* (Dimitrova, 2010, p. 407), the research methods applied gain a capital importance. What is at stake is the "ecological validity" of the study, i.e. the situations in which translation production is observed, recorded, and from which all elicitation is made, need to reproduce real-life situations. Otherwise, if the observation and recording methods force the creation of artificial testing environments, this may distort the conclusions and its applicability to the real world.

Think-Aloud Protocols (TAP)

One of the first methods to be applied by TPR projects was "think-aloud protocols" (TAP), in which translators verbalise their thinking in one of two fashions: concurrently, as they proceed with the translation, describing problems, solutions and decision processes as they tackle them, or retrospectively, by revisiting troublesome excerpts of a finished translation task. However, researchers admit that these methods are not very reliable because they depend on the capacity of the translators to describe their thinking accurately. For a discussion of how these methods relate to consciousness, see (House, 2000).

Jakobsen (2003) studied the effects that TAPs have on translation work, analysing different features, like time or segmentation. In terms of time, he identified delays due to verbalisation in the TP, but not during revision. But the most significant result of his study was on segmentation, which is the way translators move ahead processing language information by splitting it into chunks, bigger or smaller, according to criteria that are difficult to identify. The author suggests that when a researcher looks for these units in the ST, he may produce a bias towards mental processes that are related to text comprehension, whereas when he searches for those units in the TT, this may result in a bias towards processes that have to do with text production. However, his most important conclusion is that, no matter how the processes are analysed, when translators needed to verbalise their actions, for research purposes, they tended to process smaller units than when they worked uninterrupted, in their usual rhythm.

Other methods for the collection of process data

The use of computer technologies enabled a whole set of not-so-intrusive data collection methods, like logging different levels of activities performed in front of a computer. The most important software used in these projects is Translog, developed at CBS – Copenhagen Business School (Jakobsen & Schou, 1999), and later CasMaCat (Alabau et al., 2013), as shown below, in the section dedicated to post-editing. With this software, researchers could register keystrokes, activities like typing and deleting words, or moments of pause. This data could then be supplemented by other data collection methods like video recordings, direct observation, different forms of verbalisation and questionnaires.

Technologies like eye-tracking have been used to record the points on the screen on which the translator fixes his view and to register eye movements. For a discussion on the challenges of these techniques, see (O'Brien, 2009). The analysis of pauses has been associated with the cognitive effort of each translation task (Lacruz, Shreve, & Angelone, 2012).

There have also been attempts to use neurological and physiological data collection systems, such as PET (positron emission tomography), EEG (electroencephalograms) or fMRI (functional magnetic resonance imaging), in some restricted contexts, especially related to the task of switching between two languages

(Diamond & Shreve, 2010). However, not only are these methods hard to include in typical TPR research projects but they also usually defy the ecological validity required for this research (Hansen, 2013).

To improve the reliability of the data collected, it is generally recommended that a combination of methods should be used, in a methodology known as "triangulation" (Alves, 2003).

Although in TPR methods of data collection and analysis are fully integrated with the objectives that are set out, in the next sections several TPR projects will be analysed disregarding the methods used, so that focus can be kept on the themes they study.

2.2.3. Recent theories, themes and hypotheses

In TPR, there are virtually no papers that do not present some theoretical innovation in the approach to the TP, with a strong link to the methods and experiments performed. It would be impossible to present a thorough, yet critical, analysis of most studies in TPR without devoting a long part of the dissertation to both theoretical and methodological discussions. So, the option was to comment on fragments of reflections that relate to this particular research, and to leave aside the experiments and results of each study.

This section presents some of the papers that best represent the connection between new theories, hypothesis testing and method development in TP literature, as a way to highlight the evolution of the concepts presented in section 2.1 above. Two volumes include several of these papers and are a good introductory approach to this area of study: (Shreve & Angelone, 2010; Tirkonnen-Condit & Jääskeläinen, 2000).

The three phases of the TP

Most authors in TPR adopt a tripartite view on the TP. Different authors use different names for the three phases of the process, but the most common are:

- orientation (but also planning and reading);
- drafting (but also development, and generation of the target text);
- revision.

The terms "orientation", "drafting" and "revision" tend to be used as the most consensual ones, and are used by many authors, such as Dimitrova (2010).

Translator profiles

TPR projects have focused on issues like “translator profiling”: trying to identify different behaviours according to the expertise level of the translators. The main motivation for this is the fact that many translation research projects only involve students as subjects, and there was a long tradition of criticism on the lack of ecological validity of these studies. This field of research also involves the notion of “translation competence” which is related to translation learning.

The PACTE group, based in Barcelona, has two major publications in this domain. In the first one (Beeby et al., 2009), they present their model of translation competence and analyse factors like language proficiency, extra-linguistic knowledge and knowledge about translation, but also instrumental and strategic skills. In a context of problem-solving, they highlight the importance of these last two skills, which may be observable in the efficiency that translators show in using external and internal support to solve the problems that they face. The second publication by the PACTE group (PACTE Research Group, 2014) focuses on the acquisition of translation competence, but this theme will not be analysed in this dissertation.

Another example of a TPR project focused on translator profiling is “Towards a classification of translation styles based on eye-tracking and key-logging data” (Dragsted & Carl, 2013), a paper in which the authors identify individual behaviours in each of the 3 phases of the translation process. They classify these behaviours as stable and state that these can be observed across different text types, thus allowing the identification of novice vs. expert translators. Martínez-Gomez et al. (2014) extend Dragsted and Carl’s research results by applying quantitative data analysis to identify characteristics of the users.

In a different study, Michael Carl, Barbara Dragsted and Arnt Lykke Jakobsen (2011) identify different translation styles or profiles, according to behaviours in each of these three phases. These are the most relevant styles:

- Orientation phase: translators who systematically plan/orient the translation before they start typing, those who just skim through the ST before writing, and those who start translating right away;
- Drafting phase: translators who read long sequences and then translate them in one go, and translators who often backtrack while translating;

- Revision phase: some translators do “online revision”, during the drafting stage, others are constantly revising, and the rest do end revision only.

Translation as problem-solving

Another approach, not so much focused on classifying translators according to how they work, sees translation as a complex of problem-solving strategies (Dimitrova, 2005). A concept associated with this approach is “uncertainty management” (Angelone, 2010; Tirkkonen-Condit, 2000). In Angelone’s paper, a different sequence of three “translation-oriented cognitive processes” is proposed:

- Source language comprehension;
- Source language to target language transfer of meaning;
- Target language text production.

In Angelone’s paper, there is a “transfer of meaning” phase before text production, and there is no separate revision phase. The author links this to a view of how translators overcome and describe (using metacognition or consciousness) the mental processes when they are faced with uncertainty or indecision. As the author points out, this seems to be the right stage to observe and study the TP, because it is when a problem arises that the cognitive flow of the three processes is interrupted, creating the ideal situation for observation and allowing for the definition of units, processes, sequences, and other related concepts.

To analyse the reaction to uncertainty, the author presents another tripartite description of behaviours, which may occur within each of the cognitive processes above: problem recognition, solution proposal and solution evaluation. The bundle or sequence of such behaviours forms what the author terms a “cognitive translation unit”. The author describes the last part of this process as the “editing activities”, such as “additions, deletions and revisions”, which he classifies as “production-behaviours” (Angelone, 2010, p. 21).

An article in the same volume (Dragsted, 2010) discusses in detail the relation between these processes, looking specifically at the coordination between reading and production modes, contextualised by visual data, keyboard events and pause. The author uses a technique that is called “eye-key span” that associates eye fixation on an SL word with the time that spans before the equivalent word in the TL starts being typed. This data is used to study differences in processing modes between professionals and students. The conclusions mention a difference in how professionals coordinate comprehension and production in an integrated or even overlapping way, to the point of starting this coordination process before the actual translation task, and how students do this, sequentially, with longer sequences of only one of the processes, higher eye-key spans and more frequent pauses.

Although Barbara Dragstedt admits that *“we continue to be in the dark about much of what goes on in the translator’s black box”* (Dragsted, 2010, p. 59), this is a study that confirms the descriptive capacity of cognitive studies to identify human factors that help explain the TP.

2.2.4. The translation unit

It all begins with the definition of what translators work on: what is their unit of work? This notion has been discussed since the beginning of TS. Holmes suggests that earlier theorists considered the sentence as the main structure that guides translation. One such case is Hans Vermeer (1986). However, Holmes adds, as we translate each sentence we have a map of the text in our minds, and thus translation is actually a multi-level dynamic process: at a serial level, we may be working on each sentence at a time, but at a structural level, we operate at the textual map level (Holmes, 1988).

Sager goes below the sentence level, contextualising this discussion in a psycholinguistic framework: translators segment the text in longer units when they work in specialised familiar fields, whereas when they are unfamiliar with the technical language, they work with smaller segments. *"There are particular difficulties associated with the segmentation of noun phrases in technical texts into terminological units as opposed to free syntagmatic combinations."* (Sager, 1993, p. 225). Toury contributes to this view by adding the manipulation of segmentation (choosing dynamically longer or shorter units) as a matricial norm, as explained in the previous section.

In my master’s dissertation, I studied the reasons behind the higher number of words in a translation compared to its source text, by taking the noun phrase (NP) as the unit of focus. I analysed NPs that kept that surface structure in the translation, with different extensions and at different levels of the syntactic tree, as long as they contained the same content and the same head noun. Focusing on this unit proved to be a wise decision, since it contained the majority of the text transformations that lead to an increase in the number of words in translations from English to Portuguese. A summary review of this work may be seen in (Carmo, 2010).

The clause seems to be an important unit for some authors. In his book “Translation and Language: Linguistic theories explained”, Peter Fawcett quotes Anton Popoviĉ: *"The main field where the translator's decisions take place is the level of the textual microstructure"* (Fawcett, 1997, p. 64). This discussion appears in the context of the criticisms of translations that are too “literal”, or too “word-bound”. But Fawcett concludes in a way that agrees with Sager, above:

> *What professional and even novice translators actually do is relate the translation of the microlevel of words and phrases to higher textual levels of sentence and paragraph, and beyond that to such parameters as register, genre, text conventions, subject matter and so on, in a constant dialectic.* (Fawcett, 1997, p. 64)

Sager highlights the fact that the discussion on translation units is related to the notion of equivalence. He presents a definition of the translation unit from other authors: *“l'unité de traduction est le plus petit segment de l'énoncé dont la cohésion des signes est telle qu'ils ne doivent pas être traduits séparément.”* (Vinay & Darbelnet, 1977, p. 37).

For now, we can move on with the notion that there is no consensual “translation unit”, and that this discussion depends on other factors, such as the notion of equivalence.

2.2.5. Identifying units in a continuous process

In some of the studies presented above (see, e.g. Angelone’s) the authors use some units to organise the data that is collected. This section will look in more detail into a few projects in which this concept is expanded, allowing for the development of different approaches to the TP.

Dynamic definition of units

The first study discussed here, “Translation units and grammatical shifts: towards an integration of product and process-based translation research” (Alves, Pagano, Neumann, Steiner, & Hansen-Schirra, 2010), refers to nearly all the core terms in TS that were mentioned in section 2.1 above. The translation unit (TU), translation universals, equivalence, shifts and translation techniques all play a role in this study, which handles data from TP but also analyses translation products, following the tradition of corpus studies (see section 3.2.2). The study includes work with a particularly advanced annotated parallel corpus that includes multi-level alignments. The authors explain that this richly annotated corpus enables the identification of translation shifts. The absence of shifts is observable when, in a translation of two sentences, all levels of alignment (words, phrases, grammatical functions and clauses) are parallel. In contrast, a translation shift is identified when there is either an empty link (an element in one of the languages does not align with any of the elements in the other language), or when there are crossings (the only way to build alignments is by crossing from one level to the other, e.g. from word to phrase, or when a clause becomes a phrase). The authors conclude that the TU cannot be defined in single-level one-to-one alignments because translators dynamically navigate between units and levels.

The purpose of this study was to map units that are observable in the translation products to different elements in the records of the TP. The TU becomes a key item for the description of the TP: *"translation units are identified in time by pause intervals reflecting the translator's focus of attention."* As for the production side, the authors use the term "production segment", as being a "*text extract observable in target texts or in the text production process"* (Alves et al., 2010, p. 125).

Micro TUs and macro TUs

The authors explain further ahead that there are two types of TUs: micro-TUs are continuous flows of TT production; macro-TUs are collections of disparate micro-TUs and of text production segments that are linked to the same textual ST segment, encompassing everything that occurs in that segment, between the first attempt to generate a translation and the final action in the corresponding TT segment.

Both micro and macro-TUs are composed of text production segments (which include *"revisions, deletions, substitutions, etc."* or *"deletions, additions, and other possible changes"),* located between two pauses, within a defined time threshold. Special note must be taken of the inclusion of a term like "revision" (which is also used in the paper to describe one of the phases of the TP), side-by-side with simple actions like "deletion, "addition" and "substitution", in a list that also includes the vague term "other changes", all of these terms being used as examples of so-called "text production segments".

The final part of this study is dedicated to showing how the different methods of corpus and process analysis allow the authors to identify the occurrence of "grammatical metaphors" in a corpus of monolingual and parallel texts. This term was proposed by Halliday and Martin (1993) to explain the transformation of verbs into nouns, seen as a very economical procedure, particularly useful to the language of science. Grammatical metaphors, also called "grammatical shifts" by the authors, are identified in translation when the TL word that expresses the meaning of a word in the ST is not in the same grammatical category (or part of speech) as the SL word. The clearest examples of this process are "transpositions", a term used by Vinay and Darbelnet (1977) to describe, for instance, nominalisations, when the main verb of a clause in the ST is transformed into the head of an NP in the TT. In this study, evidence of these transformations is identified not just by comparing source and target sentences in the rich-annotated corpora, but also through the identification of macro TUs that contain the sequence of actions that result in these textual changes, in the logs of the TP.

Although this paper shows some terminological inconsistencies (most likely due to its multiple authoring process), the authors have published further papers that reveal the interest of this approach, such as "Modelling unpacking of meaning in translation: insights from effortful text production" (Alves, Pagano, & Silva, 2011) and "On drafting and revision in translation: a corpus linguistics oriented analysis of translation process data" (Alves & Vale, 2011), both described next.

From the three phases of the translation process mentioned above, the first two seem to be easier to isolate and study: "orientation" involves the ST and may be analysed from eye movements, whereas "drafting" is usually recorded by logging keystrokes performed on the TT side. However, it is not so easy to distinguish "drafting" from "revising", since these actions only affect the TT, they are recursive, and they may even be simultaneous. Alves and Vale (2011) describe a process that enables them to study these two stages, using the identification of micro and macro-units of translation as the basis for this.

The authors clarify that "micro-units" are flows of TT production segments (or typing and editing actions) within two pauses of at least five seconds. "Macro-units" are sets of micro-units that are related to the same ST segment. A macro-unit associated with a specific segment may include discontinuous micro-units that may even occur in different phases. For example, micro-units related to revision actions may occur in the same text segment while the translator is drafting (and he goes back a few words to replace a word) or later, when he comes back to rephrase a sentence he had drafted a long time ago. The authors identify three types of macro-units:

- Macro-units composed of micro-units (editing actions) that occur only at the drafting stage – the translator makes several changes to the same textual segment while composing the first version of the translation; in this case, the translator does not come back to edit the segments again;
- Macro-units which are drafted only once and which are only revised at a later phase – at the drafting stage, a single version is typed in from beginning to end, with no edits. However, later, at the revision phase, the same sentence is revised and edited;
- Macro-units that reveal several edits, both at the drafting stage and at the revision stage.

Then, the authors classify their subjects according to patterns of incidence of these types of macro-units. "Drafters" are translators that make most edits during the drafting phase, "Revisers" are those that make most edits during the revision phase, and "Drafter/Revisers" are those that make edits both in the drafting and in the revision phase. The latter type may then be classified as "recursive" or "non-recursive", according to the frequency of changes to the same segments in both phases.

In this study, Alves and Vale also discuss the issue of time linearity in process data logging. In order to identify dynamic movements upwards and downwards in the text (when translators change their first translation of one word because it appears in a new context below or above), one cannot simply rely on data logging, and one needs to resort to product or text analysis. However, the notion of macro-units simplifies tracking and identifying the traces of these decisions throughout the text.

User Activity Data

The next set of studies was developed at the Copenhagen School of Business, with the help of extensive databases of empirical data collected by software such as Translog and later CasMaCat.

Michael Carl introduced the concept of "Alignment Units" (AUs) to contrast, or complement, that of TUs. Carl suggests that, since *"we do not really know what TUs are",* (dynamic, multi-level process entities), we may accept AUs as more stable entities, if we consider them as "*translation correspondences in the static product data"* (Carl, 2009, p. 227). He adds that there is an isomorphism between AUs and TUs in research on MT, but that is not the case in HT. AUs are used to extract units from product data (parallel corpora), and TUs to register process data (keyboard log data). The details are then organised in a database of 4-tuples, each composed of product and process data: ST unit, TT unit, AU, and Keyboard data (process details, or TUs). The set of all this data is known by the acronym UAD – User Activity Data.

In the same year, in another paper with Jakobsen (Carl & Jakobsen, 2009), Carl discusses the difficulties in trying to use data that is collected through eye and keyboard activities to describe what goes on in the translator's mind:

> *It is, however, largely unclear how this process is organised in detail, what it is exactly that makes a text difficult to translate, how translation difficulties can be detected and how a translator could best be helped with translating or post-editing texts more efficiently.* (Carl & Jakobsen, 2009)

They further explain that UAD tries to relate textual (product) data to temporal (process) data. In this work, the database contains 5-tuples, because fixation data is added to the data mentioned in the previous work. The focus of this work is on writing and reading patterns.

A parenthesis needs to be opened here to comment on the use of the term "post-editing" in this paper. The term is used not just to refer to the practice of revising MT text, but also as the name of the final phase of the TP, as a full synonym of "revision". The same use appears in other papers and authors. Carl reuses it in a different paper in which the three "translation production" phases are: "gisting, drafting and post-editing" (Carl, Kay, & Jensen, 2010). The authors also use "MT post-editing" and "revision" in the same text, which does not help to clarify terms that have meanings so close to each other.

2.2.6. Open questions in Translation Process Research

The studies that closed the previous section discuss and look at the contribution that TPR may have on the development of translation tools. But there are other open questions in TPR.

In fact, open questions seem to be a feature of TPR. Dimitrova comments: *"process-oriented studies have so far had to a large extent an exploratory and* *hypothesis generating character"* (Dimitrova, 2010, p. 408). Séguinot (Séguinot, 2000) confirms that this is a natural feature of empirical studies: as researchers test hypotheses, new hypotheses appear.

Uncertainty seems, rather than an object of study for TPR, to be an attribute of TPR itself. Tirkonnen-Condit (2000) recalls that no one knows the final result of a translation until the process is finished. She adds that no two processes are the same, even if the task is the same, and that there are as many products as there are translators. Séguinot (2000) analyses translators' decisions, looking specifically at the points when translators do not know how to translate a certain word. She believes that this decision process underlies the whole TP, but it is very difficult to establish a connection between decisions and textual units. She admits, for example, that every time a translator sees a repeated word, he rethinks the decision he made previously and this may not be reflected in any action. Besides, a translator may rethink decisions while he is reading words that are not clearly related to what he is thinking. This means that the decision process is not only non-linear, but difficult to analyse from textual data. She even goes on to say that *"meaning on which the translation operates can be self-generated, i.e. that it arises during the course of the translation, rather than being housed in the source text"* (Séguinot, 2000). She calls these mental processes *peripheral and managerial* because these decisions are not evidenced by a textual analysis.

Despite these challenges, TPR has continued to collect data and analyse it in both products and processes of translation. But neither is this an easy task, since these data are dynamic, change over time, are based on units that link across different levels of analysis, and describe activities that are overlapping. Alves and Hurtado Albir see this as one of the achievements of the approaches of TPR:

> *The non-linear nature of the process: It neither follows a linear textual progression nor is it constrained to the sequential development of its basic stages. Therefore, it allows for regressions, i.e. recursive movements in text production, and alternations between the phases of understanding and re-expression.* (Alves & Hurtado Albir, 2010, p.34).

However, developing the theories and the methods to achieve this is an ongoing, unfinished process. A more recent paper (Carl, Lacruz, Yamada, & Aizawa, 2016b) comments that there are still no research results that allow us to know, for example, how to relate temporal data and TP data. There is no way to know, for instance, what happens during pauses and how that is connected to translation problems and strategies.

2.2.7. Outtakes of this section

TPR is a ground-breaking, fundamental, research field for TS. It covers all translation objects that may be studied, from textual products in two languages to the process data that describes how an ST became a TT. Besides, the foundations of this area of TS are solid, imposing on researchers the definition of clear objectives and appropriate methods of analysis to try to answer the hypotheses that they raise. The complexity and detail of the data collected (from single keystrokes to eye movements, including time recordings of microseconds) also demand careful and rigorous analysis.

This overview led to the collection of several concepts that will be often retrieved throughout this dissertation. One of these is the discussion around the TU, a central concept in this field because of how TS scholars handle data: do they want to segment translation products into linguistically motivated units? Then, a strong linguistic framework and reliable linguistic analysis tools are needed. Do they want to identify the units of analysis through the use of automated tools? Then, they need a theoretical background that clearly discriminates these units, and tools that follow specific rules to identify those units. However, the initial question is whether one is dealing with text units, observable and segmentable in translation products, or if one is dealing with processes or actions, which are harder to analyse and segment.

TPR demonstrated different ways to do research on translation, and it offered a view of the TP in phases, separating and clarifying mental and production actions and their products. From this view, it has shown the development of different techniques of analysis for the different phases, different objects and different subjects being studied. Section 4.1 below will discuss TPR research focused on the post-editing process. For now, let us move on to a different area of TS, which also deals with the TP, but from an instrumental perspective.

2.3. Applied Translation Studies

It is not only the process of investigating and analysing translation that is intimately connected to the tools that are used. Translation itself has evolved according to the technological context in which it is practised. Michael Cronin goes back to pre-writing civilisations to show how the relationship is so intricate as to define translation itself: *"Translation without tools simply does not exist."* (Cronin, 2003, p. 24) So, it is no wonder that, as seen above (section 1.2.2), Holmes defines a specific area within TS for the study of translation tools: Applied Translation Studies (ATS).

In revising Holmes' work of organising the map of TS, Toury (1995) suggests the need to extend the discipline, namely to amplify its "applied" branch. He refers to "bridging rules" that should support the "applied extensions" TS must establish with other disciplines. These "applied extensions" concern the instruments necessary to close the gap between Translation Theory and translation practice, in fields such as translation teaching, translation quality assessment, and the one of interest here, the development of translation aids.

Rosa Rabadán (2008) suggests that these concepts have not been sufficiently explored to have an impact on the applied practice of translation. The author explains that academics tend to consider that "applied" is related to things that should be left to users, and that do not seem to be of interest to research. Ultimately, this has an impact on users, who complain about the lack of interest in the tools that are offered to them. Rabadán suggests that this type of research should be done according to two very specific requirements: usefulness and usability. It is the failure in these two evaluation criteria that justifies the frustration with some of the tools that are proposed to translators: they are not useful, in the sense that they do not solve existing problems, or they are not usable, meaning that they do not contribute to improving what translators do. In a different paper, the same author states that these two aspects (usefulness and usability) should inform all stages of ATS research:

> *They ensure the user-centred nature of ATS by playing a fundamental role in the applied process, from needs analysis, research tools design to the analysis, and leading to the operationalization and conceptualization of the applicable parameters.* (Rabadán, 2010, p. 9)

She adds that these evaluation factors are linked to the concept of "cognetics", also known as "cognitive engineering". Byrne describes this as "ergonomics of the mind", a notion that he presents as fundamental to the evaluation of software centred on users, but also for the practice of writing and translating technical documentation (Byrne, 2004, 2006).

The sections below start by presenting literature on the tools that translators use for their everyday tasks, then analyse the different generations of these tools, highlighting their central features, and then come back to ATS research, to discuss the effects of these tools on the processes and products of translation.

2.3.1. Studying translation tools

Translation tools became an object of study in the 1980's, but gained more visibility in the 1990's, with the appearance of personal computers. The history of these tools may be found in papers like "The origins of the translator's workstation" (Hutchins, 1998) and in different sections of several of the papers presented below.

Early proposals for the design of translation tools

Martin Kay, in a paper first published in 1980, suggested the development of a *translator's amanuensis,* a device he believed would probably never be built (Kay, 1997). The description of this tool is strikingly similar to the tools translators use nowadays, although it includes some features that are not implemented in modern software applications. In several papers published in the early 1980's, Alan Melby presents plans for similar tools (Melby, 1982, 1983). In one of these papers, the term "Computer-Assisted Translation" is used for the first time. Other descriptions of translator tools in the 1980's are Sugdens' presentation of ALPS – Automated Language Processing System (Sugden, 1985), Stoll's evaluation of translation technologies (Stoll, 1988), and Olsen's account of INK TextTools (Olsen, 1988). In 1992, Sergei Nirenburg presented a model for an interactive translation tool that serves both translators and researchers (Nirenburg, 1992). Melby continued to study and propose models for translation aids throughout the 1990's (Melby, 1992; 1995, pp. 183–189) and he continues to do so (Melby, 2006, 2015).

Publications and sources of reference for translation technology

It is important to highlight some of the publications that dedicate special attention to these fundamental aspects of professional translation. Magazines and journals focused on professional translation have always devoted relevant sections to translation technologies. A few are worth a special mention. "MultiLingual" (https://multilingual.com/), a magazine which has been published since 1987 on paper and online, "JosTrans: Journal of Specialised Translation" (http://www.jostrans.org/index.php), an online journal founded in 2004, and "Revista Tradumática" (http://revistes.uab.cat/tradumatica/index) which has been published in three languages (English, Spanish and Catalan) since 2001, are the ones that dedicate close attention to themes around uses and effects of translation technology.

The earliest books dedicated specifically to translation technology had a special focus on MT. “Computers in translation: a practical appraisal” (Newton, 1992) is a good example of that. “A practical guide for translators” (Samuelsson-Brown, 1993 & 2010), first published in 1993 and followed by five editions, highlights the importance of technology for professional translation. Another popular book, with two editions, was Bert Esselink’s “A Practical Guide to (Software) Localization” (Esselink, 1998, 2000), which presents translation technology in the context of the localisation industry.

The first book totally devoted to translation technologies for translators, with only a small chapter at the end on MT, is “Electronic Tools for Translators” (Austermühl, 2001). One year later, Lynne Bowker’s book on CAT tools, “Computer-Aided Translation Technology: A Practical Introduction” (Bowker, 2002), virtually ignored MT. In 2003, a book gathered several articles on specific subjects associated with technology for translators: “Computers and Translation” (Somers, 2003). Another book worth mentioning is “Translation in the Digital Age” (Cronin, 2013), in which the author explores in detail the cultural, philosophical and political implications of the evolution of tools and technologies for translation. In 2015, an “Encyclopedia of Translation Technology” (Chan, 2015) was published, which is a vivid demonstration of the extension and importance that this field has acquired.

Some authors argue that the effects of technology on translation have been so extensive as to justify a specific discipline, a reorganisation of the map of TS and even a turn in TS, triggered by this “technologization” of translation (O’Hagan, 2012). But there are still technological areas that seem outside the scope of TS.

For ATS, MT is interesting as an extra resource, more than a tool, for translators. Although some attention has been given to MT, this was never studied by ATS scholars from the point of view of how MT works, how it emulates translation or the TP, or even of how MT tries to incorporate linguistic knowledge. In ATS, MT is, at best, discussed as a black-box that sometimes provides useful translation suggestions to complement other translation resources. A similar attitude will be taken in this chapter, as it focuses only on processes and tools that do not involve MT.

2.3.2. Computer-Aided Translation tools

The generalised use of CAT tools since the 1990’s was fundamental for the growth and stabilisation of the translation industry. These tools not only proved themselves a valuable support for the tasks performed by translators but created a huge volume of digital bilingual data that sustained the position that translation currently occupies, at the centre of the language technologies world. As will be seen below, in Chapter 3, the importance gained by such technologies as SMT could not have been reached without the availability of huge corpora of translated content, already aligned and classified, produced with CAT tools.

Commercial tools will not be discussed in detail in this dissertation, except when a specific technology needs to be commented upon. But there are several sources of reference for such analyses, as the report (LT-Innovate.eu, 2013), industry associations that publish reports such as https://www.taus.net/think-tank/reports, market consultancy services such as http://www.commonsenseadvisory.com/Research.aspx.

Introduction to CAT

A good introduction to these technologies is "Computer-aided translation" (Bowker & Fisher, 2010). In this article, one may find an explanation of most of the different terms that describe this technology, besides CAT, like TEnT (Translation Environment Tools) and TM (Translation Memories), and a description of the various modules that compose these tools, like aligners of bilingual text, concordancers for contextual word searches, terminological databases, project management modules and quality assurance features. Another feature that is clearly explained is the classification of segments in the ST in terms of their similarity to SL segments previously stored in the TM, side-by-side with their translations into the TL. Three types of segments stand out: "full matches", when the correspondence between the ST segment and the SL segment in the TM is total; "fuzzy matches", when there are similarities but not total correspondence; and "no matches", when the system has no stored SL segments that have similarities above a specific threshold. There is no consensus regarding the fuzzy match threshold, but the most common threshold applied by the industry is a 75% similarity score. This applies to ST segments in which 75% of the words are similar to a SL segment in the TM. The remaining 25% of the words in the segment have been updated and need to be checked or edited.

The translation editor

Ignacio García presents a broad view on these technologies, from the history of the development of these tools to a perspective of their future. The central piece of software in this text, and the one that this work focuses on too, is the editor:

> *... the system frontend that translators use to open a source file for translation, and query the memory and terminology databases for relevant data. It is also the workspace in which they can write their own translations if no matches are found, and the interface for sending finished sentence pairs to the translation memory and terminology pairs to the term base.* (García, 2014, pp. 70–71).

The editor application is where all the technologies converge: translators work in the editor, and it is inside the editor windows that all information taken from the data resources must be reused.

García refers to two different modes of work in the editor:

- “interactive mode”, by which translators work within a target empty window or pane, where they type in the whole translation, while getting suggestions from the TM and the termbase and deciding how to incorporate these into their TT;
- the “pre-translation mode”, in which the target window already presents either TL suggestions coming from the TM, or SL segments which had no correspondences in the TM; in this case, translators write over or edit these TL suggestions or SL segments.

These two modes are based on a study by Wallis (2006, 2008), which points to a general preference of translators for working in “interactive mode”. However, García notes that “pre-translation mode” is currently the most common mode of work.

Particular attention is given in García’s article to sub-segmental reuse, terminology extraction and the treatment of tags. Translators use the concordance feature to compensate for the limitations of CAT tools to identify translation suggestions below the segment level. García says that this method is inefficient and random, and that a technological solution is called for. That solution is called “sub-segment matching” or “advanced leveraging”. In this article, García presents some of the solutions available, with a prominent position taken up by Déjà Vu’s earliest “Assemble” proposal, and how it links to predictive writing systems. He adds that some may see this type of technology as a useful feature, but others may consider it a nuisance. At the end of the article, the author comments on the integration of all these technologies in future versions of CAT tools, including other capacities like speech recognition and linking to massive online TMs. An alternative and a complete source of reference for analyses of these tools is "State of the Art in Translation Memory Technology" (Reinke, 2013).

2.3.3. First generation of translation tools

In an article that discusses translation technologies in the context of translator teaching, Austermühl admits that CAT tools have not evolved a lot since their appearance in the 1990's: *"The way tools are accessed has changed—in many cases moving to a software as a service (SaaS), or cloud model—as has their level of sophistication, but the kind of support they give to translators, the sort of tasks they carry out, really have not."* (Austermühl, 2013, p. 327). He goes on to say that the tools that translators use, and the skills necessary to use them efficiently, are basically the same as the ones he describes in his 2001 book. Although a radical and isolated view, if one considers what goes on inside the editor windows, his statement was quite valid. But something changed in more recent years. This view will be used in this section to forcefully split what is a continuous flow of technology development into two separate generations of translation tools, each presented in a different section of this chapter.

The relative importance of the source text

Austermühl's 2001 book "Electronic tools for translators" is based on a notion of the TP divided into 3 phases, just as TPR does. In his case, the terms for the 3 phases are reception, transfer and formulation. (Austermühl, 2001, p. 13) He approaches the use of tools from reception of not only the ST but also resources like terminology databases or electronic dictionaries. A bit surprisingly, he does not devote many pages to the analysis of the work done inside the translation editor (2001, pp. 134–152). His main focus is actually on how translators use available information, like the one that is provided by TMs, to formulate their translations. The definition of translation he presents later may explain this: *"I see translation as the self-confident production of an independent target text based on a number of information sources, one of which is the source text."* (Austermühl, 2013, p. 330).

In 2013, Anthony Pym also comments on the loss of the central place of the ST:

> *In such cases, there is no one text that could fairly be labelled the source (an illusion of origin that should have been dispelled by theories of intertextuality anyway); there are often several competing points of departure: the text, the translation memory, the glossary,*

> *and the MT feed, all with varying degrees of authority and trustworthiness. Sorting through those multiple sources is one of the new things that translators have to do, and that we should be able to help them with. For the moment, though, let us simply recognize that the space of translation no longer has two clear sides: the game is no longer played between source and target texts, but between a foreign-language text, a range of databases, and a translation to be used by someone in the future.* (Pym, 2013, p. 492).

The position occupied by the ST could naturally be discussed in the context of many different theories of translation, but that is clearly not the focus of this dissertation. However, in the light of this definition, it becomes less surprising that most of Austermühl's 2001 book is dedicated to different sources of information and to techniques of how to research through them. He explains in 2013 that research is one of the most important skills for translators and translator training.

Traditional CAT tools

For a brief look at where translation editing environments were in the early 2000's, "Evaluating translation memory systems" and "Comparing Basic Features of Translation Memory Tools" (Zerfass, 2002a, 2002b) are valuable sources of information. In these two papers, Zerfass comments on the two main editing interface models: one based on the common word processor Microsoft Word linked to a separate TM package, like Trados Translator's Workbench, and the other one being the standalone editor, which showed ST and TT in two side-by-side columns. She then discusses issues that differentiate available tools, such as the ways in which they count words and identify similarities in fuzzy matching.

In the following years, translation software only took baby steps to integrate new features into their editors, playing what Zetzsche called a "*game of catch-up"* (Zetzsche, 2015). The reason for this might well be the acknowledgement that once translation tools established a good community of users, any radical change in the editing environment might have a serious impact on their commercial position. This was visible in the transition period between SDL's acquisition of Trados, in 2005, and the moment it finally launched the first version of Studio, which did not include the Trados Workbench application, in 2013. (Filkin, 2013a; SDL, 2016)

Research on how translators worked and how tools should support them continued. In Canada, a team of researchers investigated alternatives to fuzzy matching, by proposing software that looked for similarities in words and word groups (Simard & Langlais, 2001). However, this system was an alternative to TM systems, not a complement, and that is perhaps the reason why this technology did not gain the momentum it might deserve (Gow, 2003). The same team of researchers presented a technology to incorporate what was later known as "predictive writing suggestions" (Langlais, Sauvé, Foster, Macklovitch, & Lapalme, 2000), again with no extensive impact on commercial tool development.

2.3.4. Next generation of translation tools

In the early 2000's, the first proposals for new approaches to the development of translation tools appeared. The first attempts were discrete and did not gain much traction, but there was a common denominator that attracted increasing attention.

Support at the sub-segment level

In 2005, Emmanuel Planas presented a computer application that extended the capabilities of TMs, by applying linguistic knowledge to analyse the TT and the ST to identify alignments for groups of words, usually in the form of NPs and VPs (verb phrases). In his paper, "SIMILIS–Second generation translation memory software" (Planas, 2005), he shows how redundancy is higher in sub-sentence units than at the sentence level in several types of text. He sees this as an indication that tools should provide assistance below the sentence level that is typical of CAT tools. He shows how his system expands what he calls "first generation TMs", since it analyses these resources to provide more detailed suggestions to translators. Although his system did not gain wide recognition, the research focused on the main feature that CAT tool manufacturers were investigating and investing in at the time.

At the same conference, Grönroos and Becks (2005) claimed that their company had invented a completely new method for CAT, which they called "Intelligent Translation". In 2008, Mitkov and Corpas suggested that a third generation of these technologies might be based on the use of semantic information to select word-level suggestions (Mitkov & Corpas, 2008). However, none of these technologies seems to have gained enough traction power to be further developed.

Kuhns (2007) reported on several commercial technologies that explored the concept of sub-segment alignment, or, as he terms it, “advanced leveraging”. This, he said, was a brand-new generation of CAT tools, which applied technology from different MT solutions and linguistic intelligence. In 2010, at a joint workshop between researchers and users, Andreas Eisele and Caroline Lavecchia presented a suggestion for the integration of MT suggestions to fill in the gaps of fuzzy matches (Eisele & Lavecchia, 2011).

A tool that was developed in a MT lab announced itself as: “The CasMaCat project: The next generation translator’s workbench” (Ortiz-Martínez et al., 2012). Although this tool did not achieve the goal of becoming a reference for translator workbenches, it has become a fundamental tool for TPR research.

In a technology survey done in 2015, Zetzsche highlights the fact that sub-segmentation, together with advances in other technologies like web translation environments and quality assurance features, have reshaped the way we see and use translation technology. He adds that the concept of TM has changed, since we now use smaller, more specialised TMs while expecting higher quality sub-segments, and that the same happened to terminology tools, which are supposed to deliver quality suggestions as we type (Zetzsche, 2015). In the same year, Carla Parra Escartín presented a system that responded to requests from professional translators. These had requested a system that improved concordance searches, by offering translation suggestions from sub-segment fragments that already existed in TMs. The system that was presented managed a combination of TMs to achieve the best results. (Parra Escartín, 2015).

Applications that offer support at the phrase level

So, it seems that the biggest challenge for this new generation of CAT tools was not on developing the solutions for finding sub-segmental matches, but on how to incorporate them into the existing tools. In recent years, several implementations appeared, and Jost Zetzsche reports on a few, in the context of the connection between TM and MT:

> *Examples include auto-suggestions of subsegments of partial machine translation segments (Wordfast Classic /Anywhere, Trados Studio, Déjà Vu, CafeTran), some even from more than one translation*

> *engine; repair of fuzzy translation memory matches (Déjà Vu); validation of MT suggestions with the help of TM matches (Star Transit), identification of TM subsegments with MT matches (Lift); interactive MT suggestions based on what has already been entered (Lilt); automatic formatting or target texts based on machine translation lookups (Lilt); and the list could and will go on and on.* (Zetzsche, 2016b, p. 27).

Nonetheless, this still seems to be an open matter, worthy of further analysis. From 2013 to 2015, Kevin Flanagan presented his "segment recall" technology, which permitted CAT tools to align units below the segment level without the need for big resources, statistical analyses, or linguistic knowledge (K. Flanagan, 2013, 2014, 2015). In these papers, he reviews other commercial implementations of similar features, explaining how these solutions were meant to solve the limitations of concordance searches, namely the randomness in the searches, the time wasted searching for words and the fact that results were presented in full segments, making it difficult to identify the corresponding sub-segment units. His own system was later acquired by SDL and incorporated into the 2017 version of SDL Trados Studio under the name UpLift (Zetzsche, 2016c). Flanagan mentions that this type of system demands good integration into translator tools, not only to guarantee their usability but because the productivity gains may represent a good alternative to post-editing (K. Flanagan, 2014, p. 19).

In 2016, a workshop entitled "Next Generation Translation Tools" was held at the University of Swansea. Among the participants were Kevin Flanagan and Emmanuel Planas, who presented their works in a panel dedicated to "New tools and features". Flanagan introduced the implementation of his tool in SDL Trados Studio 2017, and showed how it performed "fine-grained alignment, sub-segment recall and fuzzy match repair". He revealed that the tasks the software performs to repair fuzzy matches are: "change", "deletion", "insertion" and "movement", and showed how sub-segment choices not only allow for the composition of new matches from different fragments of segments in the TM but are also integrated into the system that presents suggestions for translators as they type (K. Flanagan, 2016). Planas and colleagues presented a system concerned with identifying and managing "knowledge-rich content", i.e. content from which it is possible to make conceptual analyses. They also discussed how a system that does these analyses might be integrated with translator's tools (Planas, Josselin-Leray, & Picton, 2016).

In a recent presentation, (Carmo, Trigo, & Maia, 2016), I proposed a view of the current state and of the evolution of translation tools for translators, suggesting that they could be renamed as KAT (Knowledge-Assisted Translation) tools. For this evolution to take place, current CAT technology needs to be integrated into systems that classify and explore the textual knowledge that is contained in all translated materials and references. The knowledge embedded in these resources could then be linked with technical and human resource management tools. Common Sense Advisory presented the most recent proposal for a term for this new wave of translation tools: "Augmented Translation" (DePalma & Lommel, 2017).

2.3.5. Studying the effects of translation tools

Ergonomics

In the 2016 workshop "Next Generation Translation Tools", held at the University of Swansea referred to above, there was another theme in the limelight: "ergonomics", with two panels dedicated to it. Ehrensberger-Dow and Massey presented a project that aimed at characterising ergonomic conditions in translators' workplaces (Ehrensberger-Dow & Massey, 2016).

Sharon O'Brien presented several projects from DCU that relate the use of translation tools to the area of "cognitive ergonomics", which seems to be the most accepted name for what was called "cognetics". These projects show a close connection to some of the discussions and methods employed by TPR, and study cognitive effort, users' perception of effort, user interface design, usability and organisational ergonomics (O'Brien, 2016). One of those projects suggested the contribution of disciplines like Human-Computer Interaction, Usability Studies and Personalisation to develop answers to problems identified as arising from "cognitive friction" (Teixeira & O'Brien, 2016). Ana Frankenberg-Garcia and Joanna Gough presented at the same conference a project that aims at studying how tools may be extensions of cognition processes in translators' minds (Frankenberg-Garcia & Gough, 2016). The phrase "From black-box to tool-box" describes this intention. One of the dimensions to be studied refers to how the increasing reliance on suggestions presented by the tools may reduce the capacity to generate ideas on the part of translators.

Effects of technology on texts

Other projects aimed at studying the effects of using translation tools on texts. One of these was TRACE, which ran from 2007 to 2010 at the Universidad Autónoma de Barcelona (http://tradumatica.uab.cat/trace/trace,es). A study that resulted from this project concluded that there were more marks of interference (one of the "translation universals") in translations made using CAT tools than in translations made with no such tools, especially due to the segmentation features of these tools (Martín-Mor & Sánchez-Gijón, 2013). In another study, the "translation universal" that was analysed was "explicitation", but the paper that was published describes only the pilot study that was done (Mesa-Lao, 2011).

Productivity and quality are two of the reasons why CAT tools are used – they are said to increase both, although some argue against the quality expectation, only admitting eventual gains regarding consistency. Bowker carried out a small experiment, in which she tested and confirmed the productivity increase, but she concluded that there is a need for specific training in order to avoid a decrease in quality, when translators trust the contents of TMs too much (Bowker, 2005). Yamada also challenged the productivity gain expectation, according to the hypothesis that this depended on the relatedness of the contents of the TM to the text that is being translated. By comparing the effects of a TM that was more literal than another, he identified more visible effects on productivity in fuzzy matches, since the editing operations are finer and have a stronger impact on relative speeds (Yamada, 2011). But the consistency hypothesis still required testing.

Joss Moorkens' PhD thesis is an extensive study on TM and inconsistency. One of the first conclusions of his study is that TMs are not guarantees of consistency and that sometimes this consistency is not even required. The contexts in which inconsistency is expected occur in specific conditions: when TMs are very large, favouring recall (bridging as many correspondences as possible) instead of precision (prioritising rigour over quantity), when there are too many translators feeding the TM under time pressure, when STs are not consistent themselves, or when the terminology used in ongoing projects changes over time. (Moorkens, 2012)

Jiménez-Crespo also discusses and challenges the expectations regarding the increase in quality of, in his case, localised Web texts. Starting from the notion that TMs operate at the microtextual level, he presents two opposing views: one that believes that this leads to higher consistency and quality, due to the research and reuse of similar previous translations, and another that underlines the negative impacts of uniting translations from different projects and translators, and of the tendency for translators to not use cohesive devices (such as anaphoric and cataphoric references), so as to promote future reuse of the translated segments in other contexts. He then makes a comparison of Web texts that are not translations with others that were translated. (In the latter case, he assumes that these must have been translated with CAT tools, as that is the norm for this type of content.) The levels of consistency he investigates are lexical (looking for reuse of terms from the SL), syntactic (focusing on the forms of address of the TL – Spanish) and typographic (looking at the capitalisation of words borrowed from English). He confirms his hypothesis that translated texts are more inconsistent than texts written in the TL, at all levels, but he calls the attention to the fact that this may not be due strictly to the use of TMs, but also to factors such as the intervention of several translators in the same translation project, or inefficient editing processes. He comments that further studies of these findings might be necessary, because this may be seen as a new "translation universal", also challenging basic assumptions of TS, as the individualistic view of the translator and the unitary nature of the text (Jiménez-Crespo, 2009).

Ekaterina Lapshinova-Koltunski (2013) has built comparable corpora to study several dimensions and factors behind language variation in translation. The translation production variables analysed in this study were: professional translation, use of CAT tools by translation students, a RBMT system and a SMT system. The text features analysed include lexical density, lexical richness, the presence of pronouns, and the distribution of parts-of-speech (PoS). For most of these factors, students using CAT tools present very different results from that of professional translators. In fact, most of the time, they are at opposite extremes of the range. For example, the type-token ratio and the lexical density are the lowest for students with CAT tools and the highest for professional translators (with both MT technologies between these two extremes). However, the study is not conclusive regarding the specific influence of the use of CAT tools on the texts produced. Another study on the same lines is "The influence of translation technologies on language production" (Fantinuoli, 2016). The author studied factors like sentence length, lexical density, pronoun distribution, and nominal/verbal ratio. He confirms the hypothesis that sentences in texts translated with CAT tools are shorter than in texts translated without CAT tools. And in texts that are later revised without CAT tools the sentence length increases. The same happens with the distribution of pronouns: texts produced with CAT tools use fewer pronouns, with translators preferring word repetition as a more explicit text strategy, due to the disappearance of co-text and context caused by segmentation. The same effect was attested by the presence of more nominal structures, but lexical density did not yield such clear results.

How translators perceive technology

In 2006, three important articles on the perceptions of translators in relation to technologies were published. These were all surveys: an extensive survey of professional translators on the adoption of CAT tools (Lagoudaki, 2006), an analysis of translators' comments on technology in a translators' forum (García, 2006) and one which included an online questionnaire (Dillon & Fraser, 2006). Lagoudaki's article has some interesting results, such as the realisation that some translators have invested in CATs but do not use them, because of the steep learning curve these imply. García also comments on the learning curve, highlighting the complexity of features and functions, the unreasonable expectations of first-time users, and how this relates to the technical support, or lack of it, offered by software vendors. Lagoudaki asked users for features they wanted to see in their tools, and some of the most frequent answers had to do with sub-segment suggestions and ways to organise and facilitate searches for answers and support in reference materials and on the web. Besides, a significant percentage of users would like to be involved in tool development, which is a surprising answer for a professional community that is usually seen as not very proactive. Dillon and Fraser's survey was not so extensive as Lagoudaki's, and its main results were that young translators and translators with computer knowledge are the ones that are most in favour of the use of CAT tools.

These three projects and a few more are presented and discussed in a review that collects the most relevant empirical studies focused on the use of translation technologies: "Translation-Memory (TM) Research: What Do We Know and How Do We Know It?" (Christensen & Schjoldager, 2010). One year later, the same authors published the result of a pilot test on students' perceptions of work with CAT tools (Christensen & Schjoldager, 2011). They explain that they chose to work with students because they are more sensitive to changes in the process since they have not incorporated into their habits all the features offered by CAT tools. In this study, students reported that TM technology makes them think less for themselves, trust the suggestions too much, and even feel a certain loss of control. This has an impact on the planning stage, which is reduced and even neglected, leaving aside any macrostrategic decisions which are only possible in contexts that allow for an adequate preparation time. Comprehension was also less thorough than usual, and when drafting, instead of making microstrategic decisions, students opted for accepting solutions from the TMs. At the last stage, apparently, students dedicated more time to revision, especially to revising the contents that they had applied from the TM.

Collaborative work on the web

In a paper that is only briefly referred to in the 2010 review mentioned above, Ignacio García (2007) comments on the extent to which a new way of using TMs in collaborative platforms might change the whole work environment of translators. He makes a vivid account of the long history of gain and loss of power by translators in the shadow of translation technology, and how this depends on changes to the business models that build upon this technology. He then talks about disempowering and deskilling of translators, due to the possibility that all results of the translator's work end up in a pool controlled by clients, and to which they have no access. In 2016, a team of researchers from DCU published a paper discussing this loss of power by translators, which seems to confirm this as a relevant consequence of technology development (Moorkens, Lewis, Reijers, Vanmassenhove, & Way, 2016).

The term García chooses for this era of collaborative work is "web-interactive translation". It is interesting to note that the earliest descriptions of translator support tools (like Kay's or Melby's in the 1980's – see section 2.3.1) were presented in collaborative setups. This is because these papers were written at a period that preceded the advent of personal computers, and in which users only interacted with terminals linked to central servers. So, these new collaborative environments might be seen as a return to the primitive form of technology for translation. However, individual translators working on their desktop still seem to be the most recognised translation production environment. This dissertation will take the perspective that desktop tools are adjusted to the needs of both the individual translators and the translation industry.

The price to pay for interactivity with web tools and communities is the loss of control by translators over the process and the products they work on. The loss of control is such that García even describes this as a form of MT: *"They do not 'author' translation; they literally drive an Example-Based Machine Translation Engine (cf.*

Carl and Way 2003), filling in whatever gaps are encountered." (García, 2007, p. 66). This idea had already appeared in the 2006 survey of the same author, referred to above: in a section in which the use of fuzzy match portions in Déjà Vu is discussed, this feature is presented as the incorporation of EBMT into TM. This view is also reflected by a user, who says that, when he uses these features, he feels like he is *"editing all the time instead of translating, but that can be a good thing."* (García, 2006, p. 103) But García cannot refrain from stressing that what for some users is a useful feature, for some may be a distracting option.

Effects of technology on translation practice

In a text called "Has computerization changed translation?" Brian Mossop comments on several dimensions of the changes that technologies brought to translation practice. He makes a clear statement concerning the support computers offer to translators, in terms of the tasks they perform:

> *Computers have not, so far, directly affected the central translation processes of interpreting the source text and composing a wording in the target language; in the main what they seem to have done is speed up the activities of editing and research.* (Mossop, 2006, p. 788).

He discusses the evolution of the relationship between translators and technology more as the effects of business strategies, and of changes to how work is organised, than as effects from technology itself. He stresses that one of these effects is the fact that the words of commissioners of translations are taken as ideological, rather than economically-bound. The paper includes a reflection on how researchers discuss the terms and conditions in which translation is performed, overestimating linguistic and technological factors, while neglecting professional factors such as chunking and deadlines. As already explained, there has indeed been much research on chunking and segmentation, but not on deadlines, lack of quality of the source texts, the cost of technology, and other factors that mark the days and lives of professional translators.

As an important product of chunking, Mossop refers to the advent of "collage translation", a practice that is present in the everyday preoccupations of translators. This refers to composing one single translation product from different projects, translated by different people at different moments in time. "Research on translation tools" (García, 2009) is an overview of this area of TS, in which the author shows concerns related to some of the issues Mossop comments on. García suggests the term "hive translation" (scattering a text among many translators, or bilinguals, who produce a large amount of translated text in a short time) as a complement to "collage translation". The author also refers to "translation as utility" (in which translation is seen as a service that is always available, at the reach of a click in an online on-demand system) as another mode of work that has been enabled by technological advances.

Matthieu LeBlanc conducted an ethnographic study in which he collected data from interviews and the direct observation of professional translators at their workplace in translation agencies (LeBlanc, 2013). In his report, he presents the advantages and disadvantages of TMs. It is important to set apart the contexts of use, since most of the critical views are based on the fact that, in some companies, translators do not have the freedom to edit the contents in the TM – this creates a feeling of frustration and reduces their willingness and even ability to question and decide for themselves. Another result is that they start feeling that they retain less information since they rely extensively on the TM. The author also recalls that Pym describes CAT tools as a return to a type of equivalence that existed in the 1960's, but which is now a sort of "artificial equivalence", since it does not correspond to equivalence at the word or sentence level, but at the segment level, when the notion of segment was created by CAT tools (Pym, 2010, pp. 131–134).

The paradigmatic and syntagmatic axes

Another perspective offered by Pym in this text concerns the effect on the linearity of reading and meaning construction, which sustains the writing phase.

The way these tools embed suggestions means that the syntagmatic paradigm of writing (constructing sense in a linear way) is broken by the paradigmatic axis, when for each unit other alternatives are presented. This effect may be felt while a translator writes his translation (in interactive mode, for example) and he keeps receiving sub-segment, or predictive writing suggestions. This has an impact on the text he produces, which may lose cohesion (syntagmatic effect), in favour of consistency (built from always choosing the same options from the paradigmatic axis).

Pym discusses how technological advances defy the established powers and favour new generations, and how this is reflected in the challenges that web and collaborative platforms have presented to the whole professional environment. He finalises with a positive view on the evolution of the technological context, by saying that the openness that characterises this era may imply more dialogue and less individualistic views, which is appropriate for a profession that is founded on communication.

Cognitive effects of translation technology

In 2011, Anthony Pym, published an article with the title: “What technology does to translating”. In it, he proposes a discussion on the effects of technology on cognitive activity, social relations and the professional status of translators (Pym, 2011b). Pym argues that the cognitive dimension that is mostly affected by technology is memory because technologies externalise the sources of information and knowledge. He divides the decision process into three stages: recognising the existence of the problem, generating alternatives, and selecting the solution. The external solutions interfere with the central stage, as they show the alternatives to the translator. Moreover, he stresses that, at this stage, they may turn out to be complicating factors for the decision-making, even to the point of cutting out intuition.

Christensen (2011) also emphasises the mental effects of TM on memory. She says that a TM is like a “supplementary long-term memory”, and when translators share a TM, it is said to be an “instance of collective distributed cognition”. She proceeds:

> *Being an instance of distributed cognition, TM-assisted translation is assumed to interfere with and control translators’ mental processes. This is mainly due to the fact that a TM automatically provides the translator with translation proposals, which the translator is forced to consider before moving on. In fact, a TM can be said to offer solutions also when the translator is not aware of facing a translation problem. Having to consider and assess the retrieved matches, translators may indeed spend more time thinking about and revising previous translations than they do translating from scratch (García 2010).* (Christensen, 2011)

For the rest of her paper, she presents a review of several studies on these mental processes. One of the studies discussed by Christensen is by Tânia Liparini Campos.

In her PhD dissertation, Liparini Campos analyses thoroughly the variation in the way translators look for answers to their translation problems, according to the tools they use. Her study shows that professional translators do not go through a separate phase for “orientation” – they read the ST as they start translating. She also reveals that during drafting most of the pauses are related more to orientation than to revision, and that revision is most frequently performed as a separate stage. She uses the term “internal support” for the recourse to their own and the tools’ resources to solve translation problems, and “external support” for the use of external sources of reference. (This terminology is also used by the PACTE group in the papers analysed above, in section 2.2.3.) Her study shows how professional translators optimise looking for answers when they use CAT tools, which can be observed as a reduction in the time for revision since this is incorporated into the orientation and drafting stages. As in the study of the PACTE group, the instrumental and strategic competences are highlighted by the study. The other factor that was analysed in this study was time pressure, but she concludes that this only affected the revision phase, during which translators avoid resorting to external support (Liparini Campos, 2010).

The importance of research in translation

Researching in internal sources includes using the concordance feature to look for words in different textual contexts in the translations stored in the TM. In fact, actively using concordance search features is a fundamental skill used by experienced translators to take advantage of all the contents of their TMs. So, it is surprising that there are so few papers investigating the use of this feature. García (2014) may give a good reason for this when he observes that concordancers were developed for computational linguists, not translators.

However, research is a central concern for translators, so any tool related to research should play a central role in ATS. Nirenburg’s early plan for a “machine-aided” translation tool (1992), for example, aimed at a tool that served both researchers and translators. The exception to this is “Keeping an Eye on the UI Design of Translation Memory: How do Translators use the Concordance Feature?” (O’Brien, O’Hagan, & Flanagan, 2010), a paper which discusses how translators use the concordance features as an alternative and “manual” way to look for sub-segmental suggestions which tools do not present. The authors queried the participants in the study about the usefulness of integrating such suggestions as predictive writing supports. The conclusions of the study, although requiring further confirmation, are that these features are seen as very useful, even if they demand good implementation to be usable and avoid being a nuisance.

After this survey of different points of view and approaches that TS have offered over the years, first in terms of the principal terms of the discipline, then in analysing the TP (within DTS), and then looking at the tools used by translators (within ATS), the next sections present conclusions and discuss the grounding notions of translation and revision, as the main processes that translators are involved in.

2.4. What is "Translation" in this context?

The purpose of this section is not to present a global definition of "Translation", in its full dimension. This dissertation only focuses on the technical dimension of the translation process, analysing its evolution in the technological context that has determined so much of it in the last decades.

In an industrial environment in which translators use CAT tools for all their work, "translation", "revision" and "post-editing" are usually seen both as specific isolated processes, and as parts of the main production processes. So, it is often difficult to tell them apart. This dissertation, while admitting that they may appear together in the same workflow, analyses them as separate processes, composed of several phases, and including different tasks and subtasks.

Before analysing each of the phases of the TP, it should be stressed that the term "translation" is used as relating to a global technical process composed of phases, tasks, and subtasks performed by one person alone. The word "translating" will be used to refer to a specific task that means "translating a sentence from scratch", as opposed to editing only parts of a translation suggestion.

The term "revision" is also used to describe a whole process, performed by a person different from the one who created the translation. In order to distinguish clearly the full revision process from the last phase of the TP, the third phase of the TP will always be called "self-revision".

The "phases" in the table below are not sequential steps, and they may overlap in time. Each phase implies different tasks that are performed over different textual elements, or sources, but these are also not exclusive of each phase. As we will see, a good identification of the elements over which each task is performed is fundamental to distinguish "translation", "revision" and "post-editing", but this may not be enough to define them. Tasks are performed in CAT tools over specific textual elements (such as

ST or TT segments, TM suggestions, or, in PE, MT hypotheses) by specific subtasks. These may be broken further into "actions".

The four editing actions

Toury's "omissions, additions, changes of location and manipulations of segmentation" (section 2.1.1) are indications of "matricial norms", but Nida considers "additions, subtractions, and alterations" (section 2.1.4) to be "techniques of adjustment". Pym refers to "actions" being *what we actually observe translators doing (e.g. typing, correcting typographical mistakes, looking up terms in glossaries, etc.),* (section 2.1.5). And, as described in section 2.2.3, Angelone classifies the "editing activities" of "additions, deletions and revisions" as "production-behaviours". Finally, Alves et al. (section 2.2.5) use the term "production segments", which includes "revisions, deletions, substitutions, etc." So, it seems that it is difficult to find a consensual classification of the micro-procedures translators carry out when they are producing their translations.

In the coming sections, I claim that deleting, inserting, moving and replacing are the defining actions of "editing", and these will be referred to as "editing actions", or "edits" – observable operations performed by the translator over pre-existing text. The meaning of "editing" will be discussed in section 2.5 below.

The main concepts that describe translation in this dissertation are presented in a table below, for systematisation purposes. It is important to note that it focuses on the external manifestations of the TP alone, leaving aside anything related to the decision process that goes on inside the translator's mind. The contents of this table will be described in each of the next sub-sections, one for each phase of the TP.

Phase	**Window**	**Textual element**	**Task**	**Subtasks**
Orientation	Source	Source text	Reading	Reading
	Target	TM target text (full and fuzzy matches)		
		MT target text (no match)		
	References	Terms/Words	Researching	Researching

Drafting	Source	Source text	Reading	Reading
	Target	Empty window	Writing	Typing
		Source text		Overwriting
		TM target text		Editing
Self-revision	Source	Source text	Reading	Reading
	Target	Own target text	Checking	Validating / Editing
Result of the process		**Translated text**		

Table 1 – The Translation Process – a simplified view

This table shows that the Orientation phase only includes two tasks: reading and researching. Neither of these has subtasks.

Drafting includes reading and writing, one in the source window, and the other in the target window. According to the type of textual element in the target window, writing is divided into three subtasks: typing, overwriting and editing. These subtasks may be performed cumulatively in the same text, but not in the same segment.

Finally, Self-revision is composed of reading and checking, checking being subdivided into validating, when the segments are approved without edits, and editing, when changes have to be made.

2.4.1. Orientation: Reading and researching

Before the actual TP, a translator has collected, learnt, processed or organised, both conscious and unconsciously, vast amounts of linguistic data, in monolingual and bilingual form, and he has tested, in different contexts, strategies, techniques, macro and micro procedures to transform information written in one language into the same, or as similar as possible, information written in a different language. At each translation assignment, before or during the TP, a translator reads the text in the SL and makes some initial decisions, saving them in a sort of buffer memory.

While translating, a translator shuffles back and forth through all the information he has collected from the ST, mixes it with previous monolingual and bilingual knowledge and the translation experience acquired, to feed a process which requires the best information transfer decisions. Each piece of information, whether at a micro level (such as phonemes, rhyme and rhythm, words, lexical choice and collocation), or at the macro-level (such as paragraph coherence, text cohesion or tone) is measured and considered, in view of the explicit or intuitive decisions the translator deems appropriate for each meaning transfer challenge.

At this phase of Orientation, the amount of information required for the translation assignment will determine the rhythm of work, or the cognitive flow, that allows the translator to move from the ST to the TT without major pauses. However, when there is a cognitive challenge, translators need to stop and research, whether it is in their internal information or memory, or in the information immediately provided by the translation tool, or in external resources. This process was described in some detail in section 2.2 above, in several studies that essentially collected data from eye-tracking activities, to understand how translators manage information retrieved from these sources and to identify traces of the cognitive effort revealed.

Carmo, Trigo and Maia (2016) presented a model of the TP that starts by connecting this early stage to a knowledge management system that establishes connections between texts and human resources, using technical domain (areas of expertise) as the factor that determines the links. The next level of the knowledge management system may set up further connections around the identification of textual features in external and web resources used for reference in the research stage. This method could also be used to support the building of specialised vertical web search engines, integrated into the translation editor so as to respond to each research action from the translator, whether they are internal to the editing tool (as in concordance searches) or external (such as those that are made in generic web search engines). Some of the current CAT tools already have some form of web search from within the editor, but these are basically “favourites” or “bookmark” lists.

This proposal is similar to Séguinot’s description of translation as information and decision management (section 2.2.6), and Austermühl’s view of translation as the production of a TT based on more sources of information than the ST. Enríquez Raído’s (2014) book is a thorough study of the importance of web searches for the TP, and Muñoz Miquel (2015) studied the influence of communication with specialists as sources of reference.

We have seen above (section 2.3) that concordance searches are vital instruments to complement the results presented by CAT tools. The extensions to these features in the newest generation of CAT tools, either through predictive writing suggestions or lists of sub-segment alignments, are fundamental breakthroughs in supporting the orientation phase of work, particularly since this phase does not finish when the drafting stage begins. Quite on the contrary, concordance searches, together with terminology searches, web browsing and all other research and orientation tasks accompany the full TP. There is room for more research in this domain, which might explore and integrate other features, like those to support managing translation comments and translation queries, to take advantage of image identification technologies, and also to leverage information from social networks when these are used as information support tools.

Translation is, viewed from this perspective, essentially an information management process, which professional translators perform with a high degree of efficiency, so as to be able to give the required answers to an ongoing decision-making process that determines the words that they write.

2.4.2. Drafting: Writing over different textual elements

When the translator starts writing the translation, he is mostly and ultimately responsible for the text that will be read by the target audience. So, the focus of this stage is mainly on the writing rules and norms of the TL. The ST, as said before, is simply a source for what the translator writes. And until he is happy with the results obtained, he will draft, edit and rewrite the text, in cycles of sometimes seemingly never-ending decisions. This complex process of decision-making brings with it a natural consequence: most of the decisions of the TP do not survive in actual form in the final text. A clear consequence of this is that, if you only have the end result to work from, you will hardly ever fully understand how the ST became the TT. The notion of “macro-units” proposed by Alves and Vale (see section 2.2.4) is a solution for this problem, although it still requires a fair amount of testing.

Interactive and pre-translation modes

Section 2.1.4 comments on the two main writing modes: interactive and pre-translation. Interactive mode, when translators type their translation from scratch, is supported by the presentation of suggestions at the segment level, as the translator moves along the text. This mode was first used to describe the “interactivity” that existed between Microsoft Word and Trados Translator’s Workbench. In today’s CAT environments, most of which present a tabular view, with the ST side-by-side with the TT, it is not so common to see this mode of working.

For translators that prefer to work with interactive mode, with a clean target window and to type the whole text of their translations, CATs offer full and fuzzy matches to be copied onto the translation, searches in terminological databases and concordance features as productivity enhancements, in separate windows or panes, and translators need to copy or retrieve them from those panes using keyboard commands or mouse actions. Second-generation features, like fuzzy match repair and predictive writing, are very welcome supports in such a working environment. Predictive writing seems to be the most useful and usable of these features, since it reduces typing effort, as it fills in the rest of the word when the translator starts typing it. However, it needs to be sensitive to the textual surroundings of the words the translators type, at the risk of becoming an intrusive feature, instead of an aid.

The pre-translation mode is the standard way of working in a CAT tool because it is the most productive: the translator has the translation window prefilled with either TM matches or ST sentences (when there are no suggestions above the fuzzy match threshold). In this setting, the translator only needs to "edit" these suggestions, overwriting and replacing the words that will not survive in the target version. Christophe Declercq presents an example in which more than a third of the words in the ST extract are "named entities" which are transferred into the TT (Declercq, 2014, p. 481).

When a translator types over the ST words, instead of writing all of the translation, he will not only be more productive, but he will reduce the risks of typing mistakes. Named entities and numbers are clear examples of elements that fully justify overwriting the ST, instead of typing everything from scratch. These and similar non-translatable elements are appropriately often called "placeables" – the translator only needs to identify their position and make sure they are correctly placed. Still, it is important to stress that, even when there is a high percentage of elements that are imported from the ST to the TT, this is considered as translating. In such cases, the translation strategy chosen by the translator was to import elements from the ST, but no translator can exempt himself from the responsibility behind such a decision.

Three ways of writing

It is during the drafting phase that the translator is actually "translating", in the sense that he is doing and producing something by means of writing. In terms of technical procedures, this translating may be broken into three different subtasks: typing

TL sentences from scratch, overwriting SL sentences and/or editing TL sentences. So, the textual elements these subtasks are applied to, and the tools and techniques employed do not define the whole process, but only technical parts of it. In an industrial workflow, projects combine all these tasks, making it harder to clearly identify their borders.

The three different terms used here to describe these writing subtasks (typing, overwriting and editing) are also useful indicators of the type of support that translation editing tools should offer for different ways of working. Nevertheless, one must admit that in any working mode, writing is seldom a linear activity, composed of only typing, only overwriting, or only editing. Furthermore, this phase is frequently intertwined with the next, self-revision. The recognition of the complexity of professional and industry workflows make this description necessarily simplified. Finally, it is important to stress that at the end of the drafting phase, the text still cannot be considered a translation, but only a draft of the final result of the TP.

Editing in CAT

When he edits textual content, the translator performs sequences of actions over isolated words or sequences of words. Translators that process thousands of words in fuzzy matches every day perform these actions until they are not aware of the repetitive nature of their work. But they do not do more than:

- Deleting
- Inserting
- Moving
- Replacing.

In this dissertation, this set of actions is referred to as “editing”, in line with several authors, such as (Guerberof Arenas, Depraetere, & O’Brien, 2012) and (Declercq, 2014). The four actions described above are referred to in this paper mostly as the “editing actions”, but also simply as “edits”.

As described in the table above, a translator edits segments during the drafting phase, which relates to the editing of fuzzy matches or even full matches from the TM, and at the self-revision phase, when he checks the full translation he drafted.

It is assumed in this dissertation that most translators take advantage of all the support that CAT tools offer them, working with all the available features enabled, although, admittedly, these may be used only on an on-demand basis. The profusion of features, windows that surround the editor, popups with suggestions and windows to look for results of searches, are such that translators may feel confused and let their decision process lag behind. In such situations, interactive features such as predictive writing may present the sort of interruption that translators prefer to disable.

2.4.3. Self-revision: Reading, validating and editing

"Self-revision" is the term used by Mossop to differentiate the final phase of TP from "other-revision" (Mossop, 2014). For the ISO 17100 standard, this self-revision phase is called a "check" (ISO, 2015, p. 10).

At this final phase of the TP, the translator goes back to the beginning of the translation and reads the full TT, checking it against the ST, and following any eventual notes and comments left along the way. The translator may need to do some research again, but rarely repeats the searches done at the orientation or drafting phases. Nevertheless, he may go back to the start of the text, and reverse translation decisions, based on information collected from the whole text. The type of writing the translator does at this stage is basically editing, as made explicit in the previous section: he chooses scattered points in the text, and deletes a word here, moves other words there, replaces other words at a different point.

Several authors talk about and study different styles of self-revision. Carl, Dragsted and Jakobsen (Carl et al., 2011) suggest that there are three types of translator, according to their behaviours in revision: online revisers check their translation as they move along the draft, end revisers do the revision in one single go, once they finish writing the translation, and constant revisers revise while translating and again at a final phase. Liparini Campos (2010) has also studied this phase and shows how time pressure makes translators invest more in online revision (during the drafting stage) and less in end revision. Professional translators adapt to time pressure using supports that imply shorter time, like those provided by TMs.

Technical support to self-revision

CAT tools do not offer specific support features to revisers, whether they are revising their own or another translator's work. In fact, revisers work with virtually the same interfaces as translators, sometimes the only difference being the features they use more often.

The features offered by CAT tools that revisers use more often will be described in detail in section 2.5.3, but a quick reference to QA (Quality Assurance) is required, because these are often seen as specific for revision. These tools focus on specific details, like the accuracy in numbers and punctuation, terminology consistency or the repetition of words. However, they do not support word and phrase consistency checks, which are the most frequent tasks performed during revision.

Together with final spelling and grammar, a QA check is usually performed as the last task in the self-revision phase. Although time pressure often forces translators to limit their revision to applying these three checks, QA should not be confused with revision, because this implies a full reading of the text.

Moving away from the ST

There is another important effect of the use of these tools that needs to be mentioned. Especially while doing self-revision, it becomes clear for translators that some of their decisions are bound to factors that do not depend directly on the source segment.

Translators do not respect the source segment when the ST has problems. Consistency errors and contextual errors are quite common – when translators have access to figures or lists of parts in technical translation, these are quite simple to spot and to correct in the translation.

Another instance of not respecting the source is when segmentation breaks the alignment between two parts of the same segment. An example of such a situation is when a two-word unit is segmented into two one-word segments (for example, for inclusion in a small box in a chart or a presentation). If the correct order of the TL implies that those two words must be swapped, in order to produce a correctly ordered TT, the alignment of those two segments is inverted – the second target segment contains the translation of the first source segment, and vice versa. Besides, we have also seen above (section 2.3.5) that inconsistency is a given in translations produced in CAT tool environments, for multiple reasons, like lack of access and discussion time between translation teams, or the evolution of the preferred terminology in an ongoing project.

Finally, in the post-production stages or in transcreation projects, translators and revisers may make final decisions for style, typographical preference, visual or communicative impact, which make the ST a remote reference, with virtually no matching relationship at the word or any other level with the contents of the target segment.

In spite of this lack of alignment between ST and TT in TMs, CAT tools favour recall, which means that it is assumed that anything validated by the translator is useful in a TM. However, translators and revisers know that TMs contain several segments that, for one reason or another, cannot be reused in contexts other than the specific context in which they were used. Often, translators request that an easy "opt-out" button exists in CAT tools to mark segments that, although they are correct in the sense that they will help create a correct TT, they should not go to the TM, because they will reduce the quality of that content if it is reused on the simple assumption that everything in a TM is high-quality.

The result of these three phases of the TP is a translated text. For the ISO standard, a comprehensive revision, by a second translator, is always in order, after the full TP performed by one or more translators.

2.5. What is "Revision"? What is "Editing"?

If the terminology of the TP requires some clarification, in relation to the Revision Process (RP), things are even more complex. The main work on the RP is Brian Mossop's "Revising and Editing for Translators", a book that has seen three editions since 2001. He begins by presenting his definition of the two terms that give title to his book:

> *In this book, **revising** means reading a translation in order to spot problematic passages, and making any needed corrections or improvements. **Editing** is this same task applied to texts which are not translations.*
>
> *Revising and editing are first and foremost exercises in very careful reading.* (Mossop, 2014, p. 1).

Perhaps the two biggest challenges of revision are presented in the first sentence: first, how to identify what is needed, and second how to set the limit between correcting and improving. The next sentence includes a questionable use of the term "editing". This definition does not seem to be useful in TS since it concerns non-translations, and it does not help to understand most of its occurrences in the literature of TS. Furthermore, these concepts do not describe the reality of translation in a CAT context: *"Translators who use a Translation Memory thus need to develop a* ***reviser/editor*** *mentality rather than the mentality of a text composer."* (Mossop, 2014, p. 3) So, apparently, when matches from the TM are found, the translator becomes a reviser, or even an editor, although he is working over a text in a SL. Mossop discusses several dimensions of revision that would be worth further analysis, like the notions of "quality" and the revelations of lack of quality, i.e. "errors". He also discusses the limit at which revision becomes "retranslating", and mentions "checks" and "checking", which are to be seen as operations related to reading in order to identify problems.

2.5.1. Main terms in revision

For clarity reasons, the term "revision" will only be used for a specific process (the RP) performed by a person different from the translator, with the purpose of correcting or improving the translator's work. As stated above, when the translator revises and corrects his own work, the term "self-revision" is used. Besides, the terms "review", "proofreading" or "editing" are not used here to mean the same as "revision", although all these terms are used interchangeably throughout the literature on translation and in communication materials from the translation industry. In line with what is considered by the ISO17100:2015 standard (ISO, 2015), in this dissertation, "review" is an evaluative reading of a monolingual text, and "proofreading" is a final check to make sure all amendments suggested by revisers or reviewers have been implemented, before the final copies of printed materials are produced.

As for the term "editing", as mentioned above, it is used to describe a subtask of writing that translators perform over a TL suggestion, or a SL segment, which only requires a few changes to be ready for validation. It was used above to describe the actions carried out to update the translation of a fuzzy match from the TM, and in the rest of the dissertation, it will also be used to describe the actions performed during PE. To clarify this further, this dissertation does not follow a common set of terms used by the localisation industry, mostly in the United States, where translation assignments are composed of three steps, described by the acronym TEP – Translation, Editing and

Proofreading. In TEP, "editing" is a synonym for "revision", as this term is used in this dissertation.

The use of the gerund form of the verb (editing) as a noun corresponds to a process of grammatical metaphor, as described in section 2.2.5, and it highlights the meaning of the word as an action, instead of a result. Contrary to a translation process, which results in a translation, "editing" only exists while it is being performed. The result of the "editing actions" are not "editions": if we edit a translation, at the end of the process, we still have a translation. However, as this grammatical mechanism is not so promptly available in Latin languages, one does find in the literature references to "post-edition".

Editing is closely linked with the concept of "checking", which is the term adopted here to refer to the task of either validating or editing a segment. Editing is seen as being composed of four actions: deletion, insertion, replacement and movement.

2.5.2. Revision as mainly reading

It is important to highlight that, contrary to a translator, who can be said to be translating only when he writes, a reviser is mostly revising when he is reading. Mossop's definition cited above says so: *"... revising means reading a translation..."*, one of the chapters in his book is called "Revision: a reading task", and he stresses on other occasions that reading is the main activity performed by revisers. In fact, for most of his work, the reviser intends to simply read and validate the work done by a good translator.

Thus, the first main difference between the TP and the RP lies in the fact that there is no "drafting": the reviser can only work over a text that has already been completely translated. Even if drafting and self-revision have not been completed satisfactorily (according to different criteria) by the translator, the reviser works on a result of a full TP. So, it is assumed that writing will be performed less frequently than the other tasks: the reviser is not supposed to draft, type or overwrite, at least not as much as he will edit (insert small changes scattered through the text).

The most common view on the RP also assumes that there is no orientation, as the reviser follows mostly a good quality TT. This view could be represented by a table similar to the one that described the TP, but one that only has one phase, leaving out the "orientation" and the "drafting" phases. So, it could look like this:

Phase	Window	Textual element	Task	Subtasks
Revision	Source	Source text	Reading	Reading
	Target	Target text (another author)	Checking	Validating / Editing
Result of the process		**Revised text**		

Table 2 – The Revision Process (simplified)

However, when a reviser finds an issue, there is a pause in the cognitive flow of reading the ST and the TT, because he has to make a decision on whether to validate the translation and keep on reading, or if he should stop and amend something. From that moment on, the reviser reads not only the ST and the TT, but he also needs to research in references, to confirm the translator's decision, or to replace it with his own. So, one must consider that there is also an Orientation phase in the RP. The revised descriptive table of the RP should then be:

Phase	Window	Textual element	Task	Subtasks
Orientation	Source	Source text	Reading	Reading
	Reference	Terms/Words	Researching	Researching
Revision	Target	Target text (another author)	Revising	Validating / Editing
		Target text (another author)	Checking	Validating / Editing
Result of the process		**Revised text**		

Table 3 – The Revision Process revised

This representation is clearly a simplification. The Orientation phase, in this case, is not necessarily an initial stage, as reading and researching mix with the validating and editing subtasks. Besides, the research task is not discriminated, and different subtasks may be included here, as terminology, web and concordance searches.

In the table, revision, as a phase, appears as decomposed into two tasks: revising and checking. These may be compacted into one, but in more rigorous contexts, they may be required separately in the same project. During the revising stage, the reviser checks all segments exhaustively. He may even be required to change the status of all segments, to make sure that all segments have been either edited or validated. Then, he may need to do a second pass, with spot-on corrections, or by following the instructions of a QA process.

Different descriptions of phases in revision

In a presentation in 2014 at a conference in Copenhagen of a study on the cognitive process of revision, Huang and Minocha identified three phases in the RP, which they called “orientation”, “revision” and “final check”. They then associate these phases to three types of behaviours, concerning the proportion of reading and writing tasks between the ST and TT. In the experimental part of their study, they identified revisers who revise in a full three-phase process, others who only process two phases (some dedicating more time to the first phase–orientation–and leaving aside the final check, while others do not devote time to orientation and dedicate more time to revision and the final check) and, finally, some who process all work in one phase, revision, with no orientation or final check. (Huang & Minocha, 2014)

Edina Robin also includes an “orientation” phase in her description of the ideal RP, which is composed of five separate and sequential steps: orientation, reading, correction, presentational adjustments and language check (Robin, 2016, p. 53). Although this level of detail is useful for specific analyses, most of these steps are specifications of the main ones. Another approach might be to consider that both reading and researching were part of the orientation phase, and that the revision phase only happens when the reviser edits. However, this a very reductionist approach.

In this dissertation, researching is added to the usual tasks of reading and checking. Checking is not considered as a separate phase, but a task that is performed as part of either self-revision, or revision. It needs to be admitted, however, that in specific projects, the QA stage involves such a degree of complexity that it could be classified as a specific phase.

Revision and retranslation

Frank Austermühl’s chapter title *Revising is the new translating* (2013, p. 331) surely means that nowadays translators perform more often the role of revisers, but we may read it as if he meant that revisers often have to do a new translation from the TT they receive. Robin’s article quoted in the previous paragraph is called “The translator as reviser”. But can the opposite be said – “The reviser as translator”? Can the reviser be seen as acting in the place of the translator? For clarity purposes, when a reviser needs to replace the translation of a whole sentence, he is not translating, but "retranslating" – whatever he does, it is only possible because there is a version of a translation, even a bad one.

This is the second fundamental difference between the TP and the RP: a TP contains a full revision phase (of the translator's own work), but an RP does not contain a translation phase.

One must admit, however, that, exceptionally, a revision assignment may contain so many translating tasks as to become a "translation job". In such a situation, the reviser feels forced to either refuse the assignment, or to ask for it to be reclassified as a proper translation job. This is the typical situation when PE of MT content is disguised as a revision work, and translators do not accept the job in these circumstances.

The reviser never performs the whole 3-phase translation process, composed of orientation, drafting and self-revision. He may translate a full sentence, overwriting the translator's decision, but translating, instead of editing, cannot be a regular task in a revision process. So, adding a drafting phase to the RP does not seem to be required.

2.5.3. Tools for revision

Mossop's book has a specific chapter on computer tools: Chapter 8 – "Computer aids for checking" (Mossop, 2014), with a section for "Editing functions of word processors", where he includes spell and grammar checkers, "find and replace" features, marking changes and edits, inserting comments and comparing documents. In the final (brief) section of this chapter, "Tools specific for revision", he admits that none of these tools were developed or are specifically designed for revision work. He justifies this by saying that the main task of the reviser (identifying mistranslations by reading) cannot be automated. Mossop closes this chapter with reference to QA tools.

The support that CAT tools offer to revision is based on filtering and extending search features, like concordances and regular expressions. This is appropriate, since one of the concerns during revision and self-revision is consistency. However, it is not adequate, as one may see in a more detailed description of the process.

A recent video showing the features available for revision work in SDL Trados Studio shows that, when users set up the application for the role of "reviewer", the main changes are that the TM results pane is placed in a different position, "track changes" is automatically enabled, and the "Review" tab comes up front. This tab includes buttons related to sharing versions of the translation outside the application, comments, track changes, content filters, besides spelling and QA checks. Most of the procedures described in this video refer to using advanced filter and search actions, but other common revision tasks, like dealing with translator comments and doing final checks, or spelling and QA checks are not mentioned (SDL Trados, 2017). An analysis of other CAT tools will confirm that the approach to revision is very similar to this. memoQ 2015, for example, has the same type of features in the "Review" tab. It is important to note that spelling and grammar checks are not embedded in the CAT tool, and usually depend on word processors, like Microsoft Office, which have more advanced features.

During revision, segment statuses are changed, to reveal that they were processed at a different stage (Declercq, 2014, p. 484). At the end of a translation, all segments have the status "translated". After revision, each segment that was edited becomes "drafted". Before delivery, the translator must either revert these edited segments to the "translated" state (as is common when self-revising) or mark all the segments as "revised", which is usually done by the reviser only.

Quality Assurance (QA) features are included in CAT tools as additional guarantees that repetitive details and specific requirements are respected by the translation. These features correspond to the pre-delivery checks that Project Managers usually make after they receive a translation: checking that all numbers and punctuation marks are correct, that there are no missed translations, that terminological consistency was assured, and several other aspects. In fact, Project Managers used these tools before they were included in CAT tools, mainly as a quality control instrument. It was only when they were integrated into CAT tools that they became proper QA tools.

These features are sometimes complicated to set up, and they seem to favour recall (catching as many errors as possible) over precision, which causes a long list of false positives that may delay the reviser's work a lot. Among these false positives, one may find double spaces that were inserted in the source as indentation techniques, segments left untranslated because they only contain elements that are imported into the TT, terminological entries which are deemed incorrect just because they appear in inflected forms, and many others which are considered incorrect in most contexts, but which are correct in others. The work of the translator, when faced with these false positives, is to confirm that he checked the context, and to validate the decision of keeping them. So, rarely do translation assignments end with a "no-error" report, but instead they include a list of issues which had been pinpointed by QA tools but which are kept in the TT with the acknowledgement of the translator.

It is with these tools that the reviser must do his best to identify and eliminate errors and obvious mistakes. He must read attentively, although time pressure makes him do it at an increasing speed, as the deadlines approach. He must check repetitive reports of details, among false positives that request that he revisits different contexts and references. And, in this process, he must make strategic decisions on which are the most important details to which devote more time, like terminological choices, to guarantee that at the end he delivers a high-quality and consistent translation. But, inside the translation window, revisers see the same contents as if they were translating from scratch. For example, fuzzy matches still appear with highlights only in the updated words on the ST side, and do not show on the TT side the edits that the translator may have already made to the sentence he received from the TM.

One of the effects of this repetition of editing environments may be the loss of attention to some of the details, and a tendency to rewrite, instead of looking attentively for errors. This effect may also result from distracting features, like "track changes", which are not so much an aid as a control marker. When using this feature, it is very common for new mistakes to be introduced, like double or missing spaces and words, because the presence of word formatting, like strike-throughs, underlines and colours, in otherwise complex interfaces, adds up the distraction.

When translators self-revise, they scan through the whole translation hoping that everything is correct, either not even opening the TT segments they created, or validating them almost instantaneously. They only stop over the segments that require some type of editing, whether it is a correction or an improvement. So, any distracting and repetitive visual indications (such as false spelling mistakes that are simply words in a different language) may be disregarded as false alarms or annoying interruptions, at a time when pressure is at its highest.

Revisers working on other translator's texts are also influenced by this technological context, which may condition them to opening all segments to change their state, and so they may end up repeating everything the translator has done, rechecking all sentences and words against the TM and other references, instead of looking for the most critical errors. Mossop cautions against the tendency of revisers doing everything the translator has done, highlighting research as the task whose duplication should mostly be avoided: *"It's a good idea to have new translators write* *marginal notes indicating their reference sources, so that you do not end up repeating their research."* (Mossop, 2014, p. 194). Daniel Gouadec describes the difficulties of the relationship between translators and revisers, professionals who have to deal with each other's different opinions in an environment of desirable complementarity, but which is not always possible (Gouadec, 2007).

A new layer of text between the ST and the TT

The reviser does not approach the relationship between ST and TT as a direct relationship between two texts. Between them, is the text that the translator worked on, which included TL segments coming from the TM. To access this intermediate text, the reviser researches in a TM that was not updated with content from the translator. CAT tools may signal different translations from that in the TM, but usually this is only automatically available for full matches. So, this research is often done regularly with concordance searches for fuzzy matches.

An alternative way for the reviser to visualise the relationship between the unedited TM text and the result of the work of the translator is to activate track changes visualisation in the translation editor. However, these views do not show the complete unedited sentence and the complete translated sentence. Instead, they emphasise the overlap between the two versions and concatenate the differences. This form of visualisation hinders the reading capacity of the reviser, leading to errors that are very frequent when these instruments are used, in the concatenation points between the two versions: missing, repeated or truncated words, double or missing spaces, among others.

Tools that correctly showed this relationship between three texts (the ST, the TM unedited version, and the translator's version) would be fundamental to properly support the specificity of the reviser's work.

Revision and quality assurance

The duplication of the work of the translator, especially his self-revision, but also his researches and checks, is an expensive redundancy, for clients and for the industry. So, when time and cost pressure is too high, companies may try to reduce any of these tasks and processes to the ones that are considered priorities. Still, the opposite view is the one most defended by the industry, favouring the "four-eyes principle" (notably, it refers to "four eyes", not "four hands") as a guarantee of the quality of a product and a process that some find difficult to grasp.

The pressure on obtaining a quality product is so big that there is even the notion that a third process is required after revision. This third process may be known by different names, and it may fulfil different objectives: proofreading (in the TEP process), sign-off (to make sure the process of converting the contents of the CAT tool to other formats, or even printed form, do not insert or bring to light any mistakes), review (to verify the appropriateness of the translation for the audience), quality assurance (QA), when small details are extensively rechecked pre-delivery, or other aspects. This pressure on quality is also behind a new stage currently making its way into the routines of translation providers: assessment of the levels of quality of translations, based on standard classifications of errors. These will be discussed in the context of the evaluation of MT, but most of the procedures are becoming standard for HT too. However, since it is still not clear whether this evaluation will become a new stage in the production process, or if it will be incorporated and integrate the role of revisers, this theme will not be developed in this dissertation.

2.5.4. The product of revision

The final product of a TP is, thus, not just a translated text but also a revised text. This embodies the recognition that, during the TP, translators manage a very fine balance between the information and structure of the ST and the intention and final use of the TT. This balance is challenged by the pressures on high productivity and by the tools translators use, which often make the TT too close to the ST. Frequently, only during the revision can the second translator move away from the ST, by focusing on his mission of producing a TT that fulfils an intended use that may be different from that of the ST.

Only a text that has been translated and revised can be expected to function as an autonomous text, conveying the information it is required to convey from the ST, complying to all the rules of the TL, and fulfilling the purpose that is intended for the TT. Translations do not always have to function as if they were originals, but the process that leads to such a result must be applied to all translations. This dissertation does not discuss the concept of quality, but it must be recognised that all translations must function as autonomous texts in order to fulfil the purpose of an efficient and effective communication process across languages.

3. COMPUTER SCIENCE AND TRANSLATION

This chapter will present a brief description of Machine Translation (MT) and of the different methods that have been developed by research teams and commercial labs to produce translated content by automated methods. The main focus of this chapter will be on Statistical MT (SMT), as not only the method that is most commonly used today in the industry and academia, but also because this is the method that is more open to interactivity with translators. The application of Machine Learning (ML) techniques to solve the problems that translation poses to technology will also be described. The chapter then makes reference to the most recent research on how to incorporate knowledge collected from the correction of MT output by translators, as a means to improve the quality produced by MT systems. The chapter ends with a description of evaluation methods and how these are employed to estimate editing.

3.1. What is "Machine Translation"?

"Machine Translation" is the most common designation for any method or system that aims at producing a version of an ST in a different language by applying automated methods. As generic as this definition may be, it has helped clarify the distinction between this type of technology and other types of translation technology intended to support translators in their work. Some relevant parts of this differentiation have been blurred, especially in view of the search for interactivity between MT and HT. Still, the term "MT" describes either whole systems that produce text in a TL autonomously, or the modules that produce such text to be presented as suggestions for editing by translators.

The purpose of producing a TT by automated methods has been present in virtually all reflections on how to improve communication across languages, even before the creation of computers. John Hutchins (1986) refers to the proposals made by Leibniz, Wilkins and Descartes in the 17th century, closely connected to the idea of decoding and creating a common language for universal communication. He also refers to authors in that century that created "mechanical dictionaries", such as Beck, Kircher and Becher. In the early 20th century, Artsrouni patented a machine he called the "mechanical brain" and Trojanskij a machine that served as a mechanical dictionary in a three-stage process. In this process, a monolingual human editor would edit the ST transforming all words into their base forms and annotating their syntactic function; the machine would then convert the base forms from one language into the other; and at the final stage a human editor with knowledge of the TL only would convert the sequence of converted base forms into the normal forms in the TL. Trojanskij's view of translation and language was of a clearly mechanical process in which only piece replacement needs to be done. For a more detailed view of the contributions by Artsrouni and Trojanskij, another paper by Hutchins is a good source: (Hutchins, 2004).

Warren Weaver's memorandum is known as the first description of the use of computers for translation (Weaver, 1949). In this paper, Weaver links the cryptographic efforts during the Second World War (by researchers such as Alan Turing) to the task of deciphering a text in a foreign language. He refers to anecdotes of researchers failing in several attempts to decipher a text, until they realised in which natural language it was written. The source for a decoding system, the author points out, is the fact that all languages use the same coding pieces: letters in different combinations, frequencies, patterns and other compositions. The belief in this common structure is closely connected to the notion of the "universal language" that is at the foundation of so many language theories throughout the centuries and with particular importance in the 20th century. But the motivation to produce translation through mechanical means is strong. Weaver quotes several exchanges with other researchers that describe the problem of translation in the context of post-war:

> *A most serious problem, for UNESCO and for the constructive and peaceful future of the planet, is the problem of translation, as it unavoidably affects the communication between peoples. Huxley has recently told me that they are appalled by the magnitude and the importance of the translation job.* (Weaver, 1949, p. 5)

It is then that he writes his most famous quote: *When I look at an article in Russian, I say "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."* (Weaver, 1949, p. 5) In the mail exchange, Weaver, linguists and computer experts discuss how this task may be completed. Some of the language features discussed are: ambiguity, word boundaries, the semantic richness of simple lexical units, multi word units, and so many other concepts that mark MT's research to this day. In commenting on the difficulties of the task, Norbert Wiener, one of the researchers quoted, states: *"you must remember that in certain* *respects basic English is the reverse of mechanical (...) At the present time, the mechanization of language, beyond such a stage as the design of photoelectric reading opportunities for the blind, seems very premature."* (Weaver, 1949, p. 7) So, it seems that researchers were quite aware from the beginning of the specificities of language as an object of automation. But, somehow, translation between languages seemed to be viewed as a simpler task than language analysis. Weaver discusses the limits of "word-to-word" translation, but he suggests that for certain text types, technical mostly, this might be an acceptable strategy. The approach he envisages implies several focused methods: reducing the scope of word meanings translated, controlling the input, by making it as simple as possible, analysing the most ambiguous words (those with lexical content, like nouns, verbs and adjectives) with their surrounding context to improve the possibility of acquiring their meaning unambiguously, and considering the statistical dimension of the problem. Like many authors after him, Weaver states: *"Perfect" translation is almost surely unattainable. Processes, which at stated confidence levels will produce a translation which contains only X per cent "error," are almost surely attainable.* (Weaver, 1949, p. 12) So, the solution to the translation problem is no longer to solve all difficulties, but to tackle more attainable issues, as he puts it: not to communicate from the top of one tower to the next, but to go down to the ground and identify the routes that link the towers.

This reading of Weaver's memorandum explains many of the lines and different methods that have been attempted to produce translated text with technology. There were references to the deep rules that languages have in common, which led to the "rule-based" approaches that will be described in section 3.1.1, but the statistical features of language (words, frequencies, patterns, etc.) and statistical approaches to data problems, which are at the core of "data-driven" and "statistical" approaches to MT, were mentioned too.

A final note on MT in general, before a brief presentation of the main methods employed to do MT. Although each of these methods is developed and presented by different teams, from different schools of thought, and though they seem to be conflicting at times, there is a lot in common between them. One common feature is that all of them try to incorporate the most recent advances available in NLP research. In fact, MT has always been seen as the principal computer application for languages (Liddy, 2001). Every product of NLP research is applied in MT research, either to add features that were neglected by previous approaches, or to optimise the processes, or to improve the quality of the results. MT research usually starts with what are called “shallow features”, i.e. lexical data only – words and their frequencies, processed as simple lists created by tokenizers; they move from there to incorporating features such as parts-of-speech, produced by automatic taggers; they add syntactic analyses by parsers, semantic models and ontologies, and eventually they incorporate pragmatic and discourse features. They also employ parallel corpus processing features, such as alignment techniques to find matches between two languages. Methods and tools coming from other disciplines such as Information Retrieval, Knowledge Extraction, Data Mining, Question Answering or Summarization are either used as helpful inputs to MT methods, or as fields to which MT methods may contribute. It is not surprising then that MT is considered “the Holy Grail of NLP” (NYU - Department of Computer Science, 2011).

The next sections present the four main types of MT. The longevity of RBMT will be mentioned, and a special reference will be made to how much EBMT resembles the recent advances in CAT, especially in relation to sub-segment alignment. Finally, SMT and NMT will be presented in only introductory form. This leaves room for a whole section with a detailed explanation of the centrality of SMT, not just for this dissertation, but also for current research in MT. The technical aspects of NMT will be approached in the section about ML.

3.1.1. Rule-based Machine Translation (RBMT)

Noam Chomsky’s work on generative and transformational grammar (Chomsky, 1965) created the theoretical background that was to be explored by the first wave of MT technology. Based on the works of Chomsky and other generative grammarians, researchers applied grammar and syntactic rules to three linguistic processes: analysis (identifying the deep structures of a SL sentence), transfer (identifying the corresponding deep structure in the TL) and generation (creating a TL sentence from the deep structure). The two main components of these systems were dictionaries, which included word lists with inflection rules in each language, plus the correspondences in words between the two languages, and rule engines, which focused on the three processes (analysis, transfer and generation). The major difference between systems and architectures was on how much emphasis was put on each of these processes by each system.

This generation of MT is known as RBMT – Rule-based Machine Translation. Knowledge-based Machine Translation (or KBMT) is a designation for systems that add semantic deep analyses to RBMT systems (Nirenburg, 1989). Because of the process of going from surface to deep structures, common to all systems, these systems were associated with the theories of language universals and universal grammar, and were criticised like everything else related to those theories (Russell & Norvig, 2009, p. 291). However, the model of how to go from an ST to a TT remained untouched for several years, as syntactic rules were seen as the only proper way to perform the task.

One of the most successful projects of RBMT was MÉTÉO, a system developed in Canada for the translation of weather reports, which was active from 1976 to 2001 (Hutchins & Somers, 1992, pp. 207–220). Another extensive project was EUROTRA, which ran from 1982 to 1994 at the European Commission (Commission, 1994). However, not even these projects could be maintained forever, after the global perception of the low return on investment that was associated with RBMT. The ALPAC report (Pierce et al., 1966) inaugurated the critical take on this technology, but the difficulties in improving RBMT results after years of investment were clear even in the most ambitious projects.

Still, companies like Systran (http://www.systransoft.com/systran/) kept on investing and developing RBMT systems, and integrating new technologies to their systems. In Spain, several projects that are strongly linked to academia have proved their worth in more recent years, especially in the context of languages with close similarity. Two of those projects are Apertium (Tyers, Sánchez-Martínez, Ortiz-Rojas, & Forcada, 2010) and OpenTrad (Loinaz et al., 2006). Both these systems are open-source and have been used in successful commercial applications, namely in daily applications in the written media. The RBMT technology has also been tested as a complement to SMT systems, in so-called “hybrid systems” – see below. These success stories from a technology that has existed for such a long time have highlighted that the rationale behind it, looking for a common structure that permits the generation of a TT by the use of rules and dictionaries, should not be wasted. Although the errors that these systems produce are more visible, the knowledge about language brought about by researching these methods is invaluable.

RBMT is not known for having developed systems that interact with translators. Most of the development effort is put into building complex syntactic and transfer rules. To users, most systems only open up the dictionaries, for replacement of corresponding words, correction of PoS identified by the analysis module, or association of inflection rules. A relevant exception is PAHOMTS (Aymerich, 2005), a system that includes several editing macros that will be described in Chapter 4.

3.1.2. Example-based Machine Translation (EBMT)

In the early 1980's, researchers were looking for alternative ways to go about the translation problem. Starting from the observation that RBMT does not work well with languages that have very different structures, Makoto Nagao presented a proposal based on the process of learning a language by comparison of sets of parallel sentences (Nagao, 1984). In his model, dictionaries of correspondences between words are expanded through thesauri, and structural analysis is replaced by connections between words and groups of words. The paper ends with the author admitting that one of the challenges of such a system is how to choose the best phrases in the generation of the translation. Nagao calls this an "MT system by analogy". However, this technological framework will come to be known as Example-based Machine Translation (EBMT).

A literature review published in 1999 in the "Machine Translation" journal presented the new paradigms of translation technology that were tested at the time (Somers, 1999). The author mentions the shock caused by the presentation, a few years earlier, of the IBM SMT models. His reading of those years was of a growing number of proposals that had in common the fact that they left behind linguistic rules, and used corpora as the central element of their systems. However, this common trait was not enough to avoid confrontation between those that developed systems that were totally focused on statistical methods and those that chose approaches that relied more on theory and data.

As Somers points out, EBMT is very close to the notion of TM and CAT tools, namely because of their views on the parallel processing of an ST and a TT. In fact, the notion of "matching" is central to the two technologies. Somers discusses the problems EBMT faces in terms of selection of examples and matches, annotation, generalization from examples, and many other typical challenges of data-driven technologies. Another dimension that is discussed in this paper is the computational problems that arise when so much depends on volumes of data that increase with the amount of detail that systems have to process. The author presents several approaches to EBMT, showing that hybrid solutions are being sought to advance all MT technologies. The article finishes by pointing out some of the advantages of EBMT, the main one being the fact that it stands on real data instead of conflicting rules. And it recalls its main issue: how to guarantee scalability in view of the direct proportionality between the increase in the volume of data and the need for better technology solutions.

"Recent Advances in Example-Based Machine Translation" presents the state-of-the-art of EBMT in the early 2000's (Carl & Way, 2003). The different articles in this volume discuss the definition of EBMT, its base in corpora and the closeness to MT, the terminology used in the area, and how EBMT is used in hybrid and multi-engine systems, together with technology from CAT and RBMT. In the introduction to the volume, the authors explain that, although the sentence is the ideal TU for EBMT systems, the role of an EBMT system is to find the best matching examples through a similarity metric, which may go below the sentence-level. After identifying these matches, the systems process them in a way that recalls the editing actions seen above. However, in this case, the actions are performed by the machine:

> *Suitable sub-sequences are iteratively replaced, substituted, modified or adapted in order to generate the translation. While this replacement, substitution, modification or adaptation may be completely rule-driven or also data-driven, the transfer itself, i.e., the mapping of a source segment into an equivalent target segment, is largely guided or acquired from translation examples.* (Carl & Way, 2003, p. xix)

Andy Way wrote another review of the state-of-the-art of EBMT a few years later, and this time the focus of the comparison is on SMT (Way, 2010b). One of the points stressed by the author is that EBMT always presents the "correct" translation (according to the reference) when there is a 100% match in the data, just like in CAT, whereas SMT tries to find the closest match from a combination of chunks, rarely achieving a 100% similarity. The article discusses the scalability issue and how this relates to the integration of better resources, namely linguistically annotated data. A particular feature of EBMT is sub-sentential (or sub-segmental) data, i.e. alignment of chunks or phrases. Way discusses the availability of data with this sort of alignment, and presents several methods for obtaining it, some of them based on linguistic features (such as methods based on the identification of function words, cognates, or constituent boundaries), while others are based on features such as length, “symmetry”, “overlaps”, “sample” or edit distance, which will be presented in section 3.5.1 below. The process described by Andy Way in this paper follows several steps, each one with its own tasks and issues:

- Pre-processing – this includes filtering and smoothing the data, chunking, and aligning units;
- Search/retrieval – identifying the units in the data;
- Recombination/decoding – the most difficult task, it implies the composition of the target sentence, based on the aligned chunks retrieved from the data;
- Post-processing – which implies correcting details and improving the output.

In a different article published in the same year, these steps are simplified as: matching (searching for fragments in the source); alignment (identifying the corresponding units in the TL); and recombination (composing the TT from these units) (Way, 2010a).

The definition of what exactly makes an EBMT system is not a clear one (Hutchins, 2005). EBMT was differentiated from RBMT by doing away with linguistic rules, and from SMT by not focusing on the development of statistical instruments to identify matches in large volumes of data. This unstable identity may be one of the reasons why it never became an important model in the MT technology landscape.

EBMT never gained enough momentum, although at least one commercial CAT tool implemented EBMT technology: Déjà Vu (http://www.atril.com/node/177). Notably, translators often referred to Déjà Vu as one of their favourite tools, at least in the early 2000’s (Chan, 2015). EBMT has two features that make it very interesting for translators: its closeness to the CAT tool purpose and design, and the focus on phrase alignment.

Andy Way argues that sub-sentential alignments make EBMT more understandable for people with no statistical background (Way, 2009, p. 8). In that same paper, we find the following quotation, from another paper authored by him and Mary Hearne:

> *Thus, RBMT and EBMT focus on the best way to generate a translation for each input string, whereas SMT focuses on generating many*

> *thousands of hypothetical translations for the input string and working out which one is most likely. In seeking to understand SMT in particular, this is a key distinction: while the means by which RBMT and EBMT generate translations usually look somewhat plausible to us humans, the methods of translation generation in SMT are not intuitively plausible. In fact, the methods used are not intended to be either linguistically or cognitively plausible (just probabilistically plausible) and holding onto the notion that they somehow are or should be simply hinders understanding of SMT.* (Way, 2009, pp. 10–11)

3.1.3. Statistical Machine Translation (SMT)

After the demise of the strategy based on linguistic rules, MT lost a lot of its attraction, until new strategies brought it back to the forefront. The available technologies, together with the available volumes of data, enabled this comeback.

The term "Statistical Machine Translation", or SMT, encompasses several different methods to produce a translated text without human intervention. In these systems, statistics plays different roles. Not only because of that, but also because Statistics and Computer Sciences are two different disciplines, several authors would prefer SMT to be known as "data-driven MT" (Tiedemann, 2008).

Apart from early proposals like "Stochastic Methods of Mechanical Translation" (King, 1956) to apply probabilities to improve the selection of the best translations, SMT started in 1988, with the presentation and publication of "A Statistical Approach to Language Translation", by Peter Brown and colleagues from the IBM J. Watson Lab (Brown et al., 1988). In this paper, the authors present the outline of their proposal to sustain the search for alignments of sets of words by using only estimations of probabilities of words appearing together in a natural language text, and the estimation of matching words in two languages, as they appear in parallel (translated texts). They call the result of this second alignment process "the glossary". Language features like relative frequencies of words in sentences (some much more frequent than others) are treated as a statistical problem, with normalization factors being introduced into statistical models, iterative re-estimation (guessing and readjusting calculations in increasingly longer n-grams), distortion factors to account for changes in the order of the words in a sentence, and other methods to build a system that models language and translation as sets of probabilities. The text published by the same team five years later (Brown et al., 1993) presents a more mature proposal, with different levels of sophistication. The basic assumption of that paper is that a method that does not depend on linguistic information is applicable to different language pairs, since it may capture features that are present in any bilingual corpora.

The translation models presented by Brown and colleagues were essentially word-based. At around 2003, papers like “Statistical Phrase-Based Translation” (Koehn, Och, & Marcu, 2003) started proposing systems that extended the alignments and the models to phrases. This approach was called Phrase-Based Statistical Machine Translation (PBSMT), and it became the standard that brought stability and success to SMT.

These early architectures of SMT only dealt with lexical items, words with no linguistic annotation attached to them. However, the availability of annotated data and the assumption that this could increase the capacity of SMT to estimate the translation of words that were not in the training data led researchers to expand their systems to manage this enhanced information. In order to do that, new methods were employed, namely those that were to be called “log-linear factored” models (see below section 3.2).

At this stage, research increasingly involves ML techniques, with research papers devoting more time to the decoders that look for connections in huge amounts of multidimensional data. SMT systems may incorporate syntactic analyses, do tree-to-tree alignments and do the same with semantic data. Besides, they may align words and sub-phrases in a hierarchical way, with phrases containing sub-phrases. This gives rise to what is known as “hierarchical SMT” (Chiang, 2007).

SMT broadly deals with words and phrases (groups of words with statistical relevance, not linguistically motivated) as pieces of puzzles, and then looks for the right places for these pieces to compose correct sentences. Statistics, or ML, are the techniques to deal with the information attached to these units. This framework can be combined with other MT systems.

As the second decade of the 21st century approached, a special focus was given to “system combination”, “multi-engine systems” or “Hybrid MT”. In 2011, two joint events gathered research on the use of linguistic information for Hybrid MT, and to the application of ML techniques to optimise the different tasks involved in Hybrid MT (Labaka & Melero, 2011). In these events, researchers discussed strategies that combine techniques from SMT and RBMT to deal with problems such as translating into morphologically-rich languages, word disambiguation, and evaluation of MT. Hytra (Hybrid Approaches to Machine Translation) workshops were held in 2012, 2013 and 2014, and a book was published based on the work presented at these workshops: "Hybrid Approaches to Machine Translation" (Costa-jussà et al., 2015). This volume is an extensive analysis of research in this area, encompassing hybrid systems led by SMT and hybrid systems led by RBMT, processes to add linguistic knowledge to SMT and the inputs by ML and NLP to SMT. Hybrid or combined systems sometimes have the purpose of correcting output from one system with another, an approach that will be analysed in section 3.4.

SMT's success as a technology for translation was confirmed and popularised (some would say democratised) with the launch of Google Translate in 2006. Microsoft and other companies soon followed the lead and launched similar products and services. The hype around these technologies was such that soon the message seemed to be that the "problem of language barriers" was to be solved. In an article called "Sleeping with the enemy? Or should translators work with Google Translate?", together with Belinda Maia, I presented a broad view of the relationship between translators and this popular technology (Carmo & Maia, 2013). In this article, we discuss not only the most enthusiastic views coming from Google, but also its own admission of the limits of the technology. We also mention comments from translators that highlight that SMT's basic architecture (building translations from chunks) make it very similar to CATs, rather than a system that is capable of translating a sentence it encounters for the first time.

The next sections analyse SMT in more detail. For now, let us look at the technology that is attracting all the hype at the moment.

3.1.4. Neural Machine Translation (NMT)

This new method of doing translation with no human intervention has gained such importance in recent times that it demands a special place in this dissertation. However, this technology does not play a central role in the research carried out in this dissertation, not only because access to it still requires specialised knowledge, but also because it does not offer itself as a good framework for interaction with translators. This section briefly presents its evolution and enthusiastic reception by the translation industry, followed by more analytical views of actors in the industry and researchers.

Mikel Forcada and Ramón Ñeco are quoted as the precursors of NMT (Forcada & Ñeco, 1997). They present a system that learns translations by a mathematical function they call Recursive Hetero-Associative Memory. Except for a few papers (Kalchbrenner & Blunsom, 2013; Le, Allauzen, & Yvon, 2012), among others, only in 2014 did experiments with the application of Neural Networks (NNs) to MT start to gain some visibility. Examples are the papers published by Kyunghyun Cho and his colleagues from the University of Montréal, but other teams were also working with these technologies. (Cho, van Merrienboer, Gulcehre, et al., 2014; Cho, van Merrienboer, Bahdanau, & Bengio, 2014; Devlin et al., 2014). In 2015, Cho published a few articles in Nvidia's blog, explaining how NNs work. These also explain how the use of GPUs (graphical processing units), instead of the CPUs (central processing units) that are at the core of personal computers and servers, improve the performance of the systems (Cho, 2015a, 2015b, 2015c). But it was only in 2016 that commercial companies announced outstanding results by applying NMT techniques to the vast amounts of data in their hands. Google (Wu et al., 2016) and Systran (Crego et al., 2016) presented their systems in this year, and the world of translation technology entered a new age of euphoria. Since then, virtually all companies investing and selling services based on MT are presenting their own NMT solutions.

NNs have been used mostly as components in SMT systems, with the role of re-ranking or re-scoring translation hypotheses, but full NMT systems have also been developed. We should only talk about "NMT" when we refer to systems that apply NNs to the whole process of translation. The paper "Recurrent Continuous Translation Models" (Kalchbrenner & Blunsom, 2013) is the first example of a full NMT system.

Testing NMT

A good deal of the research on NMT has focused on testing the technology in specific contexts. Neubig and colleagues (Neubig, Morishita, & Nakamura, 2015) investigated the effects of using NNs as a method to re-rank phrases in syntax-based SMT systems, on grammatical items such as reordering of phrases, insertion/deletion of copulas, coordinate structures and verb agreement. The authors conclude that the correction of word order is one of the strengths of NMT. Rico Sennrich (Sennrich, 2016) studied how well NMT can learn specific grammatical structures. In his work, translation errors are introduced in the data, side-by-side with correct versions. Then, the NMT learning system was run, to see whether it would assess the corrected versions better than the ones that contained the errors. The errors tested were (some of which concern specific issues of the translation into German): NP agreement, noun-verb agreement, separate verb particle, polarity, and transliteration. The tests focus on character-level decoders. The main conclusion of this study is that character-level NMT manages unknown words better, but it does not behave so well with morpho-syntactic agreement, which is a long-distance phenomenon that most NLP technologies find difficult to tackle.

Bentivogli and colleagues (Bentivogli, Bisazza, Cettolo, & Federico, 2016) measured NMT against PBSMT with the focus on morphological, lexical and word order errors. Their main conclusion is that NMT is better than SMT in all factors, namely in reducing the PE effort, but that word order is the factor in which NMT gets the most impressive results. The tests performed in this study were later expanded by Antonio Toral and Víctor Sánchez-Cartagena (Toral & Sánchez-Cartagena, 2017) to nine language pairs, and involved more evaluation criteria. However, their evaluation methods were automatic only, with the authors admitting that the strategy to measure some of these (like testing fluency based on the perplexity of the systems) is arguable. The only conclusion from this study that does not confirm Bentivogli's inferences is that, in this study, NMT presents worse results in longer sentences. Another paper with a wide evaluation (30 language pairs), and which concludes that there are clear advantages of NMT systems, is also based on automatic evaluations only, assessing the speed of the systems and the quality they produce according to BLEU (a metric discussed in section 3.5.1) (Junczys-Dowmunt, Dwojak, & Hoang, 2016).

The most recent paper that assesses NMT includes not only automatic, but also human evaluation. It is an extensive work called "Is Neural Machine Translation the New State of the Art?" (Castilho et al., 2017). The paper presents three use cases, each one with different views and different conclusions. The main global conclusion is that, although automatic evaluations clearly show a preference for NMT over SMT, the results of human evaluations of both technologies are not so clear. In one of the use cases, patent information, NMT obtained better results for titles of articles, whereas SMT performed better in their abstracts. In this type of text, NMT also committed more errors of omission, and it presented fewer sentences that required no corrections. The third use case was in the ambit of TraMOOC (http://tramooc.eu/), a project that involves Online Learning (OL), a technology that will be analysed in different sections below, and this is the case with most detailed results. Although all results concerning PE effort were better for NMT, human evaluators reported that NMT errors were more difficult to identify, because errors of fluency and word order, which are more frequent in SMT, are easier to detect. NMT is, almost consensually, considered to produce more fluent translations. However, in this last use case (and also in one of the previous studies) NMT more frequently presents errors of omissions, additions and mistranslations. These are essentially, examples of errors of adequacy, which means that NMT is more capable of grasping correct units in one language than between the two languages. The authors claim that NMT may not be, just yet, at the level of quality of SMT (or above, as it is so often claimed), if you consider human evaluation only.

With NMT under the spotlight, a few critical reviews from the industry began to appear in late 2016 and 2017. "Beyond the Hype of Neural Machine Translation" was presented at a TAUS conference by the representatives of two different MT companies (Bartolomé & Ramírez, 2016). They argue that, although NMT is better than SMT with in-domain data, SMT is better than NMT according to specific criteria: it is more robust, it presents better results when trained on generic data, translators' feedback tends to be better for SMT and it presents a higher return on investment. In September 2016, Slator published an article with interviews with experts (from the industry and universities) on how they evaluated Google's system and claims (Faes, 2016). The article shows that opinions are varied and that a lot of tests are needed to establish objective evaluations. In this and other articles, industry consultant Kirti Vashee called the Google announcement a publicity stunt (Vashee, 2016b). Jost Zetzsche presented a clear and simple view on NMT for translators (Zetzsche, 2016a), and Arle Lommel comments on the impact of NMT for the translation industry (Lommel, 2017a, 2017b).

In one of these articles, Lommel discusses "zero-shot translation", a challenge Google says it is capable of achieving with its "Multilingual NMT" (Johnson et al., 2016). This ambitious goal is the capacity of a system to translate into a language pair it has never processed before, by using "transfer learning" from other pairs involving the same languages. Achieving this would mean that NMT had been able to "bridge the gap" between RBMT and SMT: a system that does not depend on specific data, like RBMT, and that, at the same time, has no explicit rules, like SMT, seems to be the ultimate goal of a real autonomous MT system.

3.1.5. The Machine Translation industry

This section is a brief overview of the current state of the localisation industry after the impact of SMT. As explained in the Introduction, the attitude to publishing research developed in academia and in industry is very different, the first publishing everything online, and the latter hiding everything under proprietary protection. So, it is not easy to see what exactly is under the hood in commercial systems, when they claim that their systems are hybrid, or when they publicise NMT systems.

However, business models are often quite transparent, and it is important to comment on a radical change in the industry with the advent of web MT. The free availability of MT, notably Google Translate, brought with it this common notion that MT is not only easy, but also cheap or even free. Commercial companies at the forefront of the localisation industry, although they saluted the efficiency and productivity of SMT systems, soon realised that it would be hard to make money out of selling translated products created by SMT. The first companies presenting free SMT services, like Google and Microsoft, did not offer translation as a service, their main business being big data management and knowledge extraction from it. Since SMT brought with it a big expansion of that data, together with a huge increase in user interaction with their online platforms, these big data companies saw huge advantages in offering these solutions for free.

The solution for companies whose business was translation technology solutions was to focus on productivity, quality and specialisation. They started developing systems adapted to specific domains, applied metrics that showed increases in productivity and quality and started selling their systems as supports to translation service companies and translators. In the era of SMT, there are many new technology companies selling MT services, but not many selling translation products directly to the final user. Instead, new services and tools appeared to focus on simplifying the adoption of MT.

Nowadays, most translation technology companies concentrate on selling MT systems linked to multilingual data management services targeted at translation service companies and multilingual companies with large amounts of proprietary data. Some use generic data, but others claim to only use in-domain data. Iconic Translation Machines (http://iconictranslation.com/), KantanMT (https://www.kantanmt.com/), Omniscien (https://omniscien.com/), PangeaMT (http://pangeamt.com) and Tauyou (http://www.tauyou.com) are some of the companies that offer this type of service.

But some companies also sell individual software packages that allow translators and translation agencies to set up their own SMT engine. Slate Desktop (http://pttools.net/) is the main product in this line. It is based on Moses (see section 3.2.1 below), and basically it provides a streamlined graphical interface to the process of training, tuning, and running a Moses SMT system. One of the main advantages offered by such a system is control over the data and confidentiality of the whole system. The MT results may then be fed to the user's CAT tool.

Slate Desktop follows the same objective as "Moses for Mere Mortals" (Machado & Fontes, 2011). This is not a commercial tool, and it does not feature a graphical interface. It is a set of scripts developed at the Directorate General for Translation of the European Commission, as a personal project of professional translators, with an easy-to-follow work process. As a particularly interesting feature for researchers, it has the possibility of analysing the result of each stage, such as the tuning of different parameters, or the possibility to look into the phrase tables. Accessing these tables may be an invaluable resource for translation process research, and for linguistic analyses of the so-called phrases or chunks.

On the side of CAT tools, it is important to note that nowadays all have plugins for external MT services, which are limited to importing one MT hypothesis for each sentence. However, some have developed extended features like "AdaptiveMT", a technology that is said to apply learning methods to continuously improve the results the MT produces.

In terms of free MT services and tools, besides Google Translate, Google offers Google Translator Toolkit (https://translate.google.com/toolkit/), an online CAT tool that provides both MT and TM, with the added advantage of a collaborative environment. However, it is not clear whether there are many users of this platform. Microsoft offers Bing Translator, but it has no editing tool for translators.

Several of the companies referred to in this section are spin-offs of research projects or are very closely linked to research. MateCAT (www.matecat.com) was developed as a research project, and its free online tool is widely used by people who want a platform that gives them MT content (machine-translated content). Lilt (www.lilt.com), a tool that will be analysed further ahead, was also developed from a university research project.

3.2. The importance of Statistical Machine Translation

This section presents the basics of the technology inside SMT, looking for the reasons for its success and for its interest as a framework to develop better tools for interaction with translators.

3.2.1. Technical overview

The IBM word-based models

The research by Brown and colleagues at the IBM J. Watson Lab presented the foundational terms and models for SMT technology (Brown et al., 1988; Brown et al., 1993). The 1993 paper describes the two probabilistic models that SMT uses to identify the best units to translate a ST into a TL: the Language Model (LgM) and the Translation Model (TrM). The LgM describes in a table the probability scores of differently sized units (from one word to a limited-length n-gram) to occur in the texts of one language. This table is used to find the probabilities of two words appearing together more often than similar-sized groupings, or than different sized groupings that include those words. The LgM of the TL is more important than the one for SL, because these probabilities are the indications that such groupings with higher probabilities may compose a fluent TT. The TrM reports in a table the probabilities that different sized units in one language will align with similar or differently-sized units in another language. In the same way that the LgM is associated with fluency, the TrM is associated with the adequacy, or faithfulness, of the translation (Jurafsky & Martin, 2009, p. 911). These concepts will be revisited in the section on the evaluation of MT (section 3.5).

All SMT systems are trained with bilingual data, which basically means extracting the two models (LgM and TrM) from parallel corpora. The results from the learning stage are saved in "phrase tables" that report the different alignments that were achieved by the system. After the system has the phrase table, an algorithm known as the "decoder" looks for the best combinations for the units that it identifies in the sentence that it is translating. These decoders are one of most studied pieces of these systems.

Brown and colleagues present the formula that describes the functioning of these systems as the "Fundamental Equation of Machine Translation" (Brown et al., 1993, p. 265). This formula also became known as the "source-channel" or "noisy-channel model", because of its foundation in communication theory. Figure 3 below is not Brown's representation but one that allows a good explanation of its functioning.

$$\hat{e}_1^I = \underset{e_1^I}{\operatorname{argmax}} \{Pr(e_1^I) \cdot Pr(f_1^J | e_1^I)\}$$

Figure 3 – The fundamental equation of MT (Och and Ney, 2002, p. 295).

A few notes to explain the formula: the letters chosen to represent the languages (e and f) are associated with the fact that the examples at the basis of Brown's papers were English (the TL) and French (the SL). These letters have remained as a convention for SMT papers to refer to SL and TL, even when those languages are not involved. Besides, the letters I and J and the number 1 placed one above the other, as indices, represent a sequence, where 1 is the first symbol and I (in the TL), or J (in the SL), is the last symbol in the sequence. This representation of the sequence means that these calculations must be done in the sequence, one symbol (or word) after another.

This formula describes the objective (ê), which stands for "the best translation", of a sequence (I/1) as being the result of a calculation that determines the best result the product of the probabilities of the units in one language [Pr(e)] with the conditional probabilities of the alignment of units between the two languages [Pr(f|e)]. So, the system calculates the probabilities of all sequences of symbols in each language [Pr(e)], and builds the target LgM. It also calculates the conditional probabilities that the symbols in SL depend on the symbols of the TL – "Pr(f|e)" is read as "the probability of f given e". This is the TrM.

Then, during decoding, a mathematical formula "argmax" builds a new sequence of words from the LgM of "e" (this sequence is represented below argmax) by searching: calculating the product of all these probabilities and choosing the sequence which has the highest probable score of not only being a good alignment of the words in the SL, but also of making up a correct sentence in the TL.

So, the problems an SMT system faces are divided into three: language modelling, translation modelling and search. The authors admit that this is not an accurate representation of the human TP:

> *"One can hardly imagine someone rifling mentally through the list of all English passages computing the product of the a priori probability of the*

passage, Pr(e), and the conditional probability of the French passage given the English passage (Pr(f|e)*."* (Brown et al., 1993, p. 265)

However, the authors stress that this description is perfectly valid from the formal point of view. In "Statistical Machine Translation", the most important reference book on SMT, Philipp Koehn takes a similar stand: although the TP in SMT does not model the human process, it is useful to describe the SMT process as if it followed a human decision process sequence (Koehn, 2008, p. 156).

Brown et al. then present and discuss 5 TrMs and the algorithms to compute them, solving new problems, but adding complexity. These models are often referred to as IBM Model 1, Model 2, etc. The authors explain that training is a process by which the likelihood of the translation is first guessed, based on a set of parameters, and then the maximum value for this likelihood is approached by applying iteratively the EM (Expectation-Maximization) algorithm (Dempster, Laird, & Rubin, 1977). This algorithm has two moments: one which estimates a value, and another that checks if the value may be improved, or optimised; then, this optimised value may be processed as the new expected value in other phases of expectation/maximisation. Brown's models 1 and 2 assume that the sentences in the two languages have the same length, but model 2 takes the order of the words into consideration. Model 1 is used to set the initial estimates, which are used for the maximization calculations of the subsequent models. In Models 3, 4 and 5, the concepts of "fertility" and "distortion" are introduced. Fertility is the number of words in the other language each word in one language is connected to, in a random alignment. Distortion is the possibility of the alignment not being made with the word in the same position, i.e. that a reordering of one of the sentences is necessary. The differences between models 3, 4, and 5 concern the way alignments are corrected, considering fertility and distortion, using only mathematical approaches. The authors show how alignments change with the number of iterations. They also admit that, while the 5 models are effective in obtaining word-for-word alignments, they may be improved, admitting that morphology and sensitivity to multi-word units could improve the system, but they stress that dealing with these language phenomena would increase the complexity of the models. The authors end the paper by explaining that, although the linguistic content of their models is scant, their intention is to incorporate linguistics in a "secure probabilistic framework" in which both disciplines could cooperate.

Log-linear models

The next major change to the representation of the SMT work is called "log-linear models", and they were presented in the early 2000's by researchers such as Franz Och and Hermann Ney, from the University of Technology of Aachen (Och & Ney, 2002). Log-linear models combine several models of features per word, and calculate them together, in a linear operation. Linear operations are those that relate the inputs (data for calculations) and the outputs (results) in a way that, if any input changes, the outputs change too (a property that is called "homogeneity"), and additions to inputs are added to outputs too (the "additivity" property").

In log-linear models, each feature model (these are essentially n-gram models, per word or phrase, and can include LgMs, TrMs, reordering models, length models, and feature models associated with grammatical information) is calculated separately and integrated into the formula as a logarithmic function. The representation below helps explain how these models work.

In the formula below, the logarithm is represented by "log" and the $f_n(\mathbf{t},\mathbf{s})$ bit is each feature model, where "t" stands for "target" and "s" for "source". Each feature in this formula is associated with a weight, which can be tuned, as a way to enhance or reduce its contribution for the final result – the symbol λ denotes these weights for each feature function. The sigma (Σ) represents the sum of the sequence of all functions, and the symbols above and below it define the limits of the sequence.

$$\hat{\mathbf{t}} = \arg\max_{\mathbf{t}} \left\{ \sum_{n=1}^{N} \lambda_n \cdot \log(f_n(\mathbf{t}, \mathbf{s})) \right\}$$

Figure 4 – A log-linear model in SMT (Ortiz-Martínez et al, 2016, p. 59).

In these models, each word is no longer a simple unit, it is represented by a set of factors. So, these models are also known as "factored" models, as they turn features to factors in a logarithm. The log-linear model allows for the inclusion of several features in a vector that simplifies the mathematical processing of the data (see an explanation of vectors in 3.3.2). The features added to each word may come from grammatical analyses or from different probabilistic calculations.

In 2004, Franz Och (already working at Google) and Hermann Ney reinforce the centrality of alignment methods to SMT, and present a phrase-based log-linear model of translation (Och & Ney, 2004). In this model, after the word alignment, there is an alignment of phrases, giving rise to PBSMT. Phrase-based systems often treat reordering in a separate model, based on the TL. The authors discuss the search methods, highlighting the fact that search is a complex process, and that only approximations that are efficient but minimise the number of errors are possible. The training of the models is iterative, with a method known as MERT – Minimum Error Rate Training (Och, 2003).

The Moses toolkit

When Philipp Koehn and colleagues presented the open-source Moses toolkit, the possibility to use these techniques became widespread in the academic and commercial worlds (Koehn et al., 2007). Available as an open-source tool, it is widely used and it was considered very easy to implement and adapt.

Moses is a toolkit made up of different modules for different tasks, such as alignment, training, LgM, TrM, decoder, tuning, and so on, which allows it to continuously evolve by incorporating new and alternative technical solutions. A researcher may, for example, include specific tools to build a LgM adapted to his data, or choose to replace the default training and tuning method, MERT, by a different one. It includes word reordering modules, modules that penalize long sentences, so as to favour the identification of smaller units (thus improving the probabilities that these may reappear in other structures), different decoders and search algorithms.

Training and testing

It is important to stress that the development of an SMT system follows a common testing process, which enables researchers to improve and adapt its results. In this process, the system is presented with a new corpus (the “dataset”), which is divided into two different sets: the "training set” and the "testing set”. The training set is usually bigger that the testing set, in a proportion that may be, for example, ¾ to ¼ of the total dataset. The TL side of the testing set will be considered the "gold standard" – the translations that the MT system will try to reproduce, or “approximate”. Training refers to the stage in which the system extracts the LgM and the TrM from the training set. Then, the source sentences of the testing set are run through the decoder, which performs the “translation”, decoding and looking for the best translations for each phrase and each sentence in the testing set, which it is seeing for the first time. When it identifies the hypotheses that maximise the probabilities of the two models, the system composes the final MT hypothesis. The final stage is an evaluation stage, in which the final MT hypotheses are compared to the "gold standard", the initial reference translations that had been removed from the testing set. This evaluation is done by applying the evaluation metrics that will be presented in section 3.5. If the results are not satisfactory, the system may be tuned and retested.

Tuning may be focused on increasing the scores that evaluation metrics give to adequacy and fluency. To increase adequacy, the decoder gives more weight to the TrM; to increase fluency, the decoder gives more weight to the target LgM. Another result that is assessed in tuning is “overfitting” to the dataset, which happens when the results are good when tests are made with the testing set (which has strong similarities with the training data), but bad when tested with different data.

The system has to be retested and the training and learning stages must be repeated quite regularly. This repetition is necessary whenever a significant amount of new data is received, when a new tuning of any features is necessary, or when the system has to be tested with a new language pair or new in-domain corpora. Although there is no translation process embedded in them, the “phrase tables” that result from these different stages of testing are sometimes called “translation engines”. That is the reason why some companies offer “in-domain translation engines”: these are the results of training stages with specific data sets.

This section described the fundamental processing inside SMT systems, but it does not extend to the most sophisticated systems that involve syntactic and semantic data, and other approaches. These additions aim at improving the results of autonomous SMT systems, not adding much to perspectives on interactivity with human translators, which is the main focus of this dissertation. In the rest of this section, SMT will be described, first, from the point of view of data problems, and then from the phases in the SMT translation process.

3.2.2. Processing data

Linguistic data is not clean, regular, uniform and unambiguous, which means that researchers cannot treat it as numerical data. This section is a broad introduction to the different issues researchers have to deal with when thinking of linguistic data.

Corpus Linguistics

Since the 1990's, Corpus Linguistics evolved as a solid discipline, with a long tradition of conferences (such as LREC – http://lrec2016.lrec-conf.org/en/), and plenty of research published. It has developed its own methods to extract knowledge from amounts of data that humans could not process at a reasonable pace, and it produced numerous resources and tools to manage corpora. Some of this research is directed to parallel and comparable corpora (Maia, 2003), both of which are used in SMT. As an example of such tools, Corpógrafo, developed at the Universidade do Porto, is a very complete tool to manage corpora, including a basic parallel text aligner and a semantic analyser for Portuguese (Sarmento, Maia, Santos, Pinto, & Cabral, 2006). Corpus Linguistics has helped bridge the gap between research in language schools and data scientists.

"Corpus Methods for Descriptive Translation Studies" discusses methods to create corpora for TS (Zanettin, 2013). Some of the author's recommendations should be taken into account by MT researchers. This paper shows a few examples of studies around the concepts of translation universals, traces of individual style and translation norms.

For example, a parallel corpus should include the identification of which language is the source and which is the translation. This does not exist often, even in one of the corpora most used in MT, Europarl. MT research often assumes that there is no direction of text transformation in a parallel corpus, and that the parallelism between the two versions is so strong that the direction is interchangeable. But inverting a text transformation may affect profoundly the information one can take from this data. Another issue concerns the assumption that a parallel corpus must represent the two languages in a quantitatively symmetrical and proportional way, when that relation may not exist between the two languages. The author also stresses that *"representativeness and comparability are conflicting goals"* (Zanettin, 2013, p. 10), which means that if one wants one's data to represent specific trends in one language, genre or text type, one will probably lose the capacity to compare that data with that of another language. The inverse effect is also true.

Data annotation

The author discusses the challenges of annotation, dividing it into structural (segmentation and tokenization), linguistic (PoS tagging, lemmatization, parsing, semantic annotation and others) and interpretative. This last type of annotation deserves a highlight:

> *Finally, by interpretative annotation I refer to all other layers of annotation based on non-linguistic categories which can be superimposed to a text and which require close human supervision and manual coding. These include the classification and annotation of translation shifts, additions, omissions etc. in parallel corpora, as well as metaphorical annotation, error-tagging, etc.* (Zanettin, 2013, p. 11)

Zanettin stresses that most studies on translation shifts (a concept introduced in section 2.1.3) are based on manual classifications of aligned pairs. One may add that manual annotations of translations do not really describe the processes employed. Instead, they provide interpretations of what those processes might have been, or they are just comparisons between two texts in different languages. He ends by expressing concerns about the balance between quantity and quality, in terms of the extension of data analyses made from corpora and the intension of the descriptions of data.

Big data

SMT is based on "big data" processing. This means that SMT systems have the capacity to deal with billions of words in multidimensional models that include numerous variables and factors. It was the availability of this amount of data that created the conditions for researchers to test sophisticated mathematical techniques and methods. This approach, that sees translation as essentially a data problem, has issues that are worth discussing.

In MT, human translated corpora are seen as the “gold standard”, as representing the reference for “quality”. However, normal variations described by the concept of “translation shifts”, such as the different options selected by different translators, are often seen as unwanted effects, they are removed from the data, or they are not considered and affect the results obtained and their interpretation. Examples of approaches that suffer from this effect may be found in several papers, such as (Bentivogli et al., 2016; Fomicheva, Bel, & da Cunha, 2015).

Problems with data

Andy Way (2010a) presents the stages of pre-processing corpora for MT as being: clean-up, segmentation and tokenization. The first two stages aim at removing troublesome features from data and separating it into sentences or segments. The function of tokenizers is to separate everything into smaller units to be processed: punctuation signs become tokens and each word too. This is not such a straightforward process, since the concept of word is not an easy one. (For example, Portuguese and Spanish words integrate morphemes that must be processed separately.) The next stages of corpora processing depend on this, but this strategy has its effect on the results produced by SMT systems. Because of tokenization, punctuation and capitalisation errors are very common in PE correction lists.

Researchers and commercial companies producing MT solutions often identify problems with specific types of text and content that does not adapt well to MT. There is the recognition that the maxim “garbage in, garbage out” applies to a technology that depends so much on its inputs. The negative effect caused by low quality data was discussed in the translation industry in the context of systems, like Google Translate, that index and reuse huge amounts of untreated data from the Web. This effect was known as “polluting its own drinking water” (Wiggins, 2011). Some companies announce their offer as being based on pre-processing and cleaning the data before applying MT. The use of controlled language in MT has also been studied in different contexts, as a means to help improve the quality of the outputs through more regular inputs (Doherty, 2012; Gough & Way, 2003).

But even “clean” regular data may not be enough. A high parallelism (and a simpler alignment) is expected from bilingual data, not just at the sentence level, but at the word and phrase levels too. There is also the recognition that specific variables, like certain languages, text genres, technical domains, and others that are associated with low parallelism between ST and TT, may reduce the quality produced by SMT. Because of this, there is a considerable interest in using “in-domain” specialised data for training specific translation models or engines, for selected purposes (Bertoldi & Federico, 2012; Turchi, Goutte, & Cristianini, 2012).

Furthermore, systems that include linguistic features need good quality annotated data (Koehn, Haddow, Williams, & Hoang, 2010). However, annotated data is rare, data annotation is a very subjective task, and automated processes (like taggers or parsers) cannot produce fully accurate data, as discussed in papers like (Manning, 2012).

On the opposite extreme, too much data has its negative effects too. In a survey of data selection methods (Eetemadi, Lewis, Toutanova, & Radha, 2015), the authors comment that, although the increase in the volumes of data was the most important factor for quality improvements, they have fallen victims of their own success. The need for automated data selection methods is three-fold: much of the available parallel data is collected through web crawling and is noisy, there is too much parallel data to train and iterate on in a timely manner, and model sizes have become too large.

With so much data, systems lose discriminative power. Safaba (a company run by Alon Lavie of the University of Carnegie-Mellon) mentioned the “data dilution effect” on its site (Safaba, 2017). This concerns the effect produced by joining different sources of bilingual data in the same learning platform in a way that each source loses its importance. For companies with large investments in managing proprietary data and terminology this may be a major loss, since they will not be able to separate their highest priority choices from everyone else’s. European Portuguese users feel this effect when using generic SMT systems that do not discriminate between the two main variants of Portuguese. For example, Google Translate more frequently offers Brazilian Portuguese choices over European ones.

Finally, for the theme of this dissertation, it is important to note that the focus on data signifies that any learning, knowledge extraction, or the results obtained from this material is based on the products of a complex process. What the SMT systems are doing is looking at a product and retrieving as much information as possible to be able to create a similar product.

3.2.3. Three phases in SMT

Early researchers in SMT admit that their systems do not model the whole process, but it is interesting to look at these systems from the perspective of the human TP. SMT is usually described as containing essentially two phases: learning followed by decoding. However, the output of an SMT process is an MT hypothesis only, not a finished translation. Even if this is not discussed in these terms, there is, in fact, in the SMT community, a consensus over the recognition that the output of an SMT system needs to be post-edited before it can be considered fit for dissemination. One could argue that even the utility of SMT output for assimilation lacks demonstration. Besides, even if this utility were established, that should not be an argument for considering that a full TP could be completed by pure SMT processes. Indeed, the acknowledgement of the incompleteness of the TP by SMT is reflected in the investment in the editing processes that are described in section 3.4 below. To cut this discussion short, let us just reinforce the notion that interaction with HT has always been behind the greatest advances in translation technology.

So, it is argued that SMT should, in fact, be seen as including three phases: Learning, Decoding and Editing. And this echoes the three phases of the main description of the TP presented in the previous chapter: Orientation, Drafting and Revision.

- **Orientation/Learning** is the phase in which information is collected from the sources, and the parallelisms that form the basis of the translation decisions are established;
- **Drafting/Decoding** is the actual production stage, in which the draft translation is written, and the passage from one language to the other happens;
- **Revision/Editing** is the stage in which the final decisions are made, all errors are corrected and the actual result of the TP is achieved.

Although too simplified, this parallelism between HT and SMT establishes a basis for a comparison between these two major approaches to translation. It shows, for example, that there is no complete TP in SMT without the last editing stage, just as there is no finished translation without the revision stage.

For this dissertation, the most important phases are Learning and Editing. Learning will be analysed in section 3.3 and Editing in section 3.4. The decoding phase is briefly described below.

Decoding

Decoding is the name given to the stage during which an SMT system runs through the phrase table and selects the units that have the highest likelihood of being the best alignments and the highest likelihood of being able to create a correct sentence in the TL. One of the most used algorithms to do this is EM (expectation-maximization). However, many others are tried and tested, some in systems that combine several of these techniques (DeNero, Kumar, Chelba, & Och, 2010).

Decoding is one of the hardest problems in NLP. Kevin Knight proved that this search problem was NP-complete (Knight, 1999). This is a term that comes from the theories of computational complexity, and it refers to the hardest problems to solve using reasonable resources. Saying that decoding is an NP-complete problem means that *"exhaustively examining all possible translations, scoring them, and picking the best is computationally too expensive for an input sentence of even modest length"* (Koehn, 2008, p. 155). The implications of this acknowledgement are that researchers look for methods that reduce the problem, or approximate the solution, instead of hoping to solve the whole problem. In such circumstances, the focus becomes the methods themselves, instead of the results. The approach used in these cases is called "heuristic search", a method that recognises that, although the best results may not be found, there are great gains in analysing the processes that bring us closer to the best solutions.

It is no wonder then that decoding is one of the most studied dimensions of SMT. This is when most systems become a "black-box", even if translators try to understand what is going on inside the process. In fact, most of the literature concerning this phase involves discussions around the demonstrations of the mathematical properties of algorithms and techniques. Papers and books talk, for example, about reducing the search space by hypothesis recombination, pruning hypotheses with low informative value from tables, organizing hypotheses in stacks, doing "beam search", a process that focuses on the best results within defined thresholds, future cost estimations, which help to identify the easiest hypotheses to calculate, and many other approaches. There are also the approaches that focus on mathematical solutions to maximize and optimize the use of complex algorithms. Just two examples of papers like these are: (Ling, Graça, Trancoso, & Black, 2012), which discusses the use of an entropy-based system to prune and reduce the search space, and (Jawaid & Bojar, 2014) that present a two-step translation process which separates lexical translation and reordering from linguistic processes, like conjugation and declination for TLs with complex morphology. But the list of papers that focus on the decoding stage would be endless.

3.2.4. SMT and human interaction

In this dissertation, SMT is viewed as the most appropriate MT method to develop interaction with human translators. The reasons for this are explained below.

First of all, one can see a parallelism between the three main phases in the human TP and the phases in SMT, as described in the previous section. Second, it is the most developed type of system that anyone can have access to, in such a way that one can look inside and appreciate its flexibility and the multiplicity of alternative techniques it offers. Such open access to technology is fundamental not only because of the availability of tools to be tested and used in many different tasks, but also because one can understand the high degree of sophistication and the breadth of solutions that it offers for the description and treatment of language and translation problems.

One of the challenges for linguists when they deal with probabilistic studies for the first time is the descriptive power of such a reductionist approach that enables it to sustain applications such as SMT, but also comprehends revolutionary technologies used every day such as Text Compression, Web Search Engines, Voice Recognition, Question and Answering communication interfaces, and many more. There are several reasons for this: probabilities deal with uncertainty, with fuzzy data, just as translated data seems to. Besides, mathematical methods include specific mechanisms to deal with some of the problems of language, such as techniques to:

- manage entropy (the degree of uncertainty in a system – the higher the number of alternatives, the higher the entropy);
- account for exponential growth of complexity (when sentences grow bigger, the search space grows even bigger – this is solved by recombining hypotheses and pruning the search space);
- adjust dependency on samples (to amplify the capacity of generalisation, even if your sample is skewed to a specific language pair or domain);
- auto-adjust or auto-tune the system, by identifying details which may lead to erroneous results, and several other applications.

Specifically in SMT, some of these techniques are applied to anticipate the possibility that n-grams are not placed in the same position in the source and in the target sentences (these are managed by algorithms and models that deal with distortion); others deal with the possibility that a source n-gram leaves a blank space in translation, or even that the best translation must be replaced by a synonym, or deal with the number of alternatives that there may be for a word or phrase (perplexity). So, it seems that SMT has enough tools to deal with most of the problems it identifies. But this does not mean that translation became a simpler process that may be performed by automated processes only.

Additionally, although SMT is often sold as a technology that would, by itself, break the language barrier, it is in fact much more open to human intervention than one might think. This is evident from its dependency on human data, which begins with the inputs it stands on, but it extends to guiding the processes of the selection of the best translation decisions, culminating with the final editing stage. Finally, it is important to stress that the most important improvements that were achieved in SMT started with the focus on sub-segmental units, or phrases, which, as we have seen in the previous chapter, is fundamental for the development of better tools to support the work of translators.

The interaction that is opened by SMT will be analysed in section 3.4.2, where Interactive Machine Translation (IMT) will be described, and it will be further explored in section 3.5.3, on estimates of editing work.

3.3. Machine Learning in Machine Translation

For the mathematical discipline of Machine Learning (ML), “learning” means the capacity to “generalize” from examples, i.e. to extract, from a set of examples, a rule that may be applied to other occurrences (Domingos, 2012). The main tasks in ML are: efficient representation of data, application to new contexts, evaluation and optimization. In order to identify rules, the best learning methods are tested on its capacity to continually, and iteratively, reduce errors in the description of the data. So, the main focus of ML is to study how to make algorithms that are adjusted to the type of problem and the data that is processed. This is a simplistic yet very powerful description of learning: training, testing and tuning different algorithms, so as to select and adjust the best for each task. This approach is used in all major currents of Artificial Intelligence, and it is the one behind its most promising achievements. The focus of ML techniques is on the capacity of algorithms to learn from data without much intervention from theoretical frameworks. This has been extensively debated by people such as Peter Norvig, cited in section 1.2.3, and a popular article coined a new expression to describe this view: “The unreasonable effectiveness of data” (Halevy, Norvig, & Pereira, 2009).

Different types and methods in ML

ML approaches problems according to specific types of tasks: classification (when a system has to learn and estimate the class to associate with a unit, like POS taggers), clustering (when the system needs to identify groupings of items), and regression (when the relationship between inputs and outputs are studied in order to make predictions). The focus of ML is also on decision processes and on search problems.

ML methods may be supervised (when the system is fed with examples and it needs to later identify other similar examples), or unsupervised, (when the system looks for regularities in the data, but with no input on what these regularities might be). Classification tasks are typical supervised methods, and clustering are typical examples of unsupervised methods.

There are different types of ML, according to the form of interaction with users. Active Learning (AL) exists when systems learn from feedback from users, like in Question and Answering automated systems. Online Learning (OL) is applied to systems that receive feedback in sequences and update the learned descriptions continuously. Multi-Task Learning applies to systems learning from two different sources in parallel, while not losing their common features. This is useful, for example, to train an engine with a generic and a specialised corpus simultaneously.

The term “deep learning” is used to describe neural network methods, because of the “hidden layers” that form part of these methods. Christopher Manning comments on the impact that deep learning has had across Computational Linguistics (Manning, 2015). As we may observe in this text, for many, the learning capacity of NMT brings with it renewed dreams of a promised land, with no language barriers. In the next sections, the focus will be on more realistic views of the application of ML to translation.

3.3.1. Different types of learning for SMT

The application of ML techniques to linguistic data is recognised as a difficult task. To translation, even more. However, MT often is presented as a success story in the ML domain and some also say that SMT is a particular type of ML (Cho, 2015b).

ML techniques have been tested in all stages of SMT, from improving the language and translation models to the incorporation of highly evolved data reduction techniques, search scope enhancement, and optimisation methods.

Phrase alignment

One of the challenges SMT meets when it builds the TrM is phrase alignment: at the same text, some of the best alignments may be between bi-grams (units of two words), whereas others may be tri-grams, or even n-grams with more words. And unigrams may align best with bigrams or even trigrams. In order to be usable, the system must find a balance between trying to achieve all the best matches, with different lengths, and the time it takes to search for all these in a big corpus.

To add to this challenge, computer scientists try to add discriminative power to these models, by incorporating corpora with linguistic annotation, be it PoS, syntactic functions, or even semantic information. These annotated corpora could be humanly built or controlled by human intervention, but that is seldom feasible, so system developers resort again to ML to build these, using classification methods but also regression ones. Finally, the models must record all this in efficient forms that enable searches in large spaces, with multiple dimensions characterising each point in the data.

ML has been an invaluable source of methods for researchers to handle these problems.

How to improve systems through ML

According to the "SMT Research Survey Wiki" (Koehn & University of Edinburgh, 2015), publications exploring the potential of ML in SMT became more frequent at around 2009. A book edited by MIT related ML and MT in its revealing title: "Learning Machine Translation" (Goutte, Cancedda, Dymetman, & Foster, 2009). Besides presenting some of the most advanced applications of ML to SMT, the book highlights the need to resort to these advanced techniques in order to deal with the complexity of translation.

In this volume, Jesús Giménez and Lluís Màrquez present an advanced method to select the best matches in the phrase table by using complex discriminative information (Giménez & Màrquez, 2009). Local classifiers are used to inform the system about contextual patterns and they are integrated into a log-linear system. The system is factored, which allows for the incorporation of several levels of grammatical information, such as lemmas, PoS, or syntactic features. The authors comment on the need to manage all features that these models contain, in a task called "feature engineering".

Feature engineering concerns the need to select the most relevant features for the different purposes and texts. This is the theme of a paper by Cristina España-Bonet and Lluís Màrquez that describes the functioning of this task (España-Bonet & Màrquez, 2010). The authors describe how MERT is the method most often used in log-linear models to attribute the right weights to the different features (factors) that accompany the data in the LgM and the TrM. The authors look for a method that does not depend so much on the training set as MERT. They achieve a training method that identifies more robust weights from the training set, in a way that makes it possible to use this training set to create translations for texts in different domains. There are other methods to improve adaptation of the TrMs to new domains. A group of researchers from the University of Lisbon, working in the QTleap project (http://qtleap.eu/) tested the use of three different strategies to improve domain adaptation when there is lack of in-domain data (Stajner, Querido, Rendeiro, Rodrigues, & Branco, 2016). They conclude that for a hybrid (RBMT and SMT) system, the best results are achieved by adding a small in-domain terminology database, whereas for a PBSMT system the best strategy is a combination of a small terminological database and a large out-of-domain corpus.

Word reordering

Another theme that deserves much attention by ML and SMT researchers is word reordering in the TT. A good example of this research is “Exploitation of Machine Learning Techniques in Modelling Phrase Movements for Machine Translation” (Ni, Saunders, Szedmak, & Niranjan, 2010). In this paper, the authors resort to Moses toolkit as a workflow management system, used to define a sequence of programs to be executed over a corpus. They replace the default Moses word-based reordering model by a ML model to solve the reordering problem. This study regards reordering as a classification problem and they call their proposal a “distance phrase reordering model”. This is integrated with the usual learning models (the LgM, a phrase TrM, and a lexicalised reordering model), all of which are processed by a beam search decoder. The new reordering model learns patterns that identify the movement of words in distances of up to 5 positions in the sentence. The features that the classifier uses to learn reordering may simply be the lexical words, or they may include PoS from annotated corpora. The authors claim that their distance phrase reordering model achieves translations which are more fluent and they propose the inclusion of more features to improve the results, or to test other techniques, such as regression.

Research into reordering has not lost its interest, and several papers on the theme were published in 2016. In "Learning local word reorderings for hierarchical phrase-based statistical machine translation" (Zhang et al., 2016), the authors test the use of one single model to learn reordering against several sub-models. Their conclusion is that using several sub-models is more efficient, but that this depends on a specific threshold above which there are no efficiency and quality gains.

This review could be much longer and detailed, but the extent of the application of ML to different tasks in SMT has already been shown. The next wave of MT technology has a strong link with ML, and thus deserves a reference in this section.

3.3.2. Neural networks and Machine Translation

For the time being, the attempts to describe NMT to "mere mortals" are not many. It may even be said that we are at the moment when proponents of a technological advance feel that they do not need to address any other community than their own (see section 1.2.3). Besides, there are many NMT architectures and many ways to employ NNs in MT. This section describes only the main concepts in NMT and it discusses whether NMT may be used in interaction with HT. The publications in Nvidia's blog by Kyunghyun Cho (Cho, 2015a, 2015b, 2015c), the webinars presented by Philipp Koehn for Omniscien Technologies (Koehn, 2016b, 2017), and Koehn's lesson on NMT (Koehn, 2016a) are some of the sources used most frequently in this section.

Neural networks and the processing of textual data

NMT is a natural evolution of the exploration of ML potential into SMT. In fact, NMT is a type of SMT, in the sense that it is a "data-driven" form of MT that depends on the contents of the data that it learns from, and it applies ML mechanisms to reproduce the contents in a second language. However, contrary to what happens in SMT, it trains simultaneously all components, making all decisions from each model concur with the others.

NNs are mathematical functions that, in their simplest form, describe strings of symbols as vectors (which is a form of identifying a set of values of the same type, a sort of “collective noun” with specific mathematical properties), and which may convert those vectors into symbols again.

This is a simple demonstration of how two 5-word sentences are converted into vectors by a simple NN – examples retrieved from Koehn (2006a):

a) but the cute dog jumped
b) but the cute cat jumped

The vector that describes “dog” and “cat” in a) and b) would be:

dog = $(0,0,0,1,0)^T$

cat = $(0,0,0,1,0)^T$

This means that the vector records all words that are associated with the word in that sentence, but only the position of the described word is marked by a “1”. These are called “one-hot vectors”. One of the advantages of vectors is that all positions can be used as dimensions in matrix calculations. Matrices are tables with values in rows and dimensions in columns, which may be the object of calculations. For example, if one multiplies a matrix with another one, we multiply the value in each cell on one table by the value in the same cell in the other table. So, the representation of words as vectors allows for very advanced calculations. Vectors may describe all words in a sentence, in a text, or in a full vocabulary, registering in an economical way their position and the words they co-occur with. Besides, they can represent their PoS and many other features. Finally, vectors are used to describe in a recursive way the inclusion of words in different-sized phrases and sentences.

In the examples above, “cat” and “dog” have the same vector, or representation, with all words surrounding them being the same. This representation does not include any probabilities, but these are always present in a NN. Probabilities describe the relative frequency of the presence of words in different phrases and sentences.

The representation that includes probabilities and feature weights, which allow for the system to choose the most relevant to identify each word and each relation, is called the word’s “embedding”. This term comes from “distributional semantics”, an area that derives from Firth’s “contextual theory of meaning”, enshrined in Firth’s famous expression: *“You shall know a word by the company it keeps!”* (Firth, 1957, p. 11). A word’s embedding relates each word with all the words that it has a strong relation with, in the training corpus.

These embeddings are similar for words that share some kind of relation, such as a morphological derivation (verb inflections, and number and gender inflections) or a semantic relation, such as synonymy. So, these embeddings allow for some kind of clustering of words, which identify the relation between “dog” and “cat” in the examples above and in much more complex and extensive contexts. Because of this, NNs have been seen as a means of revealing deep semantic relations between words, without any need for linguistic information.

The role of neural networks in NMT

It is important to recall that NNs are used in NMT to learn and predict which words are best together, as the output of the system, based on the input, which is a training corpus composed of bilingual data. When the NN runs through the training corpus, it collects all words and n-grams and their embeddings, in extensions that are much higher than was usual in SMT (up to 20-grams in some cases). There are several cycles of learning, in which the system tries to predict which word follows each word, and which phrase follows each phrase in a sequence, thus working out the relations between all units in a sequence.

The most used type of NNs in NMT are Recurrent Neural Networks (RNN). RNNs reuse previous data to predict the next symbol. The data from each word predicted is treated at a hidden layer and is added as input to predict the next word. While the NMT system learns the representations of words, together with their embeddings, it may also learn the relations between these representations in the two languages. This is learnt by a piece of the system that is called the “attention mechanism” (Bahdanau, Cho, & Bengio, 2014), which functions as a form of alignment. One cannot consider it a full alignment process because this is a mathematical process that evolves in each training cycle, and there is no point from which one can extract and review the result of such an alignment. The purpose of the mechanism is to inform the next phases of the learning process.

In NNs, we may know the inputs and the outputs, but we do not have access to the mechanisms that lead to the outputs. This information is treated as a purely mathematical one, in operations with important mathematical properties like non-linearity, which give more flexibility and autonomy to the system, since they are not restrained by homogeneity and additivity (see section 3.2.1 above). The intermediate layers between the input and the output are called “hidden layers”, which keep information to be reused in the different iterations of the learning process. One of the advantages of these methods is that it does away with, or reduces, the need to do feature engineering, because the NN adjusts the weights of each feature automatically. This auto-adjustment of features makes NNs very useful methods to re-score or re-rank hypotheses in SMT models – this is their habitual use as components of SMT systems.

In most NMT setups, two NNs are used, one with the role of "encoding" and the second one for "decoding". The first NN learns a representation of the inputs – source sentences' initial formal descriptions as vectors, with the probabilities and weights of the words. This NN runs until it has learnt all words and n-grams and their relations, which is the moment in which it can predict accurately each word in the sequence. This is when the NN "converges", i.e. it has identified a result which will not change even if there is another cycle of learning. This is when it has achieved the full representation of the training data, a vector that is the result of the "encoding" stage. The second NN will "decode", or estimate the outputs, applying this vector to produce sequences of TL words, phrases and sentences. The decoder is helped by the attention mechanism in choosing not only the most relevant sequences in terms of those that best represent the ST units, but also in building the TT sequence, by predicting the best word or phrase to come after each word or phrase.

Both NNs are trained simultaneously, so that each one feeds on the learning of the other. The training only stops when the system realises that it has achieved the point in which it cannot maximise the prediction of the sequence of symbols.

The training of a NN is a very intensive process. Systems try to capture the elements that allow them to predict complex sequences such as long sentences, or to handle words that do not appear in the training data. One of the strategies is to divide the training into mini-batches, which later must be combined, and to use the longest elements possible. One of the papers that describes a full NMT system reports the following training elements:

> *We used mini-batches of size 40, a maximum sentence length of 100, word embeddings of size 500, and hidden layers of size 1024. We clip the gradient norm to 1.0 [13]. Models were trained with Adadelta [14], reshuffling the training corpus between epochs. The models have been trained for 1.2M iterations (one iteration corresponds to one mini-batch), saving every 30,000 iterations. On our NVidia GTX 1080 this corresponds to roughly 4 epochs and 8 days of training time.*

> *Models with English as their source or target data were later trained for another 1.2M iterations (another 2 epochs, 8 days).* (Junczys-Dowmunt et al., 2016, p. 2).

Just a few notes to explain this excerpt: an "epoch" is a full learning cycle in which all examples are handled. Adadelta (Zeiler, 2012) is a method to mathematically optimize ML functions that depend on updating parameters in iterations. The clipping of the gradient norm is a technique related to the identification and update of those parameters.

In order to deal with unknown words, the main strategy for NNs is to model sub-word units, i.e. to represent words as sequences of characters and to process all tasks with these character sequences. This method is based on the assumption that there are relations between words, such as compounds, loanwords, and cognates, that are revealed by their internal structures (Sennrich, Haddow, & Birch, 2015b). This has led to the development of character-level NMT systems, which aim at being able to find a translation for words that do not appear in the training sets. This should solve, for example, the problem of new inflected verb forms.

NMT captures the coherence that binds units in a sentence better than SMT. Since this coherence is stronger in the sentences of the same language than between two sentences in two different languages, this is the probable cause of NMT being associated with better results in fluency rather than in adequacy (see above section 3.1.4). There have been, nevertheless, specific investments in improving the fluency in NMT by training systems with more monolingual data (Sennrich, Haddow, & Birch, 2015a).

Zero-shot translation and the discovery of an interlingua

One final note on the achievements that have been attributed to NMT. In the same paper in which they present the capacity of NMT to translate in a language pair for which it has no training data (see above "zero-shot translation"), the research team from Google also claims to have found an "interlingua" (Johnson et al., 2016).

The authors explain that the translation between language pairs for which there is no parallel data is made by resorting to other pairs in which each language is present. They call this type of learning "implicitly-learned bridging" – this is opposed to "explicit bridging", when the translation is done in two stages, using a well-known language (such as English) as the "bridge", or "pivot language". The "unknown" language pair that is used to demonstrate "zero-shot translation" in their paper is Portuguese→Spanish. This pair is considered unknown because translations for it are not learnt from training bilingual data with this language pair. Instead, they are obtained after NMT translation models are trained with parallel data from Portuguese↔English and Spanish↔English corpora only. It seems too clear that a system that records with such depth the relations between words, phrases and sentences, in both directions between two language pairs, will most likely contain information that is relevant to relate such close languages as Portuguese and Spanish, especially if the evaluation is based on lexical similarity scores such as BLEU (see 3.5.1).

An interlingua is an intermediate representation of the contents that are moved from one language to another during translation, in a state in which they do not belong to any of the languages. As for the claim that their NMT system had developed an interlingua, the evidence for its existence is a visualisation of the deep hidden layers of the NMT that show that there are stronger clusters of words between the two languages in the unknown language pair than between the language pairs that were trained. These stronger clusters, the authors argue, provide evidence for the claim that NMT has autonomously accessed a deep level of understanding of the relations between languages, a level that had not been specifically trained. They call this deep level an "interlingua". That seems to be a very arguable conclusion, when the graphs show clusters of word embeddings from two very close languages. This proximity is described at the deep layers of a ML system that looks exactly for these proximities, even if the training was done with data that did not anticipate those relations.

Why NMT is not appropriate for the theme of this dissertation

To finish this section, it is important to clarify why NMT is not the elected technology in a dissertation that looks at the most advanced MT systems. The purpose of this dissertation is to find methods for improved interaction between human translators and translation technologies. It may be just because we are at a stage in which researchers are too focused on developing the potential of the technical capacities of NMT to produce translation autonomously, but there are no clues that indicate that NMT will ever be opened for such forms of interaction with translators. In fact, as will be seen in section 3.4, NMT is being employed as a means of replacing human intervention even at the editing stage, and there are still no indications that they may act as supports to interactive systems.

The "hidden layers" and "deep learning" dimensions of NMT are at the core of the advantages researchers claim to have achieved. More than that, these dimensions are defining features of NMT. Furthermore, the amount of data and time NMT training involves can only be considered reasonable if at the end the system produces a high-quality result. So, there seems to be no room for "unproductive" or "inefficient" human interventions that would slow the system down, only to improve somehow the quality of the results.

One can almost say that NMT systems are one step back into black-box MT systems. NMT systems are "end-to-end" trained methods, in which learning cannot be stopped and tuned. So, their ideal use seems to be as central systems, trained on very big volumes of data that provide translations but do not receive regular feedback.

Furthermore, despite all the claims of success made by NMT researchers, it is still not so clear that the quality obtained by NMT systems is consistent across languages, domains, and other text features that SMT has already focused and developed methods to adjust to. It is not clear yet, for example, if NMT is the best method to produce translations that will be later post-edited (see section 3.1.4 above).

The TP enshrined by NMT seems to really have only two stages: learning and decoding. The whole process is sustained by elements such as the encoder, the decoder and the attention system, which process the two stages. The editing stage, which is the one this dissertation focuses on, seems to be (for now?) out of the scope of NMT.

3.4. Editing Machine Translation

In "The present state of research on mechanical translation", Yehoshua Bar-Hillel admits that *"high-accuracy, fully automatic MT is not achievable in the foreseeable future"* (Bar-Hillel, 1951, p. 2). This sentence has been cited and discussed enough, but the following paragraphs of his paper have not deserved the same attention. After this statement, the author turns to "mixed MT", a form in which a "human brain" interacts with the machine. This view of editing as an afterthought, which only merits attention after having been established that the goal of achieving high quality by autonomous machine processes is not attainable, seems to be a natural trend in the history of translation technology.

In SMT, editing is currently recognised as the necessary last stage of a process that does not produce products with the intended quality. There are, nevertheless, two approaches to this problem: this final stage too should be automated, or this stage must be performed by humans. Still, in this second approach, human intervention is seen as having the specific purpose of helping to improve the MT systems. The first approach is known as Automatic Post-editing (APE) and will be analysed in section 3.4.1. The second approach is known as Interactive Machine Translation (IMT) and its analysis appears in section 3.4.2.

3.4.1. Automatic Post-editing

The first references to PE done by automated methods appear as Hybrid MT systems. In different papers, researchers described how they added SMT capacities to correct the results produced by commercial RBMT systems, such as Systran (Dugast, Senellart, & Koehn, 2007, 2009; Lagarda, Alabau, Casacuberta, Silva, & Díaz-de-Liaño, 2009). Because the basis is RBMT, the authors called this addition to their system "Statistical Post-Editing – SPE". However, since other methods may be used with this function, the term "APE – Automatic Post-Editing" is preferred.

SMT researchers also looked for solutions to specific problems in the output of their systems. These solutions are often seen as a two-step SMT architecture. In this setup, after a first stage in which the SMT creates a MT hypothesis by the usual methods, a second training occurs, based on that output, in a "monolingual translation" stage, between the hypotheses and the reference translations in the dataset. The intuition is, that in the second stage, the system will focus on the distances between the MT hypotheses and the reference translations, eliminating the errors that were created in the first stage. So, more than a "monolingual translation", it functions as a "monolingual review".

In some of these experiments, researchers try to maintain the connection to the source sentences, so that the translated suggestions do not move too far away from the ST. The first study applying this methodology reports slight but inconsistent gains in lexical choice and reordering (Béchara, Ma, & Genabith, 2011).

APE systems usually focus on the TT side, on fluency errors visible in the MT output. One of the reasons for this lack of fluency is the lack of examples in the training data, a problem usually described as "data sparsity" and "out-of-vocabulary words". By training the systems a second time, researchers try to identify the connections between these “rare” words and other more frequent ones by manipulating the thresholds in which context information is retrieved.

Another example of work in this domain is “Can Automatic Post-Editing Make MT More Meaningful?” (Parton, Habash, Mckeown, & Iglesias, 2012). In this study, the authors try to move away from the focus on fluency issues and they tackle specific adequacy linguistic errors: missing content words, content words translated as function words and mistranslated named entities. Their system starts by incorporating rules to correct the results of a statistical system. This rule-based APE system only applies two editing actions: replace and insert. To correct the problems with this approach (for example, concerning reordering), they then use an automated system that retranslates whole sentences, incorporating corrections made by the rule-based APE system. The paper explains that this system fixes certain translation solutions (e.g. numbers, names and named entities), but it stresses that the methods to feed the corrections to the decoder must be adapted for each system. The potential of APE to improve the results of MT was also tested on improving domain adaptation: (Potet, Besacier, Blanchon, & Azouzi, 2012; Rubino & Huet, 2012).

Santanu Pal proposed a system called “Statistical Automatic Post-Editing - SAPE” (Pal, 2015) that, based on phrase alignments, identifies four types of adequacy errors: word insertion errors, word deletion errors, lexical errors and word ordering errors. Several methods of alignment are tested and combined, two of which are based on edit distances: TER and METEOR (see section 3.5.1 below).

In “Exploring the Planet of the APEs: a Comparative Study of State-of-the-art Methods for MT Automatic Post-Editing” (Chatterjee, Weller, Negri, & Turchi, 2015), Rajen Chatterjee and colleagues stress that these methods are especially important in the context of black-box systems, as a means to feedback corrections to the system. The authors discuss the two main methods for APE: monolingual, as in Simard’s paper above, in which a system does a sort of monolingual review of the result of a TM system; and context-aware, as in Béchara’s paper, in which the system is reinforced by alignments to the source words and phrases.

In 2015, a shared task (see section 3.5.1) on APE was held for the first time at the Workshop on Statistical Machine Translation (Bojar et al., 2015, pp. 28–36). The main objective of this shared task was to set a standard on how to evaluate APE systems in future competitions. Two of the systems that participated in this competition were SAPE (Pal, 2015) and FBK (Chatterjee et al., 2015). An unexpected result was reported for this experiment: none of the seven systems in the competition was able to improve on the baseline. This result is analysed in terms of the challenges posed by the input data (news texts post-edited by crowdsourcing), with the authors of the report admitting that data sparsity is a problem and that texts with a higher repetition rate may bring better results. They also refer to a variability and inconsistency in crowd-sourced data that will probably not be found in professional PE. The analyses of the types of edits identified by edit distance metrics show that all the systems made more substitutions than shifts. Another effect was visible: systems tend to make all changes they identify, sometimes deteriorating the result rather than improving it. For the next round of evaluation, the authors suggest using professionally-edited in-domain data and multiple references for training.

In 2016, APE methods became more sophisticated, incorporating methods such as NN, log-linear models and factored models for the PE learning task. Some examples of these works are: (Chatterjee, de Souza, Negri, & Turchi, 2016; Junczys-Dowmunt & Grundkiewicz, 2016; Pal, Naskar, Vela, & van Genabith, 2016).

In "Instance Selection for Online Automatic Post-Editing in a Multi-domain Scenario", Rajen Chatterjee and colleagues present an APE system that combines the best of several technologies (Chatterjee, Negri, Turchi, & Kessler, 2016). In this paper, the authors try to solve the problem of over-correction identified in the previous year's shared task. The system starts by selecting from the instances in the training data those that have a higher degree of similarity. When no instances above a certain threshold are identified, the system does not edit the words, so as to avoid the over-correction and the effect of deteriorating the results because of low-quality input data. The system builds local models (that are trained and save only data from each document and session) and global models (that save data from all documents and sessions). These global models are called a "knowledge base", and they are a fundamental feature to allow these systems to be used in real PE production sets, informing and improving the results presented to post-editors when working in CAT environments. This incorporation into CAT environments is announced as the final goal of this system, but this goal is not consistent with the orientation of the paper. If the aim of this system was to assist human post-editors, the system should not be classified as APE, but as a component of an IMT system. Besides, all tests are done with simulated PE operations, which may be appropriate for APE scenarios, but should not be considered in the case of IMT.

3.4.2. Interactive Machine Translation

The last example of an APE system presented above blurred the line between systems that aim at correcting MT output by automatic means and those that are focused on supporting human PE. This section will be a review of technologies that focus on improving the output of MT systems by extracting knowledge from results of PE work made by humans, which is also a category with blurred limits. This category includes systems identified as Interactive Machine Translation – IMT. One of the defining criteria to include an MT system in this category is the fact that these systems retrieve knowledge from post-edited texts, but they feed CAT tools with one MT hypothesis per segment with no further information, as having been generated by black-box systems. The main notion behind IMT is that the translation effort is not done by the human translator, but by the machine (hence, it is not called "interactive human translation").

In 2004, a team of researchers from Edinburgh published a paper called "Improved Statistical Translation Through Editing" (Callison-Burch, Bannard, & Schroeder, 2004). The system they presented had two purposes: to improve the translations generated by the system through the replacement of erroneous phrases according to edits inserted by translators, and to inform the system's user of the origin of the chunks the system presents, so that they could correct them. This second purpose might have meant that this system overtook the restrictive criteria presented above. However, since there are no posterior reports on this system, it is not clear whether it was developed or not.

Completing fuzzy matches with MT

As we have seen in Chapter 2, sub-segment matching and fuzzy match repair were the main concerns that CAT tool developers addressed when designing new tools. When MT researchers started investigating the connections between MT and CAT, their focus was on these problems too.

In "Convergence of Translation Memory and Statistical Machine Translation" (Koehn & Senellart, 2010), the authors present a system that completes fuzzy matches with aligned phrases from the TM. The system is built in an XML frame and the missing bits in the TM are translated by an SMT system. In 2011, a group of researchers of the Universitat d'Alacant used the same principle, but a different method, to predict and signal which words in a target translation hypothesis may be validated and which may need to be edited (Esplà, Sánchez-Martínez, & Forcada, 2011). Their methodology is sustained by evaluation metrics and has some contact points with Quality Estimation of MT, themes that are explored in the next section of this dissertation.

In the same year, a group from DCU applied discriminative learning techniques, inspired by the TM approach, to improve consistency of MT results (Ma, He, Way, & van Genabith, 2011). They call this approach "constrained translation". This approach may also be seen in Yifan He's PhD thesis: "The Integration of Machine Translation and Translation Memory" (He, 2011). The concept of "constrained translation" is applied as a way to select the best aligned phrases from a TM to improve a SMT system (Li, Way, & Liu, 2016). Marine Carpuat and Michel Simard studied the issue of consistency in SMT and confirmed the assumption that SMT systems are very consistent, with levels comparable with those of HT (Carpuat & Simard, 2012). However, they conclude that consistency is not synonymous with correction, since even consistent translations are corrected by post-editors. Mikel Forcada, from the Universitat d'Alacant, proposed in 2014 that the industry standard format for exchange of TMs (TMX) might be extended to include annotations related to the validation of the alignment of sub-segment units, as a means to support "advanced leveraging" – see section 2.3.2 above (Forcada, 2014). All these systems focus on phrases and fuzzy matches in CAT environments.

Interactive Translation Prediction

The notion of "constrained translation" is also behind the main paradigm to build IMT systems: Interactive Translation Prediction (ITP). In this case, the MT system interacts "on-the-fly" with the translator, as he types his translation. A full MT hypothesis can be presented to the translator for him to edit, but most commonly the editing window is empty. The interaction between translator and the MT system starts when the translator types the first character or a word. For the ITP system, this is called a "prefix". Constrained by that prefix, the system does a search for its end, or "suffix".

"Statistical Approaches to Computer-Assisted Translation" (Barrachina et al., 2009) is considered the precursor in this method of assistance to translators. This paper presents a statistical method that starts by building word and phrase alignment templates from the training corpus, together with the other LgMs and reordering models the system learns. Then, the search is combined with a generation algorithm that constructs the sentence from left to right, separating the sequence in different steps, in which all hypotheses are collected and one is chosen. Before interaction with the user, a word graph is created for each sentence, describing paths from each word to the end of the sentence, including alternative words at some of the nodes and alternative routes through them. Each word choice from the user constrains the selection of the next word, choosing a path in the graph.

The authors admit that presenting several alternatives at each node could be useful, but they consider that the usability of such a feature needs to be established, so they suggest that it should be implemented as an option. The paper finishes with remarks on how the system interface must be carefully built, so that users feel that they are commanding the process, instead of the system controlling their work. The authors present a list of features that should be open to tuning by the user to achieve a higher adjustment to his work methods:

> *To name just a few: The maximum length of system hypotheses, the value of n for n-best suggestions, or the "interaction step granularity"; that is, whether the system should react at each user keystroke, or at the end of each complete typed word, or after a sufficiently long typing pause, and so on.* (Barrachina et al., 2009).

ITP is a name given to different technologies that have been implemented over the years in CAT and word processing tools and which may be included under the term "predictive writing" (see sections 2.3 and 2.4 above). In this case, the novelty is the fact that these systems are being used by one type of translation technology (SMT) to support another type of translation technology (CAT). This is the method most currently used by tools that, beyond presenting full MT-produced segments to the translator, offer some kind of support based on the capacities of the SMT technology.

In spite of the proliferation of predictive writing methods in commercial tools, these will not be approached in this dissertation. This proliferation of methods might lead to the idea that this is a solved problem, and that only a few adjustments are needed to the different options available. In fact, if sub-segment support is everything translators need to be added to CAT tools, this seems to be the right way for tool development to go. Nevertheless, this type of interface presents many challenges, not only in terms of usability and definition of the interaction model, but also in terms of the technological architecture required for such a system. Besides, as we will see below, this interface does not respond to all needs. So, it is worth analysing this further.

Predictive Translation Memory

A group of researchers from Stanford University and the University of Washington decided to present a system whose name moves away from the classification of these methods as IMT. This decision to rename this approach may come from the realisation that this name put MT at the centre, but also because they consider that IMT initiatives failed, since they did not attract many users. So, they decided to call their technology "Predictive Translation Memory" (Green, Chuang, Heer, & Manning, 2014). This group of researchers started by investing in good interaction design, and soon they launched a commercially successful software that will be referred to in section 4.3.4. For now, it is important to describe a few details of their proposal. This work reveals a very careful software development project, with users at the centre, and a good analysis of the technical dimension of the TP on which to base each interface detail. The paper includes the description of interactions, in a way that is somehow related to the 3 phases of the TP (see section 2.2).

- To support the stage described as "reading of the ST", the system offers not only word translation suggestions aligned for each source word, but it also marks the words that are already translated as the translator moves ahead in the sentence.
- For the "target gisting" stage, there is a full-sentence suggestion that is updated in real-time according to what the translator types.
- As for the final stage, which they call "target generation", the user interacts with the suggestions in three ways:
    - the user accepts the autocomplete suggestion for the word he started typing;
    - the user types a word he sees in the suggestion, but in a different order, anticipating words that appear further ahead in the sentence;
    - or he accepts the whole suggestion.

These three forms of interaction have a resemblance to the four editing actions that were proposed. The authors consider, for example, that the second interaction is a form of reordering that allows the user to release from the "left-to-right" constraint. However, in reality, the user still writes words in their correct order, from left to right.

In the evaluation of their system, the authors asked translators for feedback on different types of interaction with their system. Users preferred the autocomplete feature, then the possibility to see the target full sentence, and finally the translation lookup feature based on selection of a source word. The feature that most users criticised was the word reordering feature, and the authors admit that this is a difficult action to implement in an interactive interface. The authors also compared their interactive system against a "baseline" PE situation, in which users only had one full MT hypothesis to edit. One of the conclusions is that interactive systems allow users to move farther away from the ST, but they achieve higher quality with this option. The reason for them to move away from the ST may have to do with the presentation of alternatives, but it has as an undesirable consequence the fact that this will reduce the parallelism between ST and TT.

Alternative interfaces

ITP interfaces are still a research line that garners a lot of interest, but let us analyse other interface options for editing. A team from the Universidad Politècnica de València (Domingo, Peris, & Casacuberta, 2016) proposed a system that moves away from the left-to-right editing paradigm and allows users to select, remove, replace and insert words or phrases into the translation suggestion. The authors measured the editing effort (number of keystrokes) and cognitive effort (number of mouse actions per character typed) and concluded that the system reduces the physical effort but increases the cognitive one. However, they did this evaluation in a simulated scenario, which means that all measures are estimates. In a paper published by researchers of the same university, the same evaluation method was used in a similar environment (González-Rubio, Ortiz-Martínez, Casacuberta, & Benedí, 2016). The authors devised a system that allows users to edit freely any part of the translation suggestion, and the system adapts the suggestion to these edits. In an illustrated example, a process that would require 10 edits in a regular PE environment is achievable in this system with the user only making 2 edits.

The application of NNs to this task has also been attempted, but only as methods to learn the post-edits in a more efficient way and ranking them in black-box ITP systems (with no user feedback or interaction interface) (Torregrosa, Pérez-Ortiz, & Forcada, 2016; Wuebker, Green, DeNero, Hasan, & Luong, 2016).

3.4.3. Online Learning as a method to deal with repeated errors

Behind the scenes of all these IMT systems, there are ML technologies that are capable of learning continuously from users' inputs, so as to adapt and present updated suggestions in real-time. Online Learning (OL) is the ML technology most consistently used for this task.

The most comprehensive document on the application of OL in this context is: "Online Learning for Statistical Machine Translation" (Ortiz-Martínez, 2016). The author starts by explaining that this approach allows SMT to solve the problem of batch retraining. In this study and in Green's "Predictive Translation Memory", IMT systems are compared against baseline PE scenarios. The term PE is used to describe the situations in which the user only has access to one full-sentence MT suggestion, whereas IMT describes applications in which the user sees not only a full hypothesis, but also alternatives for the words he types.

Internal repetition rates

Ortiz-Martínez mentions that the effectiveness of OL methods depends on the "internal repetition rate" of the ST. Other researchers have addressed this theme, and specific rates have been used to measure this.

In "Online adaptation to post-edits for phrase-based statistical machine translation" (Bertoldi et al., 2014), the authors comment on two often neglected text features in SMT: text consistency and repetitiveness. They suggest a metric for the repetition rate, which is based on the rate of "non-singleton n-gram types", i.e. the compared rate of repetition of groups of more than one word. They also suggest that this data is inserted into a local cache-based model, built from user feedback, that informs the OL process. The authors explain that their system applies OL in the framework of "online adaptation", in which several models of a complex system are adjusted from different inputs, weighing newer inputs heavier than older ones. The system adaptation involves both "discriminative learning", whose function is essentially re-ranking the decoder's outputs, and "generative learning", which integrates the new models' information into the decoder. The generative module allows the system to learn the local models that receive the data from the user's actions. The weight of the local and the global models may be adjusted to different purposes. All these concepts are tested in simulated environments.

Ortiz-Martínez (2016) adjusted Bertoldi's repetition rate in two new metrics, (modified and restricted) and adds a new one: "unseen n-gram fraction". The purpose is to use the ratio of words and n-grams that appear in the test set, and which did not appear in the training set to relativize the weight of simple repetitions of words and n-grams. The author then shows the seven feature functions (LgMs, length, reordering and alignment models) that are used by his log-linear TrM. The alignment is made in two directions in each sentence, and the EM algorithm is applied in a novel incremental way. This algorithm may be broken into two different steps (expectation, then maximisation), which makes it ideal for "incremental training" – this is an alternative name for OL. The author shows the advantages of such a method, obtaining more efficiency than batch learning and any other OL method. He also shows how this may be an effective method to tackle the problem of using out-of-domain training data in in-domain translation tasks, thanks to the incremental learning capabilities demonstrated.

Other approaches to repeated errors

Two innovative approaches to IMT are connected to this overview of Ortiz-Martínez's article. One is "PEPr" (Simard & Foster, 2013) a method for the propagation of post-edits in a CAT scenario with a SMT system in the background. The system is based on the notion of the internal repetition of a text, OL capabilities, consistency of SMT systems and on APE, which the authors include as a feature of interaction with users, and not as an autonomous system. The system deals with the problem of "over-correction" by admitting that all words may be left unedited, using this as a means to reduce the interference of an autonomous system by the user of the system. The main limitation of this system is the fact that it only handles local updates, not being adjustable to other documents.

The second approach to IMT that is worth mentioning tackles this limitation. In "Translation project adaptation for MT-enhanced computer assisted translation" (Cettolo et al., 2014), the introduction of the concept of "project adaptation" is a suggestion of how to extend locally-learnt knowledge (or editing behaviour) to other documents, as part of the same project. This is an approach that bears some resemblances to "domain adaptation", but which applies concepts and strategies that are closer to the realities of professional translators and translation companies. However, this is another example of an article in which simulations of editing were employed. The authors challenge the global assumption that adding more data to an SMT system always improves the output, relativizing this with the relatedness of the added data to the test documents. The results show that different domains yield different productivity gains. But the authors suggest that other variables may affect the results, such as the learning curve that depends on the user's familiarisation with the system, or the different and unevenly distributed difficulties that texts pose.

3.4.4. Outtakes from approaches to editing in SMT

We have seen above how the final stage of editing is approached by SMT: either as a process that may be learnt offline and applied automatically to improve the results of SMT, or in IMT. Often, although researchers commit to the purpose of focusing on human translators' work, the systems that are developed struggle with the models of how translators actually work. We have even seen how PE is considered a totally different thing to IMT. Either way, if not all, most of the papers published in this area only work with simulated models of actual editing work, by extracting inferences from edited results of testing corpora.

Besides, research in this area, with only a few exceptions, models translator's work as left-to-right typing. No wonder then that predictive writing is the paradigm for tools developed in this context. However, those that look for alternatives and try to move away from this constraint struggle with interface implementation for editing actions, so it seems that there is still work to do in this field.

The application of OL to SMT is a very important step towards building interactive tools that help translators take advantage of the new resources and capacities enabled by SMT. This seems to be a technology that is capable of balancing sophisticated technologies that provide translation suggestions with efficient interfaces that offer them to translators. In Chapter 4, after the literature on PE is discussed, the approaches to editing will be reanalysed, from a renewed point of view.

3.5. Evaluating and estimating Machine Translation

Most of the studies in the previous section depend on some sort of evaluation, not only to confirm the validity of the results obtained, but also to filter the best solutions that are fed to the system. The last section of this chapter is dedicated to the initiatives to evaluate MT products, by manual or automatic methods, zooming in on how these methods are being used to tap into the editing process.

3.5.1. Evaluation metrics

The term "evaluation" is used very often in this and other works on translation and MT. However, it is important to note that the interest in this dissertation is not on notions of quality or quality classification, but on how these techniques may be employed to estimate editing. Important references for this section are: Chapter 5 of DARPA's "Handbook of Natural Language Processing and Machine Translation" (Dorr, 2010), "Machine Translation Evaluation: A Survey" (Han & Wong, 2016) and the tutorial "Evaluating the Output of Machine Translation Systems" (Lavie, 2011).

In several areas of technological studies, there are competitions and "shared tasks" that regularly allow teams to compare results of their research, as a way to move each area of knowledge forward. Shared tasks are proposals for each team to put their systems to the test by using the same datasets, the same objectives, and the same evaluation rules to perform the same task. Results are presented and discussed annually at MT conferences, which gives developers a great opportunity to check alternative ways of setting up their systems. Since 2008, there has been an annual shared task on evaluation metrics applied to MT. Since 2012, there has been one on Quality Estimation, and since 2015 one on Automatic Post-Editing. This has led to a large number of publications on these themes, and to a rapid evolution of the areas. In the United States, NIST (National Institute of Standards and Technology) has maintained evaluation projects since 2001. DARPA (Defense Advanced Research Projects Agency) was one of the institutions funding research in this area, namely in the GALE (Global Autonomous Language Exploitation) project, which ran from 2006 to 2011. These institutions also hold MT evaluation challenges such as OpenMT (https://www.nist.gov/itl/iad/mig/open-machine-translation-evaluation).

To assess the quality of the products of MT, several metrics have been developed, focusing on different items and levels of what was considered "quality". Some of these methods are adapted to implementation in automated or simulated environments, and only a few are associated with evaluations made by humans. Human evaluations are usually called "human judgement" and are often used as standards, or baseline data for automatic evaluations – we often see results claiming that a system has achieved "high correlation with human judgement". In human evaluations, translators, or, very often, bilingual or even monolingual users, are asked to rate MT sentences according to different assessment scores. These may be simple "Likert scales" (Likert, 1932), but in most cases some scale around two concepts is used. The ALPAC report (Pierce et al., 1966) presented an evaluation experiment with the objective to create a standard for such a task. In this experiment, the evaluators should classify several translations in terms of "fidelity" and "intelligibility". This dichotomy survives to this day under the terms "adequacy" and "fluency". Adequacy is the coverage of the meaning of the ST by the TT. The use of the term "adequacy", with its meaning being associated with quantity, highlights a quantitative approach to translation, based on the notions of "completeness", "similarity" or "equivalence". Fluency is a classification of the degree of grammaticality of the target sentences. This duality resonates with the dual models at the basis of SMT systems: adequacy is the purpose of the TrM and fluency the purpose of the LgM. There are, nevertheless, other scales and terms, some based on rankings of translation proposals, classification of types of errors and assessment of PE effort, that human evaluators are asked to employ.

Human judgement evaluations have been criticised not only because they are expensive and time-consuming, but also because the classifications and scores vary a lot across evaluators. Human evaluation has been discussed in many papers and studies, with researchers discussing the profile of the evaluators, the type of tools used and the scales of errors used. Two examples of these studies are: (Maia, 2008) and (Popovic, Avramidis, et al., 2014).

One of the ways to classify an MT hypothesis without the need to resort to human evaluation is to measure how different it is from a reference translation, like the "gold standard" used in training and test environments. There are several ways to measure this difference, some by lexical similarity, others by computing the "edit distance".

Lexical similarity metrics

Lexical similarity measures employ the concepts of precision (the percentage of correct answers over the total answers by the system) and recall (the percentage of answers the system got correct over the global number of possible correct answers). These metrics take the same n-gram approach used by the learning models in most SMT tools. There are two main advantages to this method: they may be easily integrated into SMT tools, and they are widely available, being used as the main evaluation metrics by the SMT community.

The best-known lexical similarity metric is BLEU – BiLingual Evaluation Understudy (Papineni, Roukos, Ward, & Zhu, 2002). BLEU compares all the words in the MT hypothesis with the words in several references and gives a precision score to the MT system. The precision score is erroneously described as a simple calculation of the number of all correct words that appear in the hypothesis, divided by the length (total number of words) in the hypothesis. However, BLEU is a bit more complicated. The main idea is that BLEU focuses on adequacy, and it compares an MT hypothesis not just against one reference, but several.

BLEU starts by looking at the “word precision” that a hypothesis has, in terms of the number of words and n-grams (up to 4-grams) that match the hypothesis against the same sentence in different references. (The higher n-grams are used as a means to also account for fluency.) Then, a geometric average of all these values is calculated. To avoid a repeated word being repeatedly counted, words that are repeated more than twice are discounted. Finally, a “brevity penalty” is added to cancel the fact that shorter sentences (even incomplete ones) resulted, by default, in high precision. This penalty also intends to compensate for the fact that the system has no way to account for recall, which is not possible to calculate because there are several references.

This approach to evaluating the adequacy of a translation was said to achieve a high correlation with human judgement, and that is one of the reasons why it is still accepted as a standard, but this strategy has been criticised for its limitation (Callison-Burch, Osborne, & Koehn, 2006). The main criticism of BLEU is that the reliance on its correlation to human judgement is exaggerated, since there are several reasons why humans score differently sentences that receive the same BLEU score. One of the reasons is that BLEU only matches words in the exact same form, which means that it does not match inflected forms or synonyms, whereas these changes are considered innocuous by humans. Besides, BLEU is not sensitive to changes in word order, so it allows for ungrammatical sentences to achieve high scores. And finally, it gives all words the same weight, which implies that if an important content word is missing the sentence receives the same score as if it were an optional function word.

In the same year, the alternative NIST metric was presented (Doddington, 2002). This metric uses a different way to estimate the mean (arithmetic instead of geometric) and it includes a weight for more informative (or rare) units. In “Evaluation of Machine Translation and its Evaluation” (Turian, Shen, & Melamed, 2003), a new method is proposed, based on the “F-measure”. This measure combines precision and recall, and it is extensively used in evaluating software solutions for classification tasks in NLP, more specifically in Information Retrieval (Manning & Schütze, 1999; Resnik & Lin, 2010).

In the paper quoted above as the main source of criticism on BLEU, the authors refer to the fact that all n-gram metrics suffer from some of these problems, and that this should be taken into account and checked against the intended use of each metric. They stress that these metrics should not be used to compare systems with different technologies, and underline the fact that *"they are all quite rough measures of translation similarity, and have inexact models of allowable variation in translation"* (Callison-Burch et al., 2006, p. 8).

After looking closely at the weaknesses of BLEU and NIST and the linguistic factors at play in translation, Alon Lavie and his team proposed a new metric (Lavie, Sagae, & Jayaraman, 2004). METEOR (Metric for Evaluation of Translation with Explicit word Ordering) accounts for word precision and for word recall. However, recall is given a higher weight in the calculations, in order to achieve a higher correlation with human judgement. This metric has also proved to be able to produce better results at sentence level than BLEU. METEOR was later enhanced and it now takes into account other linguistic features. It solves the problem of multiple references by matching each reference separately and then selecting the best match for each sentence (Lavie & Agarwal, 2007). The linguistic features that the metric incorporates include synonyms, lemmas and paraphrases to help increase the unigram matching. Besides, it is sensitive to the different positions words occupy in the sentence, applying a reordering penalty to capture fluency. An added advantage of METEOR is the fact that it is tuneable for features such as precision, recall and word reordering. However, language-dependent features imply that evaluation systems are fed with data from different languages. This problem is solved by METEOR Universal, a version in which the language-dependent features are extracted from bilingual data (Denkowski & Lavie, 2014).

Edit distance metrics

Edit distance metrics are based on the principles of "edit distance". This is an estimation of the number of edits, or changes, that are required to transform one string into another. Here, "string" means a sequence of characters, but edit distance estimates are used in several fields outside NLP, such as biology and computation. The first proposals for the estimation of edit distances were in the 1960's. There are two metrics that are known by "Levenshtein's distance". One was proposed by Vladimir Levenshtein (1966), in Russia, and the other by Frederick Damerau (1964), in the United States. The latter's most usual name is "Damerau-Levenshtein", but Levenshtein's name alone is used to refer to this one too.

The metric by Levenshtein is the most adopted and the simplest. Levenshtein wrote his paper in the context of writing and correcting computer codes, for the transmission of information in binary form. The code correcting program compared a string of characters and for each misalignment it estimates whether a character was deleted, if it was inserted, or if it was "reversed", or swapped. By reinserting the missing character, deleting the extra character and inverting the swap, the proposed program could correct these errors in the code.

Damerau's paper is concerned with the correction of spelling mistakes in information retrieval systems. He identifies four types of errors:

> *An inspection of those items rejected because of spelling errors showed that over 80 percent fell into one of four classes of single error – one letter was wrong, or one letter was missing, or an extra letter had been inserted, or two adjacent characters had been transposed. These are the errors one would expect as a result of misreading, hitting a key twice, or letting the eye move faster than the hand.* (Damerau, 1964, p. 1)

This paper was written at a time when interaction with computers was based on punch cards. The spell-checking process was based on matching words, character by character, against a dictionary list. The following figure, retrieved from Damerau's paper, shows how each error is identified.

ALPHIBET ALPHABET	Words are same length and differ in only one position. They are the same word.
ALHPABET ALPHABET	Words are the same length and differ in two adjacent positions. If these are interchanged, the words match.
ALPHABET ALLPHABET ALPHABET ALPHABET	The entry word is one character longer. The first difference character is discarded. The characters following it are shifted left and the two words match.
ALPHABET ALPABET ALPABET	The entry word is one character shorter. The first difference character of the dictionary word is discarded and the characters following are shifted left. The two words now match.

Figure 5 – Spelling error identification sequence (Damerau, 1964, p.2).

In this sequence, character replacement is identified first, next transposition, then insertion, and finally deletion. The system can only deal with words that present one single error. The other 20% of errors identified in the texts are supposed to be either words with multiple errors, or simply words that do not appear in the dictionary. The method focuses on words that have more than 5 characters and it relies on the correct identification of the end of the word based on spaces and punctuation. In one of the methods tested, abbreviations are used for efficiency reasons (based on the assumption that some letters are more likely to be misspelled than others), and in this case the first and last letters are always kept.

Although this description of Damerau's method may sound too detailed, it will be necessary to interpret part of the results described in Chapter 6 below. As for terminological choices, the words "replacement" and "substitution" are used interchangeably in this dissertation (replacement being the preferred one, but substitution is often used by researchers). "Transposition" usually describes a change of position of two contiguous characters, and "shift" is used for more ample changes of position of one character. Because of the specific use of the term "shift" in TS, it was considered that the word "movement" is preferable to describe this editing action.

Edit distance metrics were created to identify and correct errors, so "word-error rate" (WER) is also a common designation for these. The purpose of these calculations was to find the "minimum distance" between the two strings, "minimum" meaning the least number of operations required to transform one string into the other. This purpose is purely methodological: imposing efficiency requirements on metrics is a way of restricting them. If there is no such restriction, calculations are endless.

Edit distance applied to translation

In "A Study of Translation Edit Rate with Targeted Human Annotation", Matthew Snover and colleagues propose a metric adapted to MT based on edit distance. The name of this metric is "TER – Translation Edit Rate" (Matthew Snover, Dorr, Schwartz, Micciulla, & Makhoul, 2006). In a footnote, the authors explain that the metric should not be confounded as "error rate", because the focus is on the edits, the operations needed to transform one hypothesis into its reference: *"Possible edits include the insertion, deletion, and substitution of single words as well as shifts of word sequences."* (Matthew Snover et al., 2006, p. 3).

The authors propose the term HTER – Human-Targeted Edit Rate for the situations in which the reference is no longer a gold standard taken from parallel corpora, but post-edited versions of MT texts. In this case, the purpose of the task is extended to trying to identify the edits made by the post-editor.

In TER, all edited words have the same cost (they count as one), including shift. The sum of these costs is then normalized (i.e. divided by the average number of words in all references). The edits are identified in two different phases: First, insertions, deletions and substitutions are calculated by an alignment function. Then, a greedy search method is used to estimate the shifts that reduce the number of insertions, deletions and substitutions, i.e. it verifies if moving a word to a new position eliminates what had been considered as two edits (a word missing and a word inserted). So, shifts are only calculated as a means to reduce the total number of edits.

This separation of deletion, insertion, and replacement for one side, and movement for another may be seen under the light of the paradigmatic/syntagmatic debate, mentioned by Anthony Pym in section 2.3.5. Alignment-based edits only deal with inserting the right unit in the right slot, according to a paradigm established with the ST. Movement, however, implies a disruption in that paradigmatic slot-filling rationale, as it implies an estimate in the syntagmatic level of the TT sentence. The rules and constraints that apply to the paradigmatic approach are no longer applicable to determine how movement is done.

The authors explain that calculating all edit distances with movement is a very complex problem. In effect, a previous paper had demonstrated that this was another search problem that could be classified as NP-complete (Shapira & Storer, 2001). So, the method applied in TER is a “greedy search” constrained by several rules so to keep the process under control. This greedy search starts by selecting sequences of words (it is not clear how long these sequences of words are) that do not match the sequence of the words in the sentence; if moving this sequence achieves the intended effect, the search stops; otherwise, the same procedure is tested with single words. This means that shift estimation privileges phrase-level edits over word-level edits. Snover and colleagues explain that this extra work compensates because this is a better model of quality, and it achieves a higher correlation with human judgement. However, they hypothesize that shifts might have different weights/costs according to length and distance.

In the experiments with TER, Matthew Snover's team explains that *"Annotators were coached on how to minimise the edit rate, while preserving the meaning of the reference translations."* (Matthew Snover et al., 2006, p. 5). The annotators that participated in this experiment were monolingual, but two annotation sessions were necessary to guarantee accuracy and fluency. The authors say that in the experiments most edits were substitutions and deletions. Shifts are usually one-word and the maximum distance moved was 7 positions. And they conclude that there is more correlation between automatic metrics and average edit rates than between human annotators against each other. A final advantage they highlight of TER is that it is easy to understand by people outside the MT community.

In 2008, the same team presented an extension to TER called TERplus, or simply TERp (Snover, Madnani, Dorr, & Schwartz, 2008). This extension gave more flexibility to the matching system and introduced tuneable weights, so that every edit is not always attached to the same cost. The method to estimate edits relies on a matching process, which is extended by allowing matches between derivations of the initial word (identified through stemming), synonyms (identified by a dictionary of synonyms) or a paraphrase (identified by a vocabulary of over 14 million paraphrases in English). These paraphrases may involve phrases (in which case, they are called "phrase substitutions"). The paper considers that these three extra features are new edits, when in fact they are modifiers of the already existing edit actions. Deletion and insertion are not modified. Substitution and shift are now allowed between words that are not matched, but which share a stemming, synonymy or paraphrase relation. Only phrase substitutions' weights have been modified, with a formula that involves the probability of occurrence of the phrase that is shifted and the number of edits necessary to transform the sentence. The authors looked for a correlation between edits and increased quality. So, they restricted, for example, shifts to movement of content words. TERp is also case insensitive, because the authors concluded that case-sensitiveness would reduce correlation to human judgement on this increase in quality. Besides, TERp has a cap on a total cost of 1 (or 100%), which TER did not have. This means that if the number of edits is higher than the number of words, TER captures this and TERp does not.

In posterior papers (Snover, Madnani, Dorr, & Schwartz, 2009), the authors mention that TERp achieves a higher correlation with human judgements on quality, especially thanks to the optimization of costs per type of edit, according to criteria such as adequacy or fluency. This implies the use of a hill-climbing search algorithm, which dynamically updates the costs of each type of edit, to optimize the parameters and achieve a better correlation with human judgement. The downside of this option is that TERp becomes much more difficult to implement than TER. This possibility to adjust the costs of each edit makes this metric useful for tuning SMT systems to correct biases towards adequacy or fluency, or other criteria that depend on the type of edits. Besides, this metric (and others) have been used as alignment components in other systems, in training modules, due to its capacity to identify words that are matched, missing, or which are misplaced.

3.5.2. From evaluation to quality estimation

Quality evaluation of MT has brought with it several discussions that have a direct impact on the development of systems and interaction with users. Confidence Estimation (CE) tries to predict the level of quality produced by MT. There are several dimensions in this line of research that are worth analysing, particularly in relation to the factors and criteria that are used to make the predictions, and how this can be applied as an aid to editing work.

In “Improving the Confidence of Machine Translation Quality Estimates” (Specia, Turchi, Wang, Shawe-Taylor, & Saunders, 2009) the authors explain that CE is essentially different from automatic MT evaluation because in CE there are no references to learn from. The ultimate aim of CE is to develop systems that MT users may use to estimate what level of quality a certain MT system will achieve for a certain document. One of the uses of such a system is to filter out MT sentences that fall below a defined quality score threshold that separates outputs that can be post-edited within a reasonable timeframe, and those that cannot. To arrive at that result, all research in CE is targeted at identifying the factors that determine the quality of the results of a MT system, i.e. the features that determine the degree of confidence of that MT system in that context. These features need to be independent of system, language and any other resource. In this paper, the authors train a system with features such as n-gram occurrence, different probabilities, type/token ratios, length, and many other “shallow” features. After the system has learnt the features and classifications of good/bad from human annotators, it tries to predict equivalent annotations for the test set. The results reported were positive, but they are presented in a context of a clear need for improvement.

In 2010, Lúcia Specia and colleagues adopted the more comprehensive name "Quality Estimation of Machine Translation" – QEMT, or simply QE, for this task (Specia, Raj, & Turchi, 2010). In this paper, they explain that QE may be used not only to filter translation hypotheses, but also to select the best hypotheses, from multiple MT systems. A few years later, the same team of researchers launched a toolkit (QuEST) to automate the collection of features (or indicators) that help determine different levels of QE (Specia, Shah, de Souza, & Cohn, 2013). At this time, the sophistication of QE systems had grown, and "confidence" was just one of the dimensions around which the features were organised. The following figure, taken from this paper, sums up the global organisation of indicators that allow researchers to automatically classify and predict the levels of quality produced by MT systems:

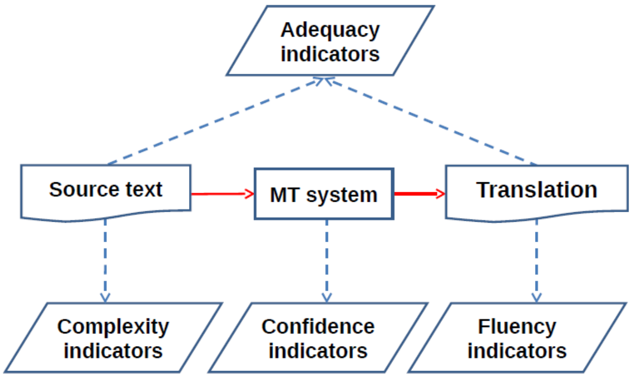


Figure 6 – Quality indicator framework used in QE (Specia et al, 2013, p.3).

The figure above shows how "confidence" is only a part of the set of indicators that may help predict the quality of a MT system. Below is a summary of these indicators, some of which may be collected directly from training sets, while others may imply linguistic analysis tools.

- **Adequacy** indicators – these are related to the alignment of ST and TT; they include type/token ratios, percentages of numbers and punctuation tokens, percentages of PoS, depth of syntactic trees, and many other features, always collected from both texts.
- **Complexity** indicators – these are concerned with the ST only, and include number of tokens, LgM probabilities, percentages of n-grams, etc.
- **Confidence** indicators – these indicators are retrieved from information on the MT system, from phrase tables, activity logs, lists of n-best

values, etc.; they include features such as scores given to MT system, number of distinct hypotheses, average size of target phrases, etc.

- **Fluency** indicators – these are similar to adequacy indicators, but they focus on the TT only.

With such a complex set of features to train and test (in some cases, systems extract hundreds of features), feature engineering and ML techniques to optimise training and selecting them are central preoccupations for QE researchers (Callison-Burch et al., 2012). However, a set of 17 features is considered an adequate baseline for most tests. In recent years, shared tasks helped research move forward at a fast pace.

In 2016, a report was presented at the LREC conference with the conclusions of 10 years of evaluation campaigns (Bojar, Federmann, et al., 2016). One of the shared tasks commented on is the one on QE. The authors describe the progression from estimation done at the system-level, to sentence-level, paragraph-level, word-level, and, only this year, phrase-level QE. All these levels imply different approaches, have different goals and require different features. Another conclusion of the report was that, although linguistic features have been tested, shallow features are the ones that seem to achieve the best results.

Phrase-level QE

Varvara Logacheva and Lúcia Specia published the first proposal for phrase-level QE, a very detailed account of the technical difficulties in setting up a learning activity that reveals much of the importance of phrases in SMT (Logacheva & Specia, 2015). They comment on the relation between editing errors and phrases – errors can seldom be captured at the word level, since they are related to word contexts. The authors hint at these distributed errors being "generated" by other errors, but that description would only make sense for RBMT systems. In PBSMT, there are no heads in phrases that command what happens around them. But the challenges appear right at the start of the task: how to segment phrases from parallel corpora? The authors discard linguistically motivated approaches and focus on the hints provided by the MT decoder, namely in the contents of the resulting phrase tables. They discuss the difficulties in doing this, because of problems in the alignments, discontinuity of phrases, and related issues. This is associated with problems in quality labelling for training: datasets used to train the estimators have quality annotations at the word level. Generalising from words to phrases, although apparently intuitive, is not such a straightforward process. The authors discuss this relation between words and phrases throughout the paper, but at the end they comment on an opposing approach: phrases may be seen as compacted sentences, so perhaps some of the processes applied to sentences are better for phrases than those that are applied to words. For linguists that have studied the complex structures of NPs, especially in the context of translation – see, for example, (Carmo, 1998) – this view comes as no surprise.

The selection of the appropriate set of features that allows the identification of quality in phrases is another challenge. The authors check the appropriateness of sentence-level and word-level features, and add a new set of features derived from the transformation of phrases into vectors, as NN does. In spite of all these difficulties, the authors conclude that in their tests the systems that were trained at phrase-level are capable of predicting simple quality labels (bad/good) at phrase-level better than word-level trained systems.

The report that presented the detailed conclusions about the shared tasks performed at the WMT workshop in 2016 has a specific section on QE at the phrase level (Bojar, Chatterjee, et al., 2016). The section on features describes the new set of phrase-level features, which includes a few that derive from analyses of alignments, such as number of unaligned target words; number of target words aligned to more than one word; and average number of alignments per word in the target phrase. The report comments on the difficulties all systems had with this task, and how the evaluation and methodology of the task is still being discussed. However, they conclude that systems that incorporate NNs as learning mechanisms are the ones achieving the best results.

Varvara Logacheva and Lúcia Specia joined Frédéric Blain in a paper that describes the evolution of this field after a year (Blain, Logacheva, & Specia, 2016). In order to solve the issues of segmentation and labelling that had been previously identified, the authors decided to use TER as one of the methods to select the bad/good quality labels for the phrases. They use this method side-by-side with a shallow syntactic analysis and the usual decoder methods, and they explain that one of the aims of a system that successfully completed this task could be as a module that provided phrase-level annotated QE data as input to a decoder. In this case, we would be in an APE scenario. However, all these approaches have issues in terms of covering all examples in the training set. Another dimension discussed is bias towards the training set, which might make these methods not applicable with test sets from a different domain.

The authors of this article asked translators to manually annotate the training set with quality classifications. The guidelines these annotators received concerned typical preoccupations in terms of linguistic annotation: where to segment, what to select, how to approach overlaps, positioning and continuity of the units. This was especially complex when annotators were trying to identify missing, inserted, moved or replaced phrases in different resources. They were presented with several texts: not just the ST, the MT hypothesis and the post-edited versions, but also other reference translations for the same texts. The results indicate that the phrase quality labels that were extracted with the support of TER were the ones that most closely correlated with the human annotations. The lack of precision and recall between the number of errors identified by human annotators and those that were actually corrected in the originally post-edited versions is attributed to a difference in the resources that were used: while initial post-editors only had access to the ST and MT hypothesis, human annotators included errors that arose from their comparison to alternatives in the reference translations. The authors concluded that all dimensions of this task need to be discussed and refined.

3.5.3. From quality estimation to editing estimation

There is a close relationship between edit distance metrics and QE, so let us zoom in on that relationship, as it reveals details about editing which will be useful for the rest of the dissertation. As we have seen up to here, quality metrics often focus on errors committed by MT systems. However, the questions of what an error is and how to identify it are not easy to answer.

Most papers that use metrics such as TER rely blindly on its capacity to identify words that are missing, or that have been inserted, replaced or moved. Then, they add a notion of effort related to the necessary number of edits, and a proportion with the time it takes to correct them, and consider that this identification of errors will help PE work. However, even issues such as segmenting and identifying the units these errors belong to are not so straightforward.

Post-editing action units

A group of researchers from different institutions proposed in 2011 a *“new unit for evaluating post-editing effort based on Post-Editing Action (PEA)”* (Blain, Senellart, Schwenk, Plitt, & Roturier, 2011). The purpose of this paper is to move forward from the notion of errors and to help correct them. The unit they present is based on the TER actions, the same as the "editing actions" used in this dissertation, but the authors state that these units are not useful enough because they only refer to "mechanical actions". They consider that the analysis of the editing process should be extended to include the "logical edits", as retrieved from post-edited material. In order to achieve that, they take PEA as a concept that helps structure the task of "learning post-editing".

The logical side of PEAs derives from the linguistic reasoning behind each editing decision, and which describes the connection between related edits in different words. An example of these connections is the gender modification of the head of a NP that "propagates" to the rest of the NP. The authors claim several advantages to this approach, linked to the assumption that this is more intuitive for post-editors and that such modifications may be learnt and applied in a single action. However, they admit that these are much more difficult to extract by automatic processes. The PEA typology includes typical linguistic units, such as NPs and VPs, editing actions applied to specific linguistic units (such as replacement of prepositions and referential pronouns), reordering, "PE errors", "unnecessary style" changes and a category for "misc". Edits to NPs and VPs are subdivided into different types of replacement actions. The tests performed show that 90% of the edits are classified as NPs, with a reasonable number of terminological changes. These classes of errors should be either manually annotated or automatically learnt, and then added as annotations to the data.

The authors of this study admit that their method depends on the repetition rates and parallelism of the data. They admit, for example, that the regular use of synonyms in specific text types might hurt the capacity of the system to learn. They tested the process in an RBMT and in an SMT system and report similar degrees of success. However, they stress that RBMT modification to this data is easier than in SMT. There is no reference to posterior developments of this approach.

Lack of correlation between edit distance measures and effort

Michael Denkowski and Alon Lavie have published a careful reflection on the difficulties in implementing MT systems that are useful for post-editors (Denkowski & Lavie, 2012). They comment on the failure of metrics to capture the effort required to post-edit a MT text. They challenge the notion of adequacy as the most important aim of the evaluation, claiming that the "acceptability" of the output revealed by this notion does not imply that the output has "utility" as an intermediate step in the TP. They then show examples of two opposite situations: when low error rates correspond to more editing effort (such as when the meaning is correctly translated, but there are so many fluency and stylistic details that correcting it becomes too cumbersome); and when there is a problem with a named entity that severely affects the meaning of the sentence, but it is fairly easy to correct.

After an experiment in which translators and students classified several sentences according to their expectations of editing effort, they divided the classification scale into two levels: usable/non-usable. They reach the conclusion that sentences with HTER scores below 0.2 (which means that 20% of the words need editing) are always classified as usable, but above that threshold some are considered usable and others are not. They refer to the need for professional translators to identify quickly sentences that require too much editing, and they close the paper asking for the research community to consider these issues in their investigations of the subject.

"Why Predicting Post-Edition is so Hard?" is the title of a paper published in 2015, reporting the lessons gained from two failed attempts at developing a system that tried to learn editing actions and apply them in an APE context (Wisniewski, Pécheux, & Yvon, 2015). Their system tried to learn edits from the training set by using edit distances and then apply these actions to the testing set. One of the problems the systems showed was the over-correction effect commented on above. Another was the "uniqueness of edits" – even the most frequent edits (e.g. insertion of Spanish punctuation) only describe a small percentage of the errors. Besides, issues that are important to correct for translators, like punctuation or case, are often neglected by automatic error identification systems. The second attempt focused on specific errors, so as to avoid over-correction, but the results were not satisfactory either, mainly because of inconsistencies in the corrections. So, uniqueness and inconsistency in PE data seem to be the main difficulties the researchers found.

3.5.4. Critical views on evaluation and editing

"A Critique of Statistical Machine Translation" (Way, 2009) presents a critical view on quality metrics. According to Andy Way, the main objectives of these metrics are: to check MT systems' improvements, to compare MT systems, and to provide elements for optimization, or tuning, of the MT systems. However, he points out that the capacity to generalise to other situations from these metrics is very limited. He then debates the correlation between quality scores and “real” quality improvements. He exemplifies with a system that is trained to be sensitive to certain linguistic features, but which fails to put the correct verb in a sentence due to immense variations of verb forms. As for the improvements gained through tuning, he calls attention to the effect of overfitting the results to the metrics that are used during training.

For this dissertation, it is important to consider quality metrics in their role of trying to reproduce the work done in PE. Since they are associated with the four editing actions that were highlighted in Chapter 2, metrics that are based on edit distances are much more adjusted to this purpose than n-gram metrics. Still, edit distances have an issue that needs to be taken into account. Edit distance metrics assume that there is a close correlation between a final version of a translation and the actual actions produced by translators and post-editors. However, that correlation is not guaranteed.

Even on a simple theoretical approach, it is arguable if methods applied *a posteriori*, on a product of a complex process, are capable of estimating adequately that process. As discussed before, the TP is based on a complex network of decisions that is virtually out of reach for researchers. This process is iterative and redundant, but the final result does not tell the whole story that led to it, even at a superficial level.

Correlation between process and product units in edit distance metrics

Quality metrics based on criteria such as good/bad or adequate/fluent, and on typologies of errors, highlight the fact that they are focused on evaluations of characteristics of products, not processes. The translation industry had been able to move the focus of quality standards in translation services from the evaluation of products to services, as enshrined by ISO 17100 (ISO, 2015). So, one may even say that evaluation of MT has taken a step back to the evaluation of such a volatile reality as the products of a TP. But other issues exist associated with this approach to editing.

From a procedural point of view, it is important to check that actions are associated with the correct textual units (see section 2.2.4). When a post-editor makes an edit, it is relevant to know not only which edit was performed, but also which words were affected by the edit. This theme has been debated in the context of TPR in Chapter 2, but we may look at it now under the light of quality metrics.

This relation between editing action and textual unit cannot be guaranteed with an analysis of any translation product, be the process made by a human or a machine. Let us discuss an example for illustration purposes. Imagine the following hypothesis and post-edited phrases:

MT hypothesis: *Our tomorrow country*

Post-edited result: *Our country today*

An edit distance analysis would, most likely, identify two substitutions: in the second position, the word "tomorrow" was replaced by "country", and in the third position "country" was replaced by "today". The TER score would then be 0.66, since two words out of three were edited.

However, a post-editor would probably not opt to write two words when he can write just one. His most likely editing option would be to move "tomorrow" to the third position, and then replace it with "today". This movement and replace action, if correctly logged, would register a semantic relation between the words that were replaced (tomorrow and today), which is useful information about the editing process.

The reason why, after identifying the two substitutions, an edit distance metric would not consider the shift, or movement action, is because movements are only considered if they reduce the number of edits and the TER score. With one movement and one replacement, the TER score is the same, so the system favours the replacement action. In most studies with TER scores (Wisniewski's paper, above, is one such example) replacements are the actions most frequently identified by edit distance metrics, and movements the least used ones. A careful analysis of the reality should be carried out to verify if this is not the result of the metric's architecture.

A few final notes on metrics like TER. These metrics are usually word-bound and do not capture edits that extend to phrases, or other linguistic units. Besides, they are often case insensitive. When MT systems and NLP tools start by tokenising all words and punctuation signals, either re-casing modules are efficient, or the need to correct errors with capitalisation and punctuation become quite regular. Quality metrics evaluations using default setups will often miss these edits. Finally, language-bound features, such as the ones that defined the evolution from TER to TERplus, are not useful when the TL is not English. If one is dealing with a TL other than English, there is no way to take advantage of the improved matching features or weight optimisation. So, TER seems to be a better choice to describe editing work in these scenarios.

In relation to QE, the first point to make concerns the implied recognition that the quality produced by MT systems does not depend only on the amount of data and the language pair, as is so often claimed by technology providers. The four dimensions of QE features (ST complexity, ST-TT adequacy, MT system confidence and TT fluency), with the added depth and variety of features tested in this research area are good evidence of the plethora of factors that contribute to a high-quality translation. The fact that these features have to be studied every time a new unit is analysed (as in the case of phrases) also shows that the knowledge about what goes on in the TP is still very limited.

The question "MT Quality Estimation for Computer-assisted Translation: Does it Really Help?" is the title of a paper which aims at confirming the assumption that QE is useful to increase productivity in CAT environments (Turchi, Negri, & Federico, 2015). The authors carried out an experiment in which they presented translators with MT suggestions with or without QE estimates, in a two-colour code that represents sentences that are expected to take longer or shorter times to post-edit than to translate from scratch. The boundary they set between the two classes is an HTER of 0.4. The answer to the title question is not conclusive, since the fact that translators knew the QE score of the segments only led to increased productivity in specific conditions: after the outliers (segments in which translators spent either too long or too little time) were removed from the analysis; in sentences with 5 to 20 words, and in sentences with HTER between 0.2 and 0.5.

The research on QE at the phrase-level is the most interesting in the scope of this dissertation, not just because of the connections to PBSMT and CAT, but also because it reveals one of the most neglected products of SMT: the phrase tables, or "MT engines", as these are sometimes called. With the notable exception of Moses for Mere Mortals (Machado & Fontes, 2011), phrase tables are out of reach of researchers and translators. Translation phrase tables, created by the trained models of SMT systems, are invaluable resources, not just for data treatment and analysis, but also for linguistic and process analysis. It is surprising that there has not been research based on phrase tables, on matters such as the ratio of linguistically coherent units versus those that have no linguistically recognised structure, and whether that ratio is correlated to the quality of the systems, or on the percentages of changes in phrase tables that an in-domain training and tuning session brings, to name just a few examples.

The paper by Logacheva and Specia (2015) cited above describes the difficulties in extracting phrase alignments from phrase tables, and comments on the limitations caused by training data that only identify edited pairs. Ideally, the use of phrase tables during PE should be linked to a "validation" mechanism, which, for example, gave higher scores to phrases that are often left unedited (i.e. validated) in PE sessions. This way, it would be easy to identify aligned phrases with different levels of quality, created not at specific annotation times, but by the regular PE work translators do.

Closing remarks

It should be clear by now that I admire the capacity to analyse complexity with simple processes. So, I do not advocate that the solutions for the shortcomings of translation technologies can only be achieved by the addition of complex levels and dimensions. Linguistic features, better annotated data, or the reproduction of the complexity in the processes, as the PEA approach seems to do, in spite of admirable attempts to bring more knowledge to the process, tend to create several implementation issues that are often the cause of the lack of attraction of the resulting systems. Translation technology is a vivid example of how much one can achieve by starting with careful observation and simple approaches.

The focus on the four editing actions as simple, yet powerful concepts seems to have been confirmed in this chapter. An added advantage comes from the fact that this concept, although being used by so much high-end research still seems to require further exploration.

One final note on the research methods described in this chapter. Although justifiable from the point of view of resource management, it was surprising to observe so much research being done with simulations of tasks that are performed in real life by translators. The most important legacy of all these years of research on MT, although obfuscated by the fantastic results achieved by automated processes, should be that translation is an activity deeply defined by human processes, and that the major advances are achieved when these human processes are integrated into research, not just as added data, but as fundamental reasoning. The next chapter describes PE as a human activity that, simple as it may seem, cannot be described by oversimplified research.

4. THE POST-EDITING INTERSECTION

As the saying goes: “All roads lead to Rome”. It should come as no surprise then, if at the end of any such road you find a complex intersection, such as a multi-level roundabout, with different types of vehicles and means of transportation cruising through, some at high speeds and others slower, possibly managed by traffic lights or other means of control of traffic flows.

Post-editing (PE) is a complex activity at which several complex routes converge. The translation and localisation industry has quickly adopted this name for a service that is described in a simplified way, but which hides a vast complexity. At the same time, researchers study its many dimensions and apply different analyses to it, unveiling characteristics that call many assumptions into question.

Each of these views reveals where it comes from and its intended destination. However, a concept that lies in an intersection between Translation Studies and Machine Translation, and to which industry and professional practice also lead, cannot be thoroughly described by following one of these routes alone.

4.1. Introduction to post-editing

In the last few years, there has been a growing number of research papers and books dedicated to the study of PE. The references that will be mentioned in this chapter come from TS university departments, from CS labs, and from collaborative projects between both sides. This has not helped clarify what is meant by PE: the TS scholars’ perspective is based on HT and on natural language production and transformation, whereas MT scholars, especially SMT researchers, take the view of translation as a computation and data processing problem.

This chapter starts from a presentation of PE history and of its most common definitions. Then, a brief review of research on PE is presented. Next, the chapter presents the two major challenges that PE poses to TS and to CS, and in particular to ML. This will be the basis for a discussion on PE that enables a new take on the development of tools to specifically support this process.

4.1.1. The evolution of post-editing

> *From the outset, it was assumed that MT would require assistance, with the concepts of pre-editing and post-editing both already mentioned in pre-Mechanical Translation literature.* (García, 2012).

Such a statement may be attested by reading the works cited by García in this paper. Yehoshua Bar-Hillel, for example, proposed "mixed MT", in which humans and machines interact, with humans pre-editing and simplifying language for the processing of machines and then post-editing its products, essentially to eliminate ambiguities, and to add "stylistic smoothing". Bar-Hillel adds the motivation for this type of interaction:

> *If the machine can produce its part in a time span comparable with that of the conventional human translator, the machine post-editor partnership may well be able to compete in time and accuracy with an all-human translator.* (Bar-Hillel, 1951, p. 3).

He suggests three different roles for the "human brains" that interact with the translation mechanism: a "pre-editor" who only knows the SL, a "post-editor" who only knows the TL, and a "bilingual editor" for someone who knows both languages. The idea of "monolingual PE" may have arisen here, but it is interesting to consider this view of the translation and editing process. The role of the pre-editor would be to reduce ambiguities in the ST: in order to be efficient, he would need to focus on the ambiguities that occur more often; to disambiguate them, all he would have to do was to check their neighbouring words. The work of a post-editor would be restricted to specific instructions, focusing on reducing the need to deal with repeated errors. With particular hindsight (this text is written before the proliferation of even transistors!), Bar-Hillel refers to the redundancy of languages and how machines may deal with it. He refers to an ideal situation in which machines could translate ca. 90% of all words and editors would only have to deal with the other 10% of the rarest words.

In 1966, the "*best known event in the history of machine translation*" took place, or so Hutchins (1996) classifies the publication of the ALPAC report (Pierce et al., 1966) A lot has been said about the report's criticism of MT and about its impact on research funding, and Hutchins' paper is perhaps the best source for such an analysis, as it shows the context and motivations for its publication. The focus of the report was on how to use public resources better and on the impact of public investments on a technology that was still in its infancy. The three axes of this evaluation (quality, speed and cost) were used to compare MT followed by PE against HT with the assistance of machines (a precursor of CAT). Besides mentioning the methods for this evaluation, the report includes a section on the receptivity of PE by translators, which, as will be seen in the next section, is a very important factor in this area. However, the focus on the criticisms presented in the report has obfuscated its recommendations, which are worth highlighting. There were two avenues of research suggested by the report: computational linguistics, which should aim at understanding language better (whether it had a direct impact on MT or not) and the improvement of translation.

These avenues of research were actually followed: MT research proceeded to explore the lines of investigation provided by Computational Linguistics, in terms of the grammatical rules of RBMT, and the concerns with the improvement of translation were an incentive to the development of CAT tools. Besides concerns with global evaluation of methods and tools, and with the usefulness of the translations produced, the list of matters that should be researched included items whose importance or technological progress may have not been so clear at the time:

> *Work should be supported on such matters as: (...) adaptation of existing mechanized editing and production processes in translation; the over-all translation process; and production of adequate reference works for the translator, including the adaptation of glossaries that now exist primarily for automatic dictionary look-up in machine translation.* (Pierce et al., 1966)

Only recently did PE become a regular activity that occupies an important part of the work translators are called in to do. So, there is the feeling that MT research never looked at interaction with human translators as a valuable input. Still, as García notes, these early researchers were very much in tune with how to drive this technology farther and make it useful through that interaction, revealing with quite a lot of detail some of its implications:

> *The first 50 years of research and practice are still surprisingly current on key issues such as when to post-edit, and ways of gauging the translatability of a text and its suitability for machine processing; how to post-edit, and whether to aim for publication quality or gisting, according to the purpose of the translation; and who the post-editor should be, whether a professional translator or a bilingual (even monolingual) subject matter expert.* (García, 2012, p. 305)

Simple definitions of PE

In the same paper, García presents an admittedly "outwardly simple" definition of PE: *"It is the process whereby humans amend machine-generated translation output to achieve an acceptable final product."* (García, 2012) Let us discuss a few more definitions of PE.

Bartolomé Mesa-Lao presented this definition in 2013:

> *Post-editing can be defined as reviewing a pre-translated text generated by a MT engine against an original source text, correcting possible errors, in order to comply with a set quality criteria in as few edits as possible.* (Mesa-Lao, 2013, p. 6).

This definition replaces "amend", used by García, with "reviewing", a term that was discussed in Chapter 2. It also clarifies that the level of "acceptable" is defined by a set of quality criteria. But the main difference between these two definitions is the addition of "*in as few edits as possible"*. Keeping the extent of edits to the MT hypotheses to a minimum is a central preoccupation in most definitions of PE. However, the roots of this may not be any intrinsic feature of PE, but the methodological constraints of edit distance estimations referred to in section 3.5.1 above.

The ISO translation industry standard 17100:2015 states that PE is outside its scope. Still, it includes PE in the initial list of terms, as "*edit and correct machine translation output* (ISO, 2015, p. 2). The option for the neutral term "output", as García does, is a way not to assume that the result of a MT process is a "pre-translated text", as Mesa-Lao calls it, or any form of translation. But the ISO standard includes a peculiar note that underlines the lack of clarity of what the ISO understands as PE:

> *Note 1 to entry: This definition means that the post-editor will edit output automatically generated by a machine translation engine. It does not refer to a situation where a translator sees and uses a suggestion from a machine translation engine within a CAT (computer-aided translation) tool.* (ISO, 2015, p. 2).

In this document, we may find PE again, together with pre-editing, in Annex F, a list of added value services that translation service providers may offer. This clarifies that the ISO only considers PE as a specific service, different from translation, revision, and any other, outside the scope of the standard. However, it is not clear what the ISO calls the process or activity by which a translator *"sees and uses a suggestion from a*

MT engine". If it is not PE, is it still translation? If that is so, maybe the standard should include some rules on this use of MT content, since it seems consensual that it has some effect on quality.

According to the previous definitions, PE's defining feature is the textual element over which the translator works, which is the result of an MT process. However, it seems that, for ISO committees, this is not enough when these results are used as suggestions in CAT tools.

The definition of PE presented by TAUS is longer, more detailed and encompassing:

> *Post-editing (PEMT) is the process of improving machine translation output or raw output. It may involve the editing, modification and/or correction of machine translation output in order to ensure that the text fulfils its intended function.*
>
> *Post-editing is done by a post-editor. The term post-editing is specifically used in combination with machine translation, hence the abbreviation PEMT, which stands for Post-Edited Machine Translation.*
>
> *Post-editing covers:*
>
> - *Correction of recurring and predictable errors*
> - *Focus on words and phrases*
>
> *Post-editing strategies:*
>
> - *Keeping changes to a minimum*
> - *Applying quick fixes*
> - *Automating certain word-processing operations*
> - *Providing regular feedback to the machine-translation team*
>
> *Post-editing facilitates faster production and will help meet the growing demand for translation. Depending on the purpose, the final outcome of the post-edited text can be achieved by light post-editing or full post-editing.* (TAUS, 2016a)

This definition confirms that the result of the MT process is commonly known simply as "output" or even "raw output" and it introduces some of the themes that must be discussed when we talk about PE, such as what to focus on during PE and how to approach it to control its extension. It also underlines that PE is closely linked with "faster production". And it highlights that much of the process depends on "its purpose", this being exactly to speed up production, just like any other automation process, as admitted on a different page of the same organization:

> *Since the overall aim of any translation automation solution is to accelerate throughput at consistent quality levels (and where possible reduce costs), ongoing efforts to optimize post-editing are focused on improving raw translation quality to maximally reduce the post-editing workloads.* (TAUS, 2010)

The definition of PE presented by this institution, which seems to speak for the whole localisation industry, includes specific strategies that modify what is usually seen as the work of translators or revisers. These strategies not only condition this work (minimum changes and quick fixes), but also expand it, with the requirement to automate operations and provide feedback. So, it seems that TAUS considers that post-editors are part of the automation and development teams, while keeping their usual work as linguists to a minimum. With this, TAUS seems to see PE essentially as a skill that focuses, not on producing a high-quality TT, but on improving MT systems.

From definition to productivity control

As an industry-led institution, TAUS is very focused on productivity metrics. This has set the agenda of much of the research done on PE and on the global perspectives about what PE is. It is thus unavoidable that a section on defining PE includes a reference to productivity. The same page of the TAUS web site says: *Publication quality post-editing should be able to output at least 5,000 translated words a day, whereas lighter editing for gisting (assimilation) can double this rate.* (TAUS, 2010)

To compare these numbers with translation and revision, let us consider that a rough average production rate for translation is 2,500 words per day and for revision around 8,000 words per day. This means that PE productivity should be around 2 to 4 times faster than translation, and that a post-editor should produce between 2/3 and 4/3 the amount produced in the same time by a reviser. These or similar levels of productivity are set as targets by industry Project and Production Managers and investigated by PE researchers.

Two PE modes: light and full

Although admitting that there are several strata of PE productivity, which depend on such factors as language pair, bilingual data used for the MT training, post-editor profile, quality of the MT output, and, above all, the purpose and intended audience of the PE product, at least two different levels of service, or modes, were proposed for PE. These are usually known as "light PE and "full PE".

In the article he wrote for "Computers and Translation", Jeffrey Allen connects these two levels with a two-fold view of the purpose of PE projects (Allen, 2003). Light PE seems to be linked to "inbound translation", or "translation for assimilation", which means that the results of this type of PE will be used in an environment where the TT is accepted as a low value text, for short periods and limited purposes, such as basic information transmission or "translating to understand", with no tight requirements in terms of quality. As for full PE, this is supposed to completely replace the translation production process and the result should have enough quality to "communicate", a purpose which can be described as "fit for dissemination", or "outbound translation". Allen discusses this level of PE in contexts where controlled language was used on the ST side, thus providing a better output, which should create the conditions for the intended result to be achieved without a very demanding PE process.

Jeffrey Allen also comments on how different PE is from revision. His focus is on the profiles of post-editors, specifically in avoiding what he describes as the "red-pen syndrome". This view is echoed in several other papers, in terms of the need to avoid "over-editing", as will be seen in the context of PE guidelines. Several researchers discuss this in relation to the demands PE poses to translator training schools. Authors like Allen also discuss other dimensions of the PE process, such as: levels of quality that are expected for each PE mode, PE guidelines, typical linguistic errors created by MT, and other issues that will be analysed in the next section.

One final point, before closing this introduction to PE: regardless of the discussion on each section below, concerning the identity of PE and its relation to translation and revision, it is important to stress that this dissertation accepts that "post-editing" is the most consensual designation for this MT output improvement/correction process.

4.1.2. Post-editing research

As we have seen in the previous section, early works on MT had already raised many questions that remain at the centre of the PE debates, like the profiles of post-editors and receptivity from translators, evaluation in terms of quality, speed and cost, how to select what needs to be edited and what should be approved, all in a context that demands high productivity even if that means low quality. These themes are discussed in this section, in a review of the many studies published in this domain.

Kring's "Repairing texts"

"Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes" is Hans Krings' postdoctoral thesis, written in 1994, but translated and edited in 2001 by Geoffrey Koby (Krings, 2001). It is considered the most important work on PE, having given an important contribution to the study of PE, namely because it set the standard for a central method to measure PE, based on three dimensions of effort.

The Introduction to the book, written by Koby in 2001, has a title that sums up nicely the wide scope of this book: "Post-editing of Machine Translation Output: Who, What, Why and How (Much)". This chapter begins by referring to the cost pressure that has led companies to reduce the high costs of translation, up to the point when they identified this task as something that did not require bilingual knowledge, as something that even monolingual personnel could perform. He refers to other research works to explain the parallelism that some authors see between revision in HT and PE in MT. The author refers to Sager when he says that, whatever language reproduction capacities the machine may have, it produces an artificial language, and human translators are required to turn that into natural language. He also refers extensively to previous works on PE that analyse the types of errors and how to correct them. It is important to clarify that all MT systems described in this book are RBMT, with a very specific type of errors. He underlines the fact that, since translators are very seldom specialists, they painstakingly try to get everything right. Where this should be seen as an invaluable resource, for many it seems to be a hindrance to an efficient process.

Koby follows other authors who claim that whenever translators try to fully replace the machine, by deleting and rewriting the translation, they are missing the point of PE, which is cutting costs. So, they should be trained to identify mistakes, and appreciate the fact that there are certain errors that machines do not make. However, these authors admit that constant exposure to errors should be avoided. As for the question on who should post-edit, it seems that there is a consensus in the fact that only translators, and the more experienced the better, can judge the accuracy of a translation, and possess technical resources to solve translation problems. The author then reviews the profile and skill-set of post-editors and he highlights linguistic and technical skills, which include typing, but also cursor positioning and the ability to use search and replace features.

Another important point needs to be made here about the book: the use of word processors to translate and edit text was still not very common when the thesis was written, so the experiments in the book were mostly done on paper. The technical details on the use of computers for the task come from the work done at the Pan American Health Organization (PAHO), with a software called SPANAM and ENGSPAN, which will later evolve to PAHOMTS, and which will be mentioned below. The authors that presented this system in several conferences in the 1980's claim that some post-editors achieved production rates of 10,000 words per day with their system, and that most translators are comfortable in working with PE after one month of work (Vasconcellos, 1987b). Koby ends the introduction reinforcing the notion that in 2001 PE had secured a place for itself in the localisation industry, and that it would become a regular part of the professional life of many translators.

In Hans Kring's introduction, which was written in 1994, he refers to Hutchins when he mentions that the term post-editor was introduced in 1950 by Edwin Reifler, and he also comments on the first references to the two modes of PE: these were initially presented by Anne-Marie Löffler-Laurian as "rapid" vs. "conventional", and later as "full" vs. "rapid", by Emma Wagner. (The author decides to adopt the terms "partial PE" vs. "complete PE".)

Both Anne-Marie Löffler-Laurian and Emma Wagner describe their experiences with the use of Systran at the European Commission (Löffler-Laurian, 1996; Wagner, 1985). Löffler-Laurian mentions the criterion obligatory/non-obligatory to separate the edits that are acceptable at one level of PE or another. Emma Wagner describes the challenge PE of MT poses to translators. She admits that highly qualified translators, such as those that work at the European Commission, start by struggling against working with these systems, but they gradually adapt to them. Many reject using the system to produce high-quality final translations (by performing a full yet cumbersome PE), but they accept working with the system for rapid PE, if that is what the final user wants.

However, Krings claims that, to his knowledge, although it is mentioned in most literature on the subject, translating for gisting does not play a major part in real life. He had previously in the text commented on "raw human translation", a non-existing task, with the same quality and efficiency parameters of MT, but which apparently no one demands in real translation practice. And he reinforces this view:

> *First, how great is the need for partially corrected machine translations? They are not plausible alternatives to full human translation and as a preliminary to a later complete post-edit even less so, because the time saving yielded by a fast post-edit would not only be wiped out by the complete post-edit, but even reversed. As a practical function of the partial post-edit, only its value as a source of raw or "gist" information remains. However, in this case the partial post-edit is competing with the raw machine translation, whose production costs are clearly much lower and which can also be assigned the gisting function. A comparison between the two modes is called for. In both cases, it is conceivable that information extraction from the texts will be disrupted as a result of the errors in the machine translation and that, as a result, the extraction process will also be lengthened.* (Krings, 2001, p. 54)

How to measure effort, with a focus on its technical dimension

There will be a discussion on the gains and losses between light and full PE below. At this point in the text, Krings stresses that most discussions and papers on PE focus exclusively on time savings, and relates time savings to a reduction in the effort applied to the task. Time effort, Krings says, on the same page, *"is the most important aspect of post-editing from an economic perspective. But the time aspect is ultimately only the obvious external form of post-editing effort."* When he refers, in the citation above, to the extra effort introduced by a more complex information extraction process, due to the presence of random errors in the MT text, he is already hinting at the cognitive dimension of effort.

The three dimensions of effort in PE are the main contribution Krings has brought to the study of translation, and of the relation between HT and MT. The three dimensions are:

- Temporal effort
- Cognitive effort
- Technical effort

Technical effort is the most relevant of the three dimensions for the purposes of this dissertation. Krings says that this dimension has received a “certain amount of attention” from researchers. The focus here is on practical details that look very simple, but reveal a lot about the task:

> *Anyone who has made extensive changes in a document using a computer knows that a myriad of mouse clicks and/or keystrokes are required. If one considers that post-editing is almost exclusively made up of deletions, insertions, and rearrangements, it is easy to see how little this kind of task has in common with normal linear writing using a word processor.* (Krings, 2001, p. 55)

The reference to “arrangements” may be understood as containing replacement and movement actions. Even so, it would seem that Krings demonstrates that this is an uninteresting dimension of the analysis. You measure mouse clicks and keystrokes, and you obtain all the information you need. But that is not the most revealing part of this citation. What this extract reveals is that these techniques are associated with a very different task from writing, both in the technical and in the cognitive dimensions. But let us leave this discussion for another section, and remain with Kring’s line of thought.

Technical effort shares with time effort the fact that they are externally observable. Next, the question arises on how to measure it. One of the possibilities is to measure the number of errors in the MT text, and that has been a major focus of attention from research, namely by adopting the edit distance approaches that were analysed in the previous chapter. After describing some of the implications of this approach, he tackles a different challenge: how to reduce technical effort.

At this point in the text, he resorts to the papers published by the PAHO team, who used an RBMT system linked to a word processor with editing macros that helped make specific corrections. The features used by this system are supports to typical editing operations in word processors, like cut and paste, search and replace, change in capitalisation, among others. Krings adds that users have the possibility to send new terms to the system's dictionary, even sending entire phrases as fixed units. The purpose of all these aids is, above all, to improve the quality of the translations that are produced by the MT system, as this is seen as the best way to reduce PE effort.

The reflections on cognitive effort have a different purpose: studying how the process happens, what demands it makes from post-editors and what kind of receptivity translators have to PE. Krings takes this as a point of departure for a very interesting review of TPR literature in the early 1990's.

The PE process

In chapter four, Krings looks more closely at the process that occurs during PE. He starts by analysing what happens during the reading stage. The fact that the post-editor has two texts to read, in two different languages, is a very different setup than the one that exists in translation. To start with, the post-editor decides which of the two texts should guide his attention, and Krings says that this depends on the quality of the MT text. He hypothesises that, during PE, reading is made on smaller units, because of the errors that are to be found, thus leading to a higher burden on short-term memory and to processing text more at the micro-level. Besides, the unexpected types of problems and alternatives that the translator finds in the MT text, different from his expectations from reading the ST, may lead to the development of strategies that only arise in that particular context. The chapter presents several very interesting questions that were worth a detailed analysis, such as: the interaction between different types of knowledge, the effects of specific tasks, the influence of technical resources and situational environment on the process, the relation between PE process and PE product, and how to approach this theoretically and pedagogically. But let us focus on the reflections on the factors that determine PE effort.

Temporal effort, being the easiest to measure, is not proportional to the number of errors found, since some errors demand a higher effort to correct than others. So, Krings says, cognitive effort should be the main measure of effort. As for the relation between cognitive and technical effort, Krings mentions that, although activities performed by the post-editor are usually seen as demonstrations of cognitive effort, in computer settings these are seldom related to one another. In fact, technical effort depends mostly on how each action is performed in the computer tool interface: some systems have very clear input methods, whereas others imply specific selection of keys to perform operations that would be very simple from the cognitive perspective.

Krings proposes "absolute post-editing effort" as a "zero value" that describes the situation in which the MT text requires no editing. Then, "relative post-editing effort" is the ratio between this value and the value for "translation effort". A value of 1 in relative PE effort means that PE took as long as a translation from scratch; below 1, it means that MT brought an increase in productivity, and above 1 means that PE was inefficient, since it took longer than translation. This value should never be 0, because at least reading the MT text takes time.

The author refers to the need to predict this effort from the analysis of source material, and he says that the focus on cognitive effort is fundamental for this, since temporal and technical effort can only be measured at the actual operation of PE.

Kring's experiments and conclusions

For the experiments devised by Krings, he used qualitative and quantitative analyses, collecting data through questionnaires and TAP during and after PE. Three main tasks were tested:

- PE with ST
- PE without ST (also known as "monolingual PE")
- Translation

These main tasks had a few variants, like English and French as TL, and PE without TAP. Besides, the subjects replied to a questionnaire and commented retrospectively on the videos that recorded their performance with the think-aloud reflections during those activities. It is important to remember that most of these experiments were done with the subjects editing texts on paper, with only a few done on the computer for comparison. The chapters describing these experiments and the data selection process are a major part of the book, and contain useful information for anyone doing research on PE.

Chapter nine, for example, describes a coding system that transforms the processes performed by the subjects into carefully segmented units. This segmentation starts by breaking the reflections that subjects verbalised during the task, and which could be analysed in the TAP recordings. Most of the units of analysis are segmented by a continuous reference to a text unit, separated by a pause or constituted by a single proposition from the subject. Other observations are defined by references to focus on specific parts of a text, and factors like writing continuity are also considered to segment a unit of analysis. The researcher then linked these segmented units to different textual units, in the different resources the subject processed, be it the ST, the MT text, the TT or the reference materials. This resulted in eight main classes of processes, which also take into account whether the subject refers to reading or writing procedures, to the task in general, or to elements that are not part of the task, such as their own attitudes and competencies. The full list of classes of processes are:

- processes not related to the task
- processes related to the task
- processes related to the ST
- processes related to the MT text
- processes related to the production of the TT
- processes related to reading/monitoring the TT
- processes related to the reference materials
- processes related to writing.

The first conclusion Krings draws from the collected data is that most of the subjects' reflections (2/3) are on the production of the TT, and this is a constant across all tasks. The second highest category is related to writing processes. This supports the notion that in any task, the translator is mostly focused on the responsibility of producing the TT. He also concludes that cognitive effort is higher for PE than for translation, after analysing features such as changes of attention focus per word. This factor is doubled by the presence of the MT text.

In chapter ten, Krings makes a detailed analysis of all these factors, but the main interest for the analysis is on the writing processes. The author mentions that these may be actually observed, and the same goes for search in references. The advantage that derives from this is that the researcher may add "hard data" to the subjects' reflections. To segment writing processes, he considers pauses of one second, and he analyses linear writing processes, and non-linear ones. In linear processes, the subjects add elements in a sequence, with pauses, but without interrupting the sequence by non-linear actions. Non-linear actions are mostly deletions and insertions, but sometimes subjects join those actions in a replacement/overwriting action, especially in short edits. He says that two thirds of all actions are linear, and one third is non-linear. The conclusions to be drawn from this analysis is that PE seems to be a more linear process than translation, when translators work on paper. Another interesting conclusion from this chapter is that it is medium quality MT text, not low quality one, the one that leads to the greatest cognitive effort.

In the concluding chapter, Krings mentions that only one-third of the MT sentences did not require editing, which makes these MT outputs "hardly usable". In terms of strategies for the reduction of cognitive effort, the book is not conclusive, because it depends on the factor that the researcher looks at: if one considers focus changes, PE seems to require more cognitive effort; however, if one considers text writing linearity, PE could be associated to less cognitive effort. The analysis of cognitive effort is hardly applicable to the currently most common technical environment, which is on a computer, with a CAT tool, that always shows the ST, and has overwriting as the main typing method. Still, Krings' work was ground-breaking and it influenced all posterior research on PE considerably.

4.1.3. Themes in post-editing research

This section aggregates research on PE around five main themes.

Professional and personal implications of PE

Geoffrey Allen said that PE introduced a new challenge to professional translators: *"the acceptance and use of half-finished texts"* (Allen, 2003). Frank Austermühl refers to work in the domain of Psychology to describe two major types of attitudes towards this type of challenge: "satisficers", for whom the correct strategy is not looking for optimal solutions but meeting adequacy criteria, and "optimizers", who commit to *"the art of finding the best choice among all choices"* and who may face an identity crisis in view of the industry pressures (Austermühl, 2013).

To study the impact this effect may have on different professional profiles, a typical line of research is to compare the performances of students and novice translators against that of professional experienced translators in PE tasks. One example of this type of research is: (Guerberof Arenas, 2014). Özlem Temizöz follows another approach: comparing the performance of professional translators against that of subject matter specialists (Temizöz, 2014).

Most of these views take on the approach, or test the assumption, that PE implies a downgrading of the skills of translators. But pedagogical views may approach this from a different angle and consider that PE is a specialised task, and that trainee translators benefit from learning to work with MT and PE. Dorothy Kenny has joined other researchers at DCU and published some of the most interesting papers in this field (Doherty & Kenny, 2014; Doherty, Kenny, & Way, 2012; Kenny & Doherty, 2014). Above all, these papers consider that these skills may empower translators and allow them to play news roles in the industry. These views have been put into practice at DCU, in innovative curricula that bring together information from TS and CS. Students learn to select data to train MT engines, produce their own MT suggestions, check their quality and tune the systems, trying to achieve better quality. A similar experiment is at work at the Universidad Autónoma de Barcelona, under the MTradumàtica project (Martín-Mor, 2017).

PE guidelines

One of the most general guidelines in PE has been given by Muriel Vasconcellos, back in 1987: *"When in doubt, don't."* (Vasconcellos, 1987a). This type of instruction is echoed in the title of an oft-cited article published in an online journal 30 years later: "Manual MT Post-editing: If It's Not Broken, Don't Fix It" (Guzmán, 2007). Emma Wagner also advises post-editors not to "embark on time-consuming research" (Wagner, 1985). These expressions refer to the already mentioned principle that PE is defined as an activity circumscribed by quality and productivity red lines, which, if crossed, will lead to loss of efficiency and money.

In 2002, Sharon O'Brien proposed the creation of a course on PE at DCU, but at the time there were no industry guidelines (O'Brien, 2002). However, the value of her work in this area was recognised and she became one of the main contributors to industry guidelines, namely to the first ones published by TAUS in 2010, and to the most recent ones, published in 2016 (Massardo et al., 2016). In Spain, Célia Rico Pérez proposed a framework for the creation of PE guidelines that focuses on more than text features, and encompasses project and process management dimensions (Pérez, 2012).

In 2010, Ilse Depraetere tested a range of different problems in PE projects and confirmed that student translators are receptive to producing "less-than-perfect translations" (Depraetere, 2010). Marian Flanagan and Tina Paulsen Christensen tested the use of PE guidelines as guidance to translator training (M. Flanagan & Christensen, 2014). They analysed the linguistic rules suggested by guidelines, and, after realising that students might not have the capacity to understand the instructions, they proposed adding detailed information to simplify the adoption and training of such guidelines. A most recent study does a comparison of different guidelines for PE work (Hu & Cadwell, 2016). This paper shows that different PE guidelines (three academic ones, plus one from TAUS, and another one from a private company) approach light and full editing differently. To compare them, they use evaluation parameters like accuracy, terminology, syntax, style, restructuring, etc. Some of these parameters are not referred to in some of the guidelines, while others are restricted by some and accepted by others. In view of the multiplicity and divergence between these guidelines, the authors comment on the need for translation service providers and clients to agree on the parameters for a PE service before entering into a commercial agreement.

Studies on measuring PE

Quality, productivity and effort metrics are some of the themes most regularly studied by PE research. The complexity and extension of work that applies these metrics would require this dissertation to dedicate an extensive section to it. However, this theme is not directly associated with the main line of this dissertation, so the option will be to focus on a very limited amount of research. The main point to make in this section is that, although the value of these metrics cannot be questioned, assumptions based on their results may reveal flaws in the theoretical grounds that sustain them.

Two PhD theses present a very good synthesis of the most relevant research on these metrics: Ana Guerberof Arena's, on the relation between productivity and quality, comparing TM fuzzy matches and MT hypotheses, and the one by Maarit Koponen, who studies PE effort (Guerberof Arenas, 2008; Koponen, 2016).

The paper "Correlations of perceived post-editing effort with measurements of actual effort" describes an extensive study with these metrics, trying to find a relation between these metrics and translators' perception of that effort (Moorkens, O'Brien, da Silva, Fonseca, & Alves, 2015).

Besides these studies, "Advancing Best Practices in Machine Translation Evaluation" presents a thorough portrait of the different metrics that may be used for the incorporation of MT into professional services (Aranberri & Choudhury, 2012). The report presents TAUS' DQF (Dynamic Quality Framework), a set of metrics for use in the industry. The report contains an important observation on the use of these metrics:

> *TER is an automatic metric that seeks to indicate the work required from post-editors. However, the minimum number of changes is*

> computationally *optimal, and may not necessarily be intuitive from a linguistic viewpoint. TER might shift a word to a different part of the sentence and thus reduce the overall change count. Yet post-editors work in linguistically motivated steps. This de-synchronization as well as the fact that post-editing time is not taken into consideration can result in the metric losing any meaningful correlation with PE effort.* (Aranberri & Choudhury, 2012, p. 8)

Although the "linguistic motivation" for translators' editing actions is not tested and demonstrated evidence, this is an important word of caution concerning the conclusions to be drawn from the use of these metrics.

There are not a lot of examples of research papers on PE using edit distances and editing actions. "Relations between different types of post-editing operations, cognitive effort and temporal effort", by a team from DFKI (Deutsches Forschungszentrum für Künstliche Intelligenz – The German Research Center for Artificial Intelligence), is a notable exception (Popovic, Lommel, Burchardt, Avramidis, & Uszkoreit, 2014). This paper presents a study of the relation between edit distances, cognitive effort and temporal effort. The authors study the four editing actions, but replacement is divided between changes in word form and in lexical choice. They also add an analysis of the distance of movement actions. All the edits were measured automatically, but quality scores were annotated manually. The authors present a simple method to estimate the average time per edit, and show that each edit takes on average 5 seconds (in the language pair French to English) or 10-15 seconds (in English to Spanish translation). They conclude that the number of editing actions is strongly correlated to the evaluation of quality, especially for sentences with errors of reordering and lexical choice. Long distance movements are also correlated to lower quality and longer PE time. Mistranslations (replacements) are the most common edits in all cases. As for editing time, it seems to be correlated to sentence length.

The effort threshold

Another line of research that is not commonly explored in PE studies is the identification of the threshold that separates useful MT suggestions from those that imply so much editing effort that it makes them useless.

In "Comparing Translator Acceptability of TM and SMT Outputs", Joss Moorkens and Andy Way confirm their assumption that response by translators to MT could be improved if MT suggestions and TM fuzzy matches took into account such a threshold (Moorkens & Way, 2016). The authors of this paper mention that preference for TM in place of MT may reflect a preference for precision in place of recall, or, simply put, for quality over quantity. So, they focus on the study of a threshold associated with MT system confidence evaluation, that they hope can break the resistance to MT by translators. In the experiment, a few translators were asked to rate the usefulness of a random selection of 60 sentences, half translated by a TM (with varied fuzzy matching levels) and half translated by SMT. The ratings were divided into 3 levels: "not usable", "useful" and "almost perfect". One of the pieces of evidence shows that, for up to 85% fuzzy match levels, respondents rate MT suggestions better than TM suggestions. They state that this is a demonstration that arbitrarily setting the threshold from which the TM fuzzy matches take over MT is a risky decision that may harm performance. They comment on the cognitive effort of identification of errors, and cite Phillip Koehn when he says that, in PE, *"most of the time is spent on contemplating changes, but very little on executing them"* (Koehn, 2009a).

Alignment and segmentation of units of analysis

It is fairly common to find, in lists of translation errors, categories that apply at the same time grammatical classes, editing actions, quantitative details and, sometimes, contextual restrictions. Several suggestions have been presented and attempts have been made to automatize such corrections, for example in the ambit of APE – see, for example (Allen, 2003, p. 314). However, these systems have not yet shown reasonable success in practice, and perhaps the main reason for this is the difficulty in basic tasks, such as aligning and segmenting these units.

We have seen above that analysis of the TP in TS and analysis of edit distances in SMT depend a great deal on the segmentation and alignment of TUs. We have also seen that this concept is not an easy one, and that these concepts become even more difficult to clarify when we add actions and time segmentation to text unit segmentation. The scenario becomes even more complex in a process that needs to fit into a CAT workflow, a MT workflow, or even both.

So, perhaps one needs to start by modelling the TP and the editing actions, and only then develop tools that integrate aids as part of the model. So, let us review the importance of TP models in the context of PE.

4.1.4. Modelling post-editing

This section presents several studies that focus on how to reproduce the PE process, so that tools may learn as much as possible from it, not to try to reproduce it, but to support those that perform it.

A model to contextualise editing decisions

In 2008, Takeshi Abekawa and Kyo Kageura proposed the construction of a corpus with the intention to capture the implicit knowledge in translators' decisions (Abekawa & Kageura, 2008). In their preparation of a system to capture this knowledge, they looked at the differences between the two usual versions of a translation: the draft and the final translation. Based on observations of translators editing draft versions and making decisions at the sentence level, they defined the sentence as the main unit of analysis. However, to gain further detail, they placed the units of modification below that level. Then, they defined a scale of different categories in which to organise the information retrieved from data: reasons for modification, aims of the modifications, linguistic operations, and primitive operations. At the level of primitive operations, we find the usual four editing actions. Their system links these classes in an integrated and formalised way. They explain that this work has a solid theoretical grounding on translation work:

> *One of the most important points we found is that translation as perceived by experienced translators has little to do with the language that linguists and computational linguists see (Kageura, 2006). Translation is concerned first and foremost with individual texts. This fact introduces a historical dimension into the translation.* (Abekawa & Kageura, 2008, p. 2003)

This is an important point to make, as so often translation technology is totally focused on linguistic processes, but pays little attention to the TP. The authors link this to the description of each translation act as included in three spheres: the linguistic, in which it shares properties with other objects or corpora in the same language; the textual, which gives context and history to decisions; and the decision-making sphere, which is individual and unique. Although this is the only account identified of this system, it seems to follow a well-structured plan for the development of a tool that intends to capture the relevant information pertaining to each edit decision.

From the detailed description of a process to its prediction

The work developed at CRITT, by different teams of researchers led by Michael Carl, in conjunction with other research centres, is a clear demonstration of how far carefully planned research focused on process analysis can lead. This section explores several works produced from this centre, culminating with the publication, in 2016, of a landmark compilation of work in the field. The focus of this analysis will be on the details revealed by editing actions.

In "Towards statistical modelling of translators' activity data" (Carl & Jakobsen, 2009), the authors present the concept of UAD (user activity data) and explain their model for the analysis of typing and reading behaviour, in a method that includes product and process data. On the process side, eye-tracking data is used to segment units from the textual data, which was collected by keystroke logs and is essentially composed of deletions and insertions at the character-level. They mention that cutting and pasting operations increase the complexity of the analysis, and add that, the smaller the units of analysis, the easier it is to capture the actions. Movement is an operation that affects longer sequences, but which is not observable in their data. They also refer to the observation that some of the keyboard activities are not linked to a particular word. The main conclusion from the data collection stage is that the difficulties in capturing a description of actions are related to the lack of alignment with ST words, and to recursive operations applied to the same units.

In the section dedicated to the discussion of the paper, the authors comment on several questions on how to assist HT processes with automated tools. The first two are: *"At what moment during the translation would the mechanical help be most welcome? Would a translator be better supported during the "linear" translation production or during the translation pauses?"* (Carl & Jakobsen, 2009, p. 136). They then discuss the distracting impact of typing suggestions and how to integrate these and other aids into translation tools. They advocate that process analysis may help identify reading patterns and develop tools for even the reading task.

A few of the ideas in this paper are reinforced in a publication by Michael Carl in the same year: "Triangulating product and process data: quantifying alignment units with keystroke data" (Carl, 2009). On the issue of the role of technology in assisting the translator, he makes the following comment:

> *While a general theory of human translation founded on translation process data may perhaps not be reached in the near future, we show*

a way of correlating and modelling the data in a quantitative manner. At this point we are still far from being able to formulate conditional probabilities over Process Data (PD) and AUs (and maybe also Translation Units) which would answer questions such as: ***given a history of PD what is the probability of the next AU$_i$ to be deleted, inserted or modified?*** *The extent to which we can answer this and related questions will ultimately determine the success of integrating advanced translation aids with human translation activities.* (Carl, 2009, p. 245) [my emphasis].

This extract clearly states the purpose of designing translation systems that do not just support typing, but approach editing as a set of actions that should be modelled independently.

The time has come for translation tools to be predictive

In 2016, a few signs seemed to indicate that technology and research might have reached the evolutionary state required to pursue the goal of predicting editing actions.

The CRITT TPR-DB is a database of data collected from 500 hours of recordings of translation activity, annotated and compiled at CRITT. The database includes 200 different annotations on AUs, TUs, time, keystroke and eye-tracking, at word and character level, thus enabling a very ample spectrum of research studies. The data was collected with Translog-II (Carl, 2012) and CasMaCat (Alabau et al., 2013). This data fed many research projects and many publications.

In 2016, several articles that represent the state-of-the-art of the technology were assembled in a volume entitled: "New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB" (Carl, Bangalore, & Schaeffer, 2016).

In the introduction to the volume, the editors state:

We are now at a stage in the development where translation research becomes predictive. The records from keylogging software and eye-trackers make it possible to address Holmes' (1972) second main objective, to "explain and predict" translators' behaviour: at present, ***we have all the necessary tools to address the challenge of building a model of human translation which makes specific, falsifiable predictions regarding the process and the product of translation.*** *Perhaps the most fundamental question in this regard is to determine*

> *the mechanisms underlying the production of translations which are common to all translators. This babelian question attempts to find, on the one hand, the cognitive processes which are shared among all translators during the translation of diverse language combinations; it is the quest for linguistic and cognitive universals of translation. On the other hand, from a utilitarian viewpoint,* ***having a model which can predict translators' behaviour makes it possible to design translator assistance just when it is needed. It will allow us to automate those aspects of the translation process that save mechanical effort****, so that the translator can dedicate their full attention to those aspects which cannot be automatized.* (Carl, Bangalore, et al., 2016, p. 4) [my emphasis].

This bold statement gains form in the different articles in the book. Chapters 3, 4 and 5 of the book will be described below. Chapters 6 and 7 focus on the use of reference materials and how users interact with them. Chapters 8 to 11 go deep into the cognitive processes, discussing concepts such as priming of syntactic and semantic units, and activation of TL during ST reading, in a sort of anticipated mental TP. The final chapters discuss issues such as the differences in focus (on bigger or smaller units, according to experience) and criteria for the segmentation of units.

Modelling revision

In section 2.5.3 of this dissertation, there is a reference to the need for tools that correctly model the relationship between three texts: the ST, the unedited TM text, and the translator's version. In revision of PE, this is even more important because the translator works with a text that is fully converted into the TL, either by the TM or by the MT. Besides, the transformations that he operates on this TT are more detailed, and more difficult to identify. So, for a model of a PE text to be accurate and provide useful information to the reviser, it should describe this transformation with more detail.

ML methods that support the PE process

Online and Active Learning (OL and AL) are the preferred methods for MT systems to learn and predict technical PE effort. The authors of chapter 3 of the book apply these techniques to ITP methods, and test them against each other (Ortiz-Martínez, González-Rubio, Alabau, Sanchis-Trilles, & Casacuberta, 2016). OL is used to learn from user actions and incrementally adapt the statistical models in the SMT system. AL is used to determine which sentences need editing, as a way to balance quality and effort. The authors comment that the potential of these methods is still not implemented in CAT tools and they test these methods with CasMaCat. Their conclusions are that OL does reduce typing effort, and AL does allow for a higher quality for the same effort.

The authors comment on the need to define incremental update rules for OL systems. An incremental version of the EM algorithm, together with a log-linear model trained on seven features, are the fundamental pieces to perform this.

AL is applied on a separate task and system. In this case, a simple IBM Model 1 is estimated. The sentences that the system selects for PE are the ones that present the highest levels of uncertainty. After these are post-edited, the system is retrained by incorporating what it has learnt from those new examples, and a new estimation of uncertainty filters the only sentences that need PE, from the next set of sentences. Thus, the system reduces the volume of sentences that require PE, so that the user does not have to deal with sentences with repeated problems. In spite of the positive results obtained with this ML technique, this method is not connected to an interactive editing mode.

In the article by Ortiz-Martínez and colleagues, typing effort was measured by the number of source words per time (speed) and number of keystrokes (which they call “edits”) divided by the total number of words (which they call “effort”). The authors comment on the observation that users of an ITP with OL spend more time outside the CAT tool, and that when this reading time is removed from the time logs, the productivity gains of OL are much clearer. However, 40% of the users were faster using a non-OL ITP system, although at the price of more effort, in terms of number of keystrokes. These users were slower with the OL suggestions, because they felt the need to check the suggestions presented by the system on the web.

PE effort and strategies to reduce it

In the next chapter of the book, Fábio Alves and his team present a study on the impact of ITP on the three dimensions of PE effort (Alves et al., 2016). They test two hypotheses: a) with ITP, PE time will be shorter, and b) the number of edits, as measured by TER, will be smaller. One of the interesting conclusions is that higher TER scores are not associated with higher fixation time. Since the two measures are used to elicit PE effort, such a lack of correlation challenges the assumption that any of these measures is an evidence of effort.

In chapter 5, Alabau and colleagues used CasMaCat in two different use scenarios: in traditional PE environment (with just one full MT hypothesis, and no interaction) and in an ITP scenario with OL (Alabau et al., 2016). The experiments extended over a long period (6 weeks) and the authors report reductions in typing time. However, this conclusion is hindered by a behaviour that the authors observed: users often disregarded the suggestions and kept on typing and overwriting them, just to avoid the cognitive interruption to the action flow. The authors mention the need to control this overwriting behaviour, however it is not clear whether this behaviour actually affects speed.

At the end of the experiment, results on acceptability of the system were not very positive. In fact, all users, except one who gave very positive feedback, said they preferred to post-edit without interactivity. Sentences change dynamically: as the translator types, the rest of the sentence in the suggestion adapts to the new words. This interferes with the cognitive flow considerably. One user proposed that these suggestions might appear in a separate window as a form of avoiding that interference. The typing experience and the efficiency of the subjects were factors that contributed to this negative result, in a reaction that is similar to that of experienced drivers' tendency to disliking driving with automatic gears for the first time. So, the authors suggest that time is all it takes for translators to get used to and start taking advantage of the suggestions. Then, they posit the hypothesis that novice translators may be more receptive and adapt faster to the new interactive techniques.

Everything above supports the notion that, although the technology is available to model and predict PE work, the issue of how to offer that in a useful and usable way is an open matter.

4.2. First challenge: understanding post-editing

When you are driving and you arrive at a complex intersection for the first time, it is not always easy to understand each car's behaviour, the next step they take and how to make your way through it. In such cases, a good understanding of the global picture is useful. To access the bigger picture of PE, one needs to understand how MT works, but a good theory of the process is also required.

This section discusses the challenges that PE poses to Translation Theory. To achieve a good description of what PE is and what it involves, one needs to resort to TS, but also to MT theory. This section reviews definitions of PE, and it discusses the relation between this activity, translation and revision. Then, it clears away a few notions that are often associated with PE, but which should probably be put in a different section. And it ends with a description of a few details that are only visible in PE practice, as it happens every day in translation companies and freelancer workspaces.

Before we move on into the intersection, it is also good to know our destination. The proposals for a theory of PE that this dissertation presents will only be discussed in the next chapter. But the identification of the main lines is called for now, so that the route that we must take in the intersection is clearer. This is the definition of PE proposed in this dissertation:

> **Post-editing** is a generic name that describes a set of tasks by which a translator modifies language content that has previously been converted from a Source Language into a Target Language by a Machine Translation system, in order to make it conform to the objectives defined for the Target Text. The set of tasks required for the modification of the machine-translated content may include translating, editing and revising.
>
> **Post-editing** may be identified as being composed only of editing, but this is only possible if the purpose of the Target Text can be achieved by performing only the four technical actions (deleting, inserting, replacing and moving) over the machine-translated content, within a defined editing effort threshold.

4.2.1. What post-editing is not

The discussion of what exactly PE is will be the subject of the next chapter. For now, let us take an approach similar to pre-editing and clear our way through a few misconceptions about PE, before we move on.

There should not be such a thing as "monolingual PE"

There are many papers that consider and discuss a mode of PE called "monolingual PE". Oliver Culo and colleagues call this "editing", or "blind editing", in contrast to bilingual PE (Culo, Gutermuth, Hansen-Schirra, & Nitzke, 2014). The motivations for this are varied, including the need for voluntary help in non-commercial scenarios, but the main motivator is reducing the human resource cost in typical HT environments. In this mode, an MT text is either post-edited by a bilingual user with no access to the ST, or by monolingual users with no knowledge of the SL. An example of this situation may be found in (Koehn, 2010). In this paper, Koehn shows that monolingual users have been able to present results in which, at the most, 35% of the sentences were considered correctly translated. He concludes that this reveals that: a) MT systems bring across enough meaning to enable monolingual PE; and b) a monolingual user with good language skills and domain knowledge can perform PE with MT assistance at the level of a bilingual translator. It is arguable if you can draw such conclusions from the results, but they were widely adopted by research and the industry. However, the main point of interest lies in the assumptions of the TP that sustain such a mode of work.

These approaches to the TP basically assume that the translation is fully completed by the MT system, and that PE is a very secondary task, and an undemanding one in terms of language skills. Furthermore, it assumes that PE is a type of revision, performed over a complete translation. The implications of such a view are numerous. Let us discuss three of these implications, which reveal that this is a flawed perspective of the TP, and that monolingual PE should not be considered as a reasonable process to produce a translated text.

Several authors look at the four editing actions as "adequacy errors", i.e. errors of meaning transfer (Pal, 2015, p. 13). However, as other authors recognise: *"...since adequacy and fluency are not independent, degrading the fluency of a sentence can often negatively impact the adequacy as well."* (Parton et al., 2012, p. 114). The inter-relation of these two dimensions is indirectly revealed by a comparison of two different MT technologies. As (Castilho et al., 2017) comment, SMT and NMT, although both relying on ML methods applied to bilingual data, produce very different types of errors, with NMT producing more adequacy errors than fluency. The challenges these different types of errors pose (adequacy errors being more difficult to identify, as Castilho's study shows) is evidence of the inter-relation between adequacy and fluency, and of the need for specific skills to be able to post-edit a text competently.

The second dimension relates to the simplest description of the TP, which divides it into three phases, one related to the ST, the other related to the transfer process and the last linked to the TL. A simple intuition from this description is that errors may be produced in any of the phases of the process. So, no error-reduction strategy can claim to achieve its purpose with efficacy if it cannot access errors that were produced in any of these phases. Monolingual PE cancels the possibility to spot errors in two of the three phases: the one that derives from the interpretation of the ST, and the one that does the transfer process. So, this should not be considered as a reasonable error-reduction technique.

Finally, as we have seen in Chapter 2, revision is a fundamental part of the TP. One of the main reasons for this is the increasing parallelism between the ST and the TT that results from processes much conditioned by technology. For the TT to comply with its purpose, it is often necessary to break some of that parallelism and let the TT move away from the ST, and this can only be achieved at the revision stage. When revision is done by a professional, he knows how to balance all the conditions for a safe journey: clear traffic lane markings imposed by the ST, a vehicle that he knows well, and good planning for the time of arrival. This image easily reveals the risks of monolingual PE: without access to the ST, or knowledge of the transfer process, the monolingual post-editor only knows the destination and arrival time of the journey. He is basically driving a car for the first time in a road he does not know, with not even clear lane separation markings.

So, I argue that monolingual PE should be simply dismissed as an unreasonable approach to the TP and to PE, and that PE should always be regarded as a bilingual task that implies knowledge of the TP.

The distinction between "full PE" vs. "light PE" is not solid

The distinction between these two PE modes is often made in terms of "expected quality", accompanied by the common assumption that the quality level is negotiated between client and translation service provider. Some even argue that the definition itself of PE depends on this negotiation of the quality level. However, this is another grey area that is worth discussing.

For a discussion on quality levels to be possible, they have to be distinguishable in terms of objective metrics. On account of this, there have been numerous attempts to build and apply error lists and to define levels of admissible errors for each quality level. In fact, in research papers and industry guidelines, these two levels are always defined either in terms of general instructions or as evaluative interpretations from results of post-edited tasks. However, for a rigorous discussion on quality levels to be possible, not only must these categories be consensual, but they must also be related to a method of work. Words and expressions used in this domain show that this is clearly not an area where clarity reigns: “medium quality”, “good enough”, “not ideal” are common expressions which show how far we are from measurable and objective definitions of quality.

Another problem with these quality levels is that they are focused on the elusive dimension of the quality of the product, which had, as referred to before, been erased from the quality discussions in the industry, in favour of the most objective notion of service processes. The notion of product quality associated with a quantifiable number of errors comes from metrics such as TER. However, one needs to remember that this was not the purpose for which they were created, but that the purpose was to study how the text transformation from the draft MT to the result of PE is processed. As the authors of TER noted, the “e” in TER should be “edits”, not “errors” (see section 3.5.1 above).

The importance of not focusing on errors takes us further into the professional dimension: a translator needs to be able to counter a claim, on objective terms, not on a classification of errors, which may not hold in a discussion between two experts advocating in favour of two contending sides, but by proving that he did everything within his reach, following industry-approved processes, to do his job. So, until we have standards that describe how PE must be done, these two quality levels are not bringing any clarification to the provision of PE as a language service.

One final note on these two quality levels of PE. Krings pointed out that the usefulness of light PE is questionable. As a process to produce “translation for assimilation”, light PE does not seem to add enough value to make it a commercially interesting service, in view of the fact that most of this need is adequately served by free online MT services, like Google Translate (Koehn, 2009a).

In my view, the distinction between light and full PE has not proved its usefulness, neither as two distinct services of equal commercial value, nor for research.

One may see it as just a reflection of the pressure on productivity that comes all the way from the birth of MT, with the reduction of human effort that it envisaged.

For research, theoretical discussions on these two levels bring with them too many blurred classes and concepts, such as those that separate "obligatory" from "optional" edits, the opposition between "grammar" and style", and, above all, the distinction between "correcting" and "improving". These ill-defined dualities are some of the open themes in TS that its closeness to CS seems to have only intensified, and which are shaking the stability of the efforts to bring objectivity to PE and TS.

Evaluation of MT output and MT improvement should not be part of PE

We have seen in section 4.1.1 that TAUS considers that assessing the results of MT and helping improve MT systems is part of the role of a post-editor. This has led to an increase in the number of quality reports and evaluation systems that are integrated in the tools that translators use – the TAUS Quality Dashboard (TAUS, 2016b) is currently implemented by all major CAT tools. As mentioned in section 2.5, it is not clear whether this will integrate the full package of services that translators provide. However, this should not be confused with PE, and instead it should be seen as a separate service, for the reasons explained below.

There are several examples of goals, rules or standards that everyone else should try to achieve, but which were created in very specific situations. The "10,000 words per day" productivity with a MT system, for example, was reported in a special context, with hired translators at PAHO. However, the technological and work context in this institution is privileged and it cannot be easily reproduced elsewhere.

What we have currently in the translation industry is quite different. Most of the translation effort is currently being performed by freelance translators, who are paid on a word basis, in a market that puts a massive pressure on cost-cutting. PE is a service already conditioned by this, with a very narrow price scale, so it is not sustainable to expect new time-consuming tasks to be added to it. There have been discussions of changing the price rate, from words into time (Zetzsche, 2016b), and that is a change that needs to be considered. However, that does not disguise the fact that PE is already an unrecognised specialised and complex task, and adding new skills to it only obfuscates this.

Besides, improvement of MT systems is an expensive investment, in terms not only of human resources, but also of access to data and technology frameworks. The costs of tasks that are not directly connected to the production of translated products can only be supported by the companies that have access to those resources.

PE is not limited to editing

According to the traditional view of PE, the post-editor only has to perform the final stage of the TP, which implies adapting the way the information is conveyed by the TT in order to make it conform to the rules and norms of the TL. These low-level editing tasks are deemed adequate because everything related to the TP should have already been performed by the MT system. However, as will be discussed later in this dissertation, (section 5.2) PE implies a lot more than just editing.

The identification of PE with editing is so restrictive that it could be used to distinguish PE from revision. For example, in PE, research is virtually not allowed, or highly controlled, whereas it must be recognised as a fundamental task in revision.

So, either PE is seen as an ancillary task, complementary of a TP, and, in that case, it would be a synonym of editing, or it is an autonomous process, from translation, and from revision. I argue that we should leave the designation "editing" to describe a technical dimension of translation, revision and PE, as discussed in this dissertation, and we should consider PE as a full service or process.

PE is not a form of revision

The majority of papers on PE consider that this activity is a form of revision. This notion was born at the same time as PE, in the early days of MT technology. The arguments in favour of it are strong. To name just one, several recent papers on PE show that most of the time spent on PE is in pauses, not on actual technical action (Green, Heer, & Manning, 2013; Koehn, 2009a; Ortiz-Martínez et al., 2016). This could put PE on the side of revision, as it more associated with reading that with writing (see section 2.5). However, I argue that the technological evolution and the current context in which PE is performed challenge that view, as will be described in section 4.2.2. The focus in this section is only on the factors that distinguish PE from revision.

The first argument against PE being identified as a type of revision is the argument presented by translators when confronted with a "translation" that needs to be revised and then realise that it has actually been machine-translated. Translators properly refuse the classification of such assignments as revision on the argument that these jobs imply more translation than revision.

Moreover, if we compare PE with the revision tasks in the TP, we can see that PE is neither a self-revision, nor the revision of a translation made by another translator. It is clear to see why it is not a self-revision: the post-editor did not translate the MT text. And it is not revision either, since the textual elements with which the post-editor works are not "translations", but "MT hypotheses", "MT outputs", or "translation suggestions". The post-editor must recognise the special status of the text he works on when he accepts a PE job. If he approaches an MT text as if it had been translated by a human translator, and accordingly expects to work on a high-quality translation, he will most likely do a bad job and miss most of the detailed errors the MT text contains.

So, the conditions to have an RP are not met by PE, the most important one being that the post-editor does not have a translated text to revise. But the same happens at the end of the PE process.

As result of the PE process, it is often assumed that one obtains a translated text, as in (Forcada & Sánchez-Martínez, 2015, n. 1). Such a reading not only implies that, during PE, the translator performed a TP, but it also defeats the idea that the translation had already been completed by the MT system. This degree of redundancy, with a translator doing a translation for the second time, would be unacceptable in such a demanding environment in terms of efficiency.

However, it is not proven either that, at the end of the process, there is a revised text. During PE, there is often no self-revision, especially during tasks whose specifications restrain production time and expected quality, and in which such a task would be considered redundant. In fact, in current industrial workflows, it is not uncommon for post-edited texts to be fully revised before delivery to clients, because that is the only way to guarantee their fitness for the purpose.

DARPA's "Handbook of Natural Language Processing and Machine Translation" describes the GALE program, which is considered a good practice standard for many procedures involving MT. In this handbook, the description of PE production is a two-stage process that includes a revision pass by a second translator (Dorr, 2010). Finally, in one of the most complete experiments with interactive editing modes within the CasMaCat project, which was performed with the collaboration of a commercial company, the outputs were revised by different translators after the PE tasks, (Sanchis-

Trilles et al., 2014). The purpose of this revision was not evaluation, but it was intended as a quality assurance step. The authors explain that: *"We will refer to translations as produced by PE or (A)ITP as (edited) draft translations (as opposed to raw MT output), and to their final versions after review as final translations."* (Sanchis-Trilles et al., 2014, p. 224).

So, the arguments against PE being a form of revision are strong too.

One note on this apparent obsession with classifying and differentiating the results of different processes, or processes that include, or do not include certain phases. Although this approach may seem exaggerated, it derives from good process management practices: if a different process, or a change in a process, does not result in a different product, then one should not have a separate category for it. Having different processes clearly described in terms of what distinguishes them, and each resulting in clearly different products, simplifies production flows, and improves system and process efficiency.

Post-editing is not correctly modelled by predictive writing

Several studies in IMT test ITP against PE (see section 3.4). In these experiments, PE is considered to occur when the translator only has one MT hypothesis per sentence to edit, whereas in ITP a translator types the translation with the help of dynamic predictive writing suggestions. This should be a good indication that predictive writing is not the most appropriate model for supporting PE work, or at least not the best model for all editing actions.

When a translator is doing PE, he tries to spot as many errors as he can in one reading of each MT segment. Then, he decides on how he will correct those errors, in the places where they appear. So, he clicks on scattered words in the segment and applies different techniques to correct the segment.

This is one very important difference between writing a translation and editing it: if a translator writes a sentence from left-to-right, even if he has suggestions for each word he types, the translator is generating the whole translation in his brain, and the interactive aids he gets depend on the sentence he has generated in his brain. In IMT systems based on ITP, the suggestion that he sees ahead changes as he is typing new words that had not been predicted. So, his mental processes are divided between his mental representation of what the translation should be and the metamorphosed suggestions he sees ahead of his writing. This shows that predictive writing modes may require specific conditions not to interfere with the mental processes of translation. These mental processes are very different in editing, and different tools are required.

After an analysis of the difference in revision methods between novice translators and experienced ones, Michael Carl and colleagues state the following:

> *If certain translation and post-editing strategies turn out to be more successful than others, as in the case of our professional translators, then they should presumably be taken into account in the design of translation support tools. Under this assumption, a MT post-editing tool seems to be better grounded than a translation completion tool [1], which would mix drafting and post-editing phases, as we have observed in novice translators.* (Carl et al., 2010, p. 11)

In 2016, a study in which Michael Carl participates showed that scattered editing, smaller production units, and a special eye focus on the target side are features that characterise translator behaviour during PE (Carl, Lacruz, Yamada, & Aizawa, 2016a).

These are sufficient arguments to argue that PE is not adequately supported by ITP or predictive writing systems. However, these methods are important supports to, at least, two editing actions: insertion and replacement. So, it is still relevant to continue studying these interface methods.

4.2.2. A description of post-editing in practice

Post-editing with CAT tools

PE is often described as a task performed from the beginning to the end of a text. However, in a typical translation workflow around the use of CAT tools, that is a rare situation.

Before CAT, translation aids came in the form of dictionaries and reference works, like encyclopaedias and specialised manuals. The syntagmatic axis was created by the translator as part of a TT generation process that was only interrupted to look for a word and, maybe, rephrase a sentence. With CAT tools, this process not only became faster, but, like Pym says (see section 2.3.5), the paradigmatic axis took over: each sentence, each word has a position that is defined by its alignment to others, be it the source, or alternative sentences and words with which it shares some similarities. A modern CAT tool presents suggestions for terminology, it allows searches of words in context, and it shows predictive writing suggestions. If one considers the fuzzy match composition of most recent tools, then it is just another resource that contributes to the construction of the TT.

One of the effects of this environment is that working modes have become more sophisticated. The duality interactive vs. pre-translation mode that Wallis and García mention (see section 2.3.2) no longer describes adequately all the possibilities translators have.

In this context, the inclusion of suggestions from MT is not a major disruption, but a natural evolution. MT text is just another resource that translators have to read, in order to compose their translations. The increased cognitive reading load has to be balanced with the decreased technical load. Instead of writing, or overwriting the whole translation, the translator should be able to edit it at specific points. But usually this can only be achieved in parts of MT texts, not from beginning to the end.

Editing and translating in PE

Let us recall Table 1, presented in section 2.4, to focus on the specific types of segments presented by CAT tools and how they are processed in terms of the writing task and its subtasks:

Phase	**Window**	**Textual element**	**Task**	**Subtasks**
Orientation	Source	Source text	Reading	Reading
	Target	TM target text (full and fuzzy matches)		
		MT target text (no match)		
	References	Terms/Words	Researching	Researching
Drafting	Source	Source text	Reading	Reading
	Target	Empty window	Writing	Typing
		Source text		Overwriting
		TM target text		Editing
Self-revision	Source	Source text	Reading	Reading
	Target	Own target text	Checking	Validating / Editing
Result of the process		**Translated text**		

Table 4 – Technical details of the Translation Process in CAT

The discrimination of the type of matches in CAT tools is a fundamental guide to understanding the working modes that have replaced the duality between the interactive and pre-translation mode presented in section 2.3.2. A more detailed description than the table above might discriminate how each type of match behaves in all phases of the TP (orientation, drafting and self-revision). For instance, fuzzy matches are read during orientation, they are edited during drafting and they are revised in self-revision. But let us concentrate on the drafting stage. For a more refined description, a discussion on percentages that define and separate each fuzzy band might also have to be made. However, these vary a lot in the industry, and the values presented below are quite common.

Choosing segments from TM or MT

As we have seen before (see section 3.1.2, on EBMT), SMT, and any other form of MT that relies on estimates, can very rarely produce 100% matches against the translations in the TM (or the training corpus). So, the industry will not replace TM for the leverage of these cases. If we go down a level in the fuzzy match scale, to the highest level of 95%-99% match band, we find segments which usually require small editing, in very repetitive patterns: correcting tag placement and punctuation (usually due to changes in text coding and segmentation filters), or replacing a word here and there. It is also very difficult for SMT to compete with the competence of CATs in dealing with such cases.

For the intermediate fuzzy bands (from 94% to 75%), the advantages are not so clear. Fuzzy matches from the TM are still preferred, especially in complement with composition methods that replace unmatched bits in the sentences, but SMT may present enough quality to replace some of the lowest matches. Ana Guerberof Arenas reported productivity gains from MT use in a high band, between 80-90% fuzzy score (Guerberof Arenas, 2008). The work that translators do in these fuzzy matches is the traditional editing: the CAT tool has a pane with a visualization of the edited parts in the source segment, the previous target segment is already in the target window, and the translator edits this target segment making sure that the edited parts are updated.

The sentences that have match scores below 75% are considered “no match”, and in CAT tools they were left for the translator to translate. This is usually the biggest bulk of each project, and this is where MT replaces the TM. Typically, in CAT tools, the ST is inserted into the target window and the translator overwrites this text. In a

CAT tool fed by MT system, the translator will find the MT text in the target window, for him to edit.

Translations for internal repetitions and fuzzy matches (segments that are repeated or are similar in the ST) are propagated throughout the text. In the case of internal fuzzy matches (sentences that repeat with small variations within the text), these are propagated but they need to be edited, or recomposed.

The following table sums up the previous description, presenting the content of the target window and the task performed by the translator. The only level at which things change with the arrival of MT is highlighted.

Segment	CAT	Task	CAT+MT	Task
Full matches	TM full match	Check	TM full match	Check
Fuzzy match	TM fuzzy match	Edit	TM fuzzy match	Edit
No match	Source segment	Overwrite	MT hypothesis	Edit
Internal repetitions	TM propagated	Check	TM propagated	Check
Internal fuzzies	TM recomposed	Check	TM recomposed	Check

Table 5 – Types of segments in CAT and CAT with MT.

So, the only level at which there are changes with the arrival of MT is that of "no match". Truth be told, this is usually where the highest numbers of words in each project lie. This is also where the two of the writing modes were selected: either users overwrote the ST, or they typed the TT from scratch.

But the no match segment band is not homogenous: it includes everything from segments in which no words match the TM, to segments in which up to 74% of the words are matched against a segment in the TM. With MT content, the users can no longer type their translation and they cannot overwrite the ST either: they have to edit the MT text. So, this is where editing is supposed to take over from translating.

One must admit that, within such a wide range of similarities in no-match segments, there is the possibility that a fuzzy match with a higher similarity retrieved from the TM might require less editing than an MT hypothesis. But even if one accepts that all segments within this range are to be pre-translated by MT, it should be clear that there is a high probability of a wide variation in the editing work that will be necessary to transform these segments into quality translations.

Self-revision and research

The focus in this dissertation is always on production, on the so-called drafting phase. However, as has been repeatedly mentioned, these phases are not separate from one another, and one cannot assume without discussion that PE does not go through any of these phases. When most descriptions of PE speak about limits to research and do not refer to any revision, there is the underlying assumption that none of these tasks is performed. But that cannot be true in reality.

Research needs are not proportional to the purpose of the TT, so research effort cannot be gauged from a task specification based on this. Research needs are proportional to the difficulties posed by the contents that the translator works with. As we have seen in previous sections, both on the TP, and on PE, quite often, the more available references there are, the more queries and doubts translators have. So, an MT hypothesis may not reduce research time, and in fact it may even increase it, because translators need to check more things. This may also lead to increased inconsistency, because the translator is reading too many things, and making different decisions in each context.

The same goes for the revision phase. The frustration that is usually associated with this activity is not so much derived from a high self-esteem, but from the notion that some of the problems in the final TT can only be attributed to how the work was done. The way the work is done is the sole responsibility of the translator. And the translator knows that if there were an analysis of each decision, he would have to admit that he should have paid attention to whatever mistakes there are in the final result. So, while he is doing a PE task, or at the end, all the available time he may still have will be applied to a strict priority list of issues that need to be revised and checked before delivery. Consistency may be one of the last items on that list, but when quality checks are done, all eyes are attentive to problems like that.

So, it should be acknowledged that research and revision are part of the PE process.

Fragmented content

Let us bring here another issue that affects the results that may be achieved through this more economical way of producing a translation. Most metrics are based on numbers of words, matched, edited, in correlation to time, and so on. This leads to several assumptions that are challenged by the practice of PE.

One of these assumptions is the correlation between difficulty and segment length. There are several metrics that correlate text complexity with number of characters or syllables per word, and other similar indices, but as one can understand from text compression, that is the kind of metric that is language-bound, and cannot be generalised from one language to another (Fonseca & Alves, 2016).

However, in general, these metrics are based on the analysis of proper texts. And, as Pym has mentioned in "Translation Skill-Sets in a Machine-Translation Age" (Pym, 2013) text integrity is long gone, and nowadays translators work with fragmented content, coming from content management databases that distribute separate segments to webpages, manuals, advertising material alike, even software strings.

Due to its association with economic gains, PE is regularly used in projects that are made up of this type of fragmented content. In these projects, there cannot be a proportional relationship between length and difficulty. The main difficulty comes from the lack of context for each decision. There may be high productivity gains in such projects, because translators restrict the research time devoted to them and, unwillingly, validate all segments that show no particular grammatical and "meaning" problems. The reason why they validate these translations is because even if they could research and revise, they would have no context or support material to improve on the decisions.

So, fragmented content is at the root of a misconception of the value of PE, based on the relation between little time devoted to research and revision. In reality, the fragmentation of content is removing value, linguistic and commercial, from the transfer process between languages. With this, it also cancels the possibility of extracting accurate knowledge of what is going on in the decision process, and it invalidates conclusions such as those that relate difficulty to sentence length.

Post-editing and the translation industry

It was mentioned before that the translation industry does not publish information on its technological advances. It is, nevertheless, very vocal about the changes that it would like to see in the market. At present, the industry is facing complex challenges: it tries to offer competitive high-technology services, without losing the quality that has given it its major successes. In trying to maintain that balance, it is pushing the pressure on to the translation service providers, by trying to reduce costs while requesting the same quality. So, there are a lot of mixed messages, some of which seem to indicate that companies are willing to jeopardise the quality they were offering, in favour of not losing the productivity gains brought by technology.

In the section on revision, reference was made to the "four-eyes strategy". Translation service companies will only give up on offering these high-quality levels of services if, or when, they are sure that the eyes of the only "human in the loop" have the adequate conditions to offer a good final piece of work. Otherwise, their business may no longer be sustainable, because translation will have become a service with a very low value, in which the major costs will no longer be human. So, the focus on the development of adequate tools for translators may be the strategy to allow for everyone to cross the PE intersection safely, and to continue to make sure all messages are understood.

Translation, productivity and quality

Translation practice has always been a demonstration that productivity and quality do not have to be incompatible. Translation has always been defined by efficiency, as Sager pointed out. Indeed, an inefficient TP does not comply with its purpose, and experienced translators are specialised in making the best decisions in balancing efficiency and quality. When a translator crosses a line from PE guidelines, his only sin will be over productivity and money loss, because it is unlikely he will ever be accused of presenting "too much quality". And, as translators know that achieving constant increases in productivity depends mostly on using the right tools, they feel that it is their responsibility, and no one else's, to look after translation quality. This is a preoccupation that is strongly felt in the training of new translators.

Educating for PE

The discussion on whether the translators that do PE should be called "post-editors" or not is not a very important one, and in this dissertation both names are used interchangeably. But the most important point is that educating new generations of translators to include PE in their many skills should be regarded as a specialisation, not as a downgrading of skills. The "educate not to translate" that the PE current paradigm seems to defend is a step in the wrong direction.

In "Translation Skill-Sets in a Machine-Translation Age", Anthony Pym discusses the influence of MT on translation teaching (Pym, 2013). He begins by admitting that the way TS specialists viewed technology has had to evolve, so that they could encompass the increased importance of technical skills over the years. The author also explains that there is a shift from generative skills to selective skills, to identifying the best solutions for translation problems in a multiplicity of resources. This skill became so important that knowing how and where to find knowledge is more important than retaining it. So, let us focus on how research is helping translators manage all the resources at their disposal.

4.3. Second challenge: supporting post-editing

After you understand how you need to get through the complex intersection to your destination, you need to be sure that you have all the instruments at your disposal to get to the other side, and avoid your car stopping in the middle of the intersection. You also need to make sure you pay attention to everything around you. So, the second challenging level is whether the tools you use give you adequate conditions to perform your task, and how you actually use them.

In this section, we review how editing is seen from the tool development point of view. As mentioned before, ML plays a very important role in meeting this challenge, especially in terms of the inclusion of MT content into CAT environments.

4.3.1. The need for specific editing interfaces

When a translator approaches an MT sentence, he takes it as a sentence already written in his TL. So, he approaches it by reading it fast. In fact, maybe he just scans through it, not really reading it from left-to-right, but just looking for the existing errors that he must spot. He does this in an internal conflict between being pushed forward by externally-imposed productivity goals, but being dragged back by his trained and required focus on small quality details. This conflict does not bring with it the clarity that so often is described by error lists: these lists bear the underlying assumption that in this process it is possible to use an objective red line that simplifies the decision of dismissing one error and picking up another. On the contrary, in the process, the translator quite often makes decisions he would like to revisit, if he could.

After he has cleared through this conflict and identifies an error, he decides how he will correct those errors, in the places where they appear. So, he clicks on scattered words in the text, and he applies different techniques to correct the words that transport that error. He does this with spot-on, surgical interventions, based on detailed keyboard and mouse selection actions over the words selected.

There is one very important difference between writing a translation and editing it: if you are writing a translation from left-to-right, even if you have suggestions for each word, the whole translation is being generated in your brain. And the word "generation" here is Chomsky's notion present in the generative grammar, and the role syntax has in this: all syntactic relations that construct the surface of a sentence come from an abstract notion of its intended meaning and purpose, and these are first and foremost being produced in the translator's brain. That is the reason why there is such a strong interference in terms of cognitive load when this generation process has to be confronted with a sequence of ever-changing suggested completions for the sentence. In editing, the generation process is not totally triggered, as happens in simple reading tasks.

To be "editable", the sentence presented by the MT system must be good enough for the translator to only worry about certain scattered points that may be corrected through the application of well-directed actions. However, that is not enough. The interface elements, i.e. the mechanisms that build the communication between the systems and human actions, must also be improved in order to provide the necessary conditions for the editing to proceed in an efficient way.

The need for improved interfaces

In the TS literature, there have been several calls for improved editing interfaces. A very recent example appears in Adrià Martín-Mor's paper, in which the author suggests several ways in which a system like MTradumàtica may help new editing interfaces (Martín-Mor, 2017).

Lucas Nunes Vieira and Lúcia Specia presented in 2011 "A Review of Translation Tools from a Post-Editing Perspective" (Vieira & Specia, 2011). In this paper, they compare nine translation toolkits, from commercial ones to others conceived for academic research. They compare their offers according to ten criteria that go from interface intuitiveness, to the existence of spell/grammar/style checkers, and the combination and identification of sources of content. In terms of interface, the authors focus on features that include:

- colour codes to identify fuzzy match levels,
- side-by-side visualisation of source and target,
- keyboard shortcuts,

- find and replace functionality,
- concordance searches,
- presentation of statistical figures, related to fuzzy match scores and MT text reuse.

In their analysis from the point of view of PE needs, a tool stands out: Déjà Vu. There are several reasons for this choice: the fuzzy match repair functionality (built from a close integration of TM and MT); the fact that suggestions are retrieved from different sources; the mechanism that sorts these suggestions, based on fuzzy scores, but also on project management features, such as client and text domain.

The authors present a list of features that, in their view, are desirable in a toolkit adapted to PE. These include: support at the sub-segment level (like on-screen phrase alignment); control features, like tracking of keystrokes (which may be implemented in a similar way to a word processor's "track changes", so as to assist revision), and the indication of quality scores of MT suggestions.

In 2015, a team from the University of Malaga presented a proposal for the classification of commercial CAT tools, in terms of the integration of MT and TM technologies (Zaretskaya, Pastor, & Seghiri, 2015). This is a good source for a state-of-the-art analysis, as it includes technologies such as the recent "adaptive MT". In this dissertation, there is no analysis of this technology, because it concerns a procedure that runs in the background, and our focus is on the interface.

Joss Moorkens and Sharon O'Brien conducted a survey on "User Attitudes to the Post-Editing Interface" (Moorkens & O'Brien, 2013), which had a surprising result during the data collection: in 2013, 40% of the respondents used Microsoft Word to post-edit MT content, not taking advantage of either CAT tools or MT potential. The replies to questions on which features the users would like to have in their PE tools was also surprising due to the fact that there were virtually no answers specifically directed towards PE, revealing that the distinction between the two activities, from the technical point of view, is not so clear for translators. Most replies voiced general preoccupations with translation aids, such as the propagation of changes, global search and replace features, and the like. In the section on the use of shortcuts, and the use of MT content, respondents mentioned as useful the possibility to delete a whole MT suggestion with one simple action. They commented on the possibility of using shortcuts associated with language-specific actions, such as correcting word number and gender, and on how this has to be well balanced in terms of usability.

The survey also addressed the combined use of content from TM and MT. Respondents showed interest in visual representations of confidence scores, which is related to the apprehension over MT quality. They also recognised the usefulness of indications on the provenance of the suggestions, a theme that has been studied by Carlos Teixeira (Teixeira, 2014). The authors also comment on the possibility of combining MT and TM to compose matches from sub-segment pieces, a type of match they call MTM – Machine Translation/Translation Memory match. One of the questions asked specifically from which level of fuzzy match the users would prefer to see an MT suggestion to a TM fuzzy match. The lowest fuzzy match level (65%) was the one most users chose. However, in total, ca. 60% of the users said that this threshold should be in a value below the 75% fuzzy match score. The final theme referred to in the survey concerns the reuse of the data from translation work by translators to improve the client's MT systems.

This survey was later extended with interviews and the full report from the experiment was published in 2017 (Moorkens & O'Brien, 2017). The purpose of the whole study was to create a list of specifications for PE tool design. The interviews focused more specifically on PE features, with the global concern of the balance between feature richness and simplicity being one of the global themes. The study addresses specific linguistic changes. For some languages (Portuguese being the example provided), word order is a recurrent problem, but users recognised the difficulty in implementing this edit in a way that compensates over current keystroke-based actions.

Let us now look at how different tool development projects addressed these challenges and, in the last section, discuss which challenges are still open.

4.3.2. Early proposals for editing interfaces

The integration of CAT tools and MT systems started quite early. In 1996, Mattias Heyn presented a sketch of how MT content could be integrated into Trados, both in a batch and in an interactive mode (Heyn, 1996). The author recognises that an interactive system has implementation issues, particularly because of conflicts between the MT lexicon, and the CAT tool's terminological database.

In 1997, Martin Kay's "The Proper Place of Men and Machines in Language Translation" describes the famous "Translator's Amanuensis" (Kay, 1997). This is a "text editor" which acts as the central application in a system that takes "baby steps" before presenting MT sentences that fully replace human translation. These baby steps are essentially editing aids, which allow the user to click on a word, a phrase or a sentence, and act upon it. Little by little, the translator replaces the MT suggestion by his own choices. After the translator's selection, the system gives him support through a dictionary and, if requested, it may present linguistic information on the word selected. The changes made by the translator may be stored temporarily, or permanently, and shared with other translators. Kay emphasises the main advantage of this system: "*This somewhat unconventional procedure has the advantage of making it possible for the machine to maintain detailed linkages between the original and the translation so that it has a detailed idea of what corresponds to what."* (Kay, 1997, p. 14). Besides these alignment capacities based on user's actions, the system also anticipated propagation of phrase translations, supported by morphological rules, and lists of alternatives. Kay underlines that such a system gives power to the user. In his model, man and machine cooperate not just to produce a good quality translation, but also to improve over time, with the machine being capable of distinguishing that the newest decisions override older ones. In his project, the translator is the sole validator of the produced content. This tight link between translator and his computer editor seems to have been lost in the wave of big data, and with it the capacity for the system to learn from users' actions too.

In 2000, a team from Canada presented TransType, which presented itself as: *"the embedding of a statistical translation system within a text editor (...), a system that watches over the user as he or she types a translation and repeatedly suggests completions for the text already entered"* (Langlais, Foster, & Lapalme, 2000). This system is very often presented as a good example of an early IMT tool. The project was shut down in 2005, and one of the main criticisms it had received was the incapacity to learn from translator's edits and to eliminate repeated errors (Macklovitch, 2006).

Another project that garnered a good reputation was PAHOMTS, as mentioned before (Aymerich, 2005). This system is based on an RBMT engine, and it serves only three languages: English, Spanish and Portuguese, in different combinations. The interface is based on Microsoft Word macros and keyboard shortcuts, and it allows users not only to add and correct entries in the dictionary, but also to perform other automated actions. Examples of these actions are: looking up words in the dictionary, moving words left or right, deleting words, creating NP compounds, deleting articles, replacing adverb endings, and so on. This system was used for more than 30 years, becoming an integrated environment in which translators and engineers worked and developed the system together. In 2009, it featured several modules, which included not just the RBMT system, but a request management module, a feedback mechanism and research tools. The authors that presented the system in 2009 considered that the interface between the system and the translators was the main piece of this environment: *"This is the most important step, as the main cost of translation is largely determined by how efficient an environment is provided to the translator."* (Aymerich & Camelo, 2009).

4.3.3. Editing tools for research purposes

Caitra (Koehn, 2009a, 2009b) is seen as a model for the development of interactive editing tools. Caitra presents suggestions retrieved from the MT phrase table underneath the typing window. Using a beam search, it presents the different alternatives for each word or phrase, ranked in colour-coded boxes. The user edits the sentence by either typing his own words, or selecting and editing the suggestions below. The system also shows the results of the edits made, in a typical "track changes" interface. In the end, Caitra collects information on all edits, namely in terms of keystrokes and time per edit. This data is invaluable for the researcher to compose translator's profiles and to study efficacy of editing methods.

There are several other PE tools that are essentially data collection tools. Examples of these are Translog (Jakobsen & Schou, 1999), Translog-II (Carl, 2012) and CasMaCat (Alabau et al., 2013), which have already been mentioned in the ambit of TPR projects, but others like PET (Aziz, Castilho, & Specia, 2012), iOmegaT (Moran, Saam, & Lewis, 2014) and Transcenter (Denkowski, Lavie, Lacruz, & Dyer, 2014) should also be mentioned. Most of these tools collect user activity data (UAD) and they also have quality scores for user assessment of MT output.

MateCAT (Federico et al., 2014) also began as a research tool, but it soon became a popular online tool, which simplifies the use of MT in translation projects. The interface is that of a typical web application, with no specific support for editing, but it is complemented by a host of project and text resources management, like different analyses, TMs and glossaries.

CasMaCat has the same basis as MateCAT, but instead of developing as an online, commercial tool, it was focused on research and development of interactive features. Its advanced data collection functionalities have enabled numerous projects, especially in the areas of TPR and ITP. Besides, it has also given rise to other software packages, with different purposes. One of these is Thot, a toolkit that presents itself as an alternative to Moses, and which includes not just SMT tools, but also IMT features, incremental estimation of models and phrase alignments (Ortiz-Martínez & Casacuberta, 2014). The toolkit is available in the open-source model and it may be downloaded here: https://daormar.github.io/thot/. Another project that derived from CasMaCat is SEECAT, a study on alternative interfaces with CAT tools, namely by the incorporation of voice recognition as alternative to keyboard interfaces (García-Martínez et al., 2014). In this study, PE was modelled as a text unit replacement action only.

Still, above all, the biggest contribution of CasMaCat is the wealth of research publications the participants of the project produced, and the questions and assessment of editing features it involved. The final papers and reports that the consortium published are filled with invaluable inputs to this domain.

The paper "Interactive translation prediction versus conventional post-editing in practice: a study with the CasMaCat workbench" (Sanchis-Trilles et al., 2014) presents the result of an experiment with different interfaces and interactivity modes. The authors explain that they approach ITP as a work mode completely different from PE, and they use the term "editing" as the generic name for both. They test the use of an e-pen as an interface for use with mobile devices, and they concluded that this input device is appropriate for the editing of few errors, as with high fuzzy matches, or when a translator revises the result of a PE process. CasMaCat may be used with its basic interface, similar to MateCAT's, but more advanced features may be added to it. The list of advanced features includes:

- intelligent autocompletion
- prediction rejection and presentation of alternatives
- search and replace with propagation of results
- colour-coded visualisation of MT confidence scores
- limited prediction horizon, which the user may expand
- presentation of the word alignments with the source

- highlighted visualisation of the user's edits (to discriminate those from the system's suggestions).

The authors of this report recognise that the results, in terms of increase in productivity, were not as good as expected, but they associate this to the longer learning curves that complex systems imply.

One of the teams that worked in APE has recently presented a PE tool that features a very complete logging system to record user's actions in terms of the four editing actions, among many other options. The name of this software is CATaLog and it was presented as a means to collect data for project management, TPR and to train an APE system (Pal, Zampieri, et al., 2016). There are interesting interface features, like the colour codes that identify whether a word came from a TM or a MT system, but its strongest point is the comprehensive logging system. The system saves a separate log for each editing action (deletion, insertion, substitution and reordering). The system identifies these edits through TER analyses and keystroke data, which includes mouse selection, cursor positions, cut and paste operations, shift-arrow selection mouse clicks, etc. Besides the words that are edited, the logs include their position. The system uses a cost estimate for the editing operations, and considers that insertion and substitution are the ones with the highest cost.

4.3.4. From research to tool development

Several of the projects commented on in the previous sections made their way to full commercial tools. One such case is Lilt. In "The Efficacy of Human Post-Editing for Language Translation", the team of researchers that launched Lilt analyses the process from the typical tripartite view from TPR (Green et al., 2013). They compare two translation modes: unaided (although it is not very clear in which conditions), and post-edit (which is done with a "barebones" editing interface).

Their own collection of data from the experiments shows that PE is a good strategy to deal with MT content, and it reveals different translator behaviour. They classify the interaction pattern as "more passive", meaning that there are fewer interaction events with the software, and that users spend more time on pauses than on text generation. This behaviour is relatively constant, which does not allow for the identification of the three phases that are apparent in the translation task. Another effect of a different behaviour in PE is that edit distances between ST and TT are smaller in PE, which seems to reveal that the translator's choices are primed by the MT hypothesis. Some of the implications of these conclusions on interface design are worth emphasis:

- systems should not be designed for just one mode of work, inspired only by a gisting mode, a drafting, or a revision one, as these intermingle in PE;
- words, phrases and reorderings should be used as additional training data;
- the different user actions should be encouraged by the system in a way that enables a specific recording of their occurrence.

The way the authors implemented these conclusions in Lilt has already been discussed in section 3.4.2, in the context of IMT. ITP seems to be the paradigm they chose for their tool.

Interaction with the editing actions on a mobile device

Joss Moorkens, Sharon O'Brien and Joris Vreeke are involved in a project of a PE tool which applies the lessons learnt from their research projects in a novel environment: mobile devices. The application that was developed in this project is called Kanjingo, and it was presented in two different stages of development (O'Brien, Moorkens, & Vreeke, 2014; Moorkens, O'Brien, & Vreeke, 2016). The first paper discusses the motivations and objectives of the project, plus the trials and tribulations of the development and use of the application in a smartphone, the main ones being related to the size of the screen and the very simple forms of interaction. The application shows a sentence at a time, the ST at the top, and the MT hypothesis with each word in a separate tile in the centre of the screen, with action buttons above and in front of each word. The target translation appears after the user scrolls to the end of the screen, to see the whole sentence. Figure 7 below illustrates this interface in its implementation in 2016.

Figure 7 – Kanjingo's editor (O'Brien, Moorkens and Vreeke, 2016, p. 61).

The interface shows that the application is too dependent on the initial tokenisation, as it shows each word separately. This raises typical issues reported by users, such as problems with spacing, capitalisation and punctuation. The action buttons are linked to the two simplest editing actions: the (+) sign above each word allows the user to insert a word in that position, whereas the (–) sign deletes the word it is attached to. To edit the word (i.e. to replace it by a different form) the user may tap on the word and enter new characters. There are also references to swipe movements in one of the figures of the 2014 paper, but these do not seem to have been implemented. Ordering can be achieved by selecting the word tile and moving it to the desired position. So, the potential of the visual and mobile interfaces was successfully explored to model all four editing actions.

The version that was launched in 2016 allowed for the word replacement to be made by voice. Users' feedback was very positive, taking into account the characteristics of smartphones, but the authors admit that better implementations will be explored on tablets. Users' comments also referred to the need to see sentence contexts and to re-editing submitted sentences. The implementation of each feature was also commented upon, with users referring to the usefulness of editing phrases instead of just words, namely for easier reordering, and the need for character-level replacement, in order to avoid having to type a whole word just to change its capitalisation or spelling.

Kanjingo is an excellent and innovative example of an implementation of the four editing actions in an easy-to-use interface. However, it also shows that the problems to solve in such a scenario are not trivial.

Interaction with the editing actions on a keyboard

In 2016, there was another presentation of an interface that allows the user to apply the four editing actions, with specific keyboard shortcuts. The tool is called LFPE – Learning from Post-Edits (Simianer, Karimova, & Riezler, 2016), but, although it is available in GitHub, it has not been released or adequately documented for use outside the research lab. Information on the keyboard shortcuts, for example, can only be found in a help document inside GitHub: https://github.com/pks/lfpe/blob/master/inc/help.inc.php.

Like Kanjingo, LFPE shows words inside tiles. However, tiles may contain not just individual words, but also phrases, and there is a visual representation for the alignments between words and phrases of ST and TT, as illustrated by the following figure.

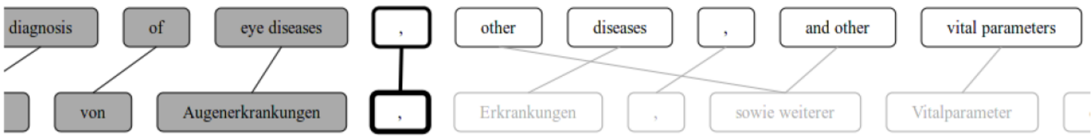


Figure 8 – LFPE's editor (Simianer, Karimova, & Riezler, 2016, p.3).

The alignment with phrases is a necessity due to the agglutination process of word formation used in German, but one may see in the figure that agglutinated words are not the only items aligned to phrases. Users interact with the system by clicking on a tile and highlighting the aligned word or phrase, and then use keyboard shortcuts: Shift+M to move that word to another position, and others to delete, replace, change or to insert a word next to the one selected. A user may also delete the alignment link of a word and establish a link with a different word. So, the interaction revolves around the concept of alignment and the four editing actions.

4.3.5. Open challenges for editing interfaces

The throughput of translation is nowadays often only limited by the capacity of the interfaces between humans and technology, as recognised by analyses of typing speed, mobile input interfaces, and voice recognition technologies. Technological development must be based on correct readings of the processes, at the risk of not being useful (in the sense that it solves problems) or usable (in the sense that it improves processes).

The current challenge for PE is not so much on improving the quality of the MT output, but on giving translators proper conditions for their job, including PE. However, the current views of how translators work are too focused on an ancient notion of equivalence (like Pym says), and one needs to look at translation shifts again, and allow for systems to accept and integrate them.

Sharon O'Brien approaches these challenges in a text entitled "Translation as Human-Computer Interaction" (O'Brien, 2012). As a counterpart to HCI – Human-Computer Interaction, she presents the term TCI – Translation Computer Interaction for a global name for these approaches. She discusses the challenges of interaction design, which studies the friction between humans and devices. After discussing the role of translators in the development of the tools they use, she focuses on the effect that the use of technology has on the definition of the TU.

CasMaCat's "Final Report on Interactive Editing" (Koehn, Saint-Amand, & Alabau, 2015) is a very detailed report on the different experiments and evaluations that were carried out in this project, correlated to methods to interactively support the work of translators, which they classify as a sort of "puzzle solving". The report is divided into different methods, and they would all deserve a detailed exploration, since the technical challenges are big and CasMaCat researchers spared no efforts in addressing them. However, we will focus on the section in which the authors analyse the capacity of the system to learn and predict edits.

Koehn and colleagues refer to the component that tries to learn and predict edits as "automatic reviewing". An analysis of edits in the test set identified a high number of replacements, followed by insertions and deletions – they use the Levenshtein distance, which does not detect reordering, or movement. Of these cases, they decided only to try to predict the insertion and deletion of content words. They use an eager search method, which results in much higher recall values than precision. They admit that *"However, it is not clear, how meaningful these metric scores are to judge the viability of the method practical use."* (Koehn et al., 2015, p. 18) After a comparison with an evaluation made through a community platform, they assume that the system reaches a precision of a mere 10%. So, they conclude that this is too low for them to recommend its implementation in practice, as it would lead to too many false alarms, frustration and lack of use. This is an early attempt to use such a system, so future attempts should not be excluded.

In “A general framework for minimizing translation effort: towards a principled combination of translation technologies in computer-aided translation”, Mikel Forcada and Felipe Sánchez-Martínez present a thorough analysis for the development of interactive systems that apply both MT and TM (Forcada & Sánchez-Martínez, 2015). Their focus is on how to select the best technologies in a complex CAT environment. They mention that such systems may be so complex that they may lose their interest as aids. The framework they present is clearly still a sketch that needs to be tested in practice. Furthermore, their suggestion is for a batch system, which needs to be expanded to online learning systems.

To close this section, it should be stressed that the recognition of the complexity of the challenges in this domain, by experienced researchers who, like Philipp Koehn and Mikel Forcada, have proved their capacity in dealing with very complex problems, is a strong evidence of the hard work that lies ahead. So, any presentation of alternative views, suggestions for improvements, claims or hypotheses brought forward in a work such as this dissertation can only be done with the humbleness to admit that the problem may just be too big for one’s capacities. Still, it is worth trying to tackle some of its issues.

4.4. Clearer views to the intersection

This chapter was an attempt to condense several views on PE, with different origins and destinations, that meet in the same intersection and try to make their way by respecting anyone else’s routes. At the end, the intersection is still cluttered with cars, but it seems easier to see which way to go and which tools to bring. This is a brief description of the routes taken in this chapter and what they showed us.

There is no way back from the integration of CAT and MT technologies into a single work environment, and there are strong benefits for translators from it. The input translators can bring to the discussion should come from their analysis of the processes they use, because only with that knowledge will they make sure the tools they use will be adapted to them.

The technology that enabled SMT to generate reasonable TL sentences is now being focussed on the task of helping translators do their jobs. It has been demonstrated that there are long avenues to explore with this technology. Moreover, it has been shown that analyses of PE place researchers at a great observation point regarding the technological transformations that may benefit translators.

Here are some of the technical details that should be further studied as a means to achieve the expected transformations of translation tools:

- unit selection and segmentation processes, like chunking, n-grams, or tokenizers, very often define the capacity for the tools to deal appropriately with phrases, a fundamental unit in the processing of translation data;
- flexible alignments may expand the learning capacities of tools, which nowadays are conditioned by the search for optimal equivalence;
- editing actions describe most of the work that happens in the tasks that may be described as editing;
- systems should learn not only alignments but also the actions that users apply to these aligned or misaligned units;
- user activity data is invaluable as support for translators' work.

The next chapter discusses the implementation of these features, and it presents a proposal for a system that supports translators properly, by learning from their actions and providing specialised information to each user.

5. INTERACTIVE POST-EDITING

Sustained by the views and analyses of the literature, research and experiments from both TS and CS, this chapter presents the theoretical grounds and practical questions that should enable the development of new tools adjusted to the specificities of the processes that have been analysed.

5.1. A few lessons and some guidance from previous chapters

From Translation Studies, central notions were collected, such as equivalence, translation shifts, efficiency (a central concept for translation in an industrial environment), besides analyses of the TP and methods of collection of data from processes.

The search for equivalence has sustained several views and theories of translation, mostly based on product data analysis. The identification and segmentation of equivalent units became a central concern, as it enabled the classification of units and the identification of patterns, leading up to the prediction of processes.

In a movement in the opposite direction, the notion of translation shifts may have been an argument for TS to focus on the analysis of creative processes. However, this notion is of major importance even for the mechanical treatment of translation data.

This dissertation has selected the technical dimension of the translation work as the preferred object of analysis, focusing on a particularly detailed manifestation of this dimension. This option was an attempt to contextualise and understand the practice of PE. The four editing actions have led us from Gideon Toury's first reference to very specific procedures in the context of translation norms, to computer interfaces that try to guide the translator's work with MT content in mobile devices. Such an extensive analysis based on a small detail is surely an evidence of its importance.

Of course, editing and the four editing actions are not a complete description of the TP, nor even of PE, but they show a dimension, which, when neglected, may lead to biased and misconceived approaches to such a complex phenomenon. The effects of this approach are of utmost importance when one is concerned about the support that is given to translators.

The identification of the phrase as a central TU for current translation tools has brought with it a new look at TP and PE. For example, it was demonstrated that the more experienced translators are, the longer the units they process. However, PE work has the effect of reducing the size of the units that translators process, because of the type of content they work with. PE tools will be useful if they offer a type of support that allows experienced users to extend their processes of work to at least the phrase level, or even across sentences.

A current difficulty in the processing of translation data is capturing long distance phenomena, such as gender or number agreement in morphologically rich languages. An example is a complex subject at the beginning of the sentence that links to an adjectival phrase at its end, or to an anaphoric reference in a separate sentence. If users are allowed to flexibly edit these occurrences, perhaps even in the same sequence of actions, there will probably be ways to link these actions and identify the relation between these elements.

The technologies that are used nowadays in domains such as Quality Estimation (QE) and APE (Automatic Post-Editing) allow for the identification of some of the traces of these connections. However, these technologies are focused on retrieving knowledge from product data, and miss the information from the processes.

Instruments of process analysis have to be more flexible than those for product analyses. Processes are not as stable as data, they are not easy to segment, and they are very sensitive to variation, according to task specifications, instruments used, and the individuals that perform them. Adding linguistic analysis to this data may often be a frustrating process too, as the effort put into research is not compensated for by gains in performance.

The lack of aligned units, at the sentence, the word or the phrase level, is often seen as measure of lack of success. However, as Alves and colleagues indicate, experienced translators manipulate levels of segmentation. This means that equivalence must be found at higher levels. So, if there is no alignment possible at the word or even phrase level, this perhaps means that the translator built an alignment at the sentence level, or at the text or message level. In between these extremes, are the translation shifts, viewed not as errors of alignment but as strategies to solve problems. Carl and colleagues mention the lack of isomorphism between alignment units and translation units exactly because of dynamic processes such as this manipulation of segmentation.

Translation, as we have seen, like text writing, is processed on two axes: the paradigmatic and the syntagmatic. Because we have a ST to look at and dictionaries that establish relations between words in one or two languages, there is a strong tendency to look at it only from the paradigmatic point of view. It is editing most of all that may be processed essentially at the paradigmatic level: the translator looks for gaps and glitches in the sentence flow, and finds the unit from the same paradigm of the flawed unit to correct the issue. Still, not even that is a simple process, because the decision process behind it may be motivated by information from a different point in the text. The notion of micro and macro units of analysis suggested by Fábio Alves' team is a good instrument to capture the dynamics of these processes.

CAT tools are very powerful instruments, not only because they have been a fundamental tool in helping translators increase their productivity while improving consistency and abiding by other requirements in industrial contexts, but also because they permitted the production of vast amounts of data that has fed powerful MT systems. The time has come for the technologies in those MT systems to offer their functionalities as support for the work of translators. This can be achieved by helping with sub-segment alignment, but more can be achieved.

By studying SMT and ML, one is impressed by the plethora of approaches, methods and techniques to tackle small and big problems. Often, there is more than one way to do the same task, such as learning patterns, extracting best ranked examples, or estimating behaviours. So, there is a large margin for testing procedures, but also a large margin for making mistakes.

Research in MT should not be so focused on errors. The "error-free" translation is just a chimera. Current evaluation methods are comparative measures, not means to eliminate errors. To complement them, there is the need for more flexible process-aware criteria. The simple replacement of the word "error" by "edit" embodies this much more positive attitude.

It is my belief that translation is still, and will continue to be, an activity centred on people. So, even if the human element is just a piece in the loop of a mostly automated process, it is from him that all knowledge emanates, and he is the one that most benefits from it. His role is not secondary, even in PE processes: it is up to the translator or post-editor to validate, decide, prioritise and create hierarchies of pieces and parts of textual knowledge. Without this work, all we have is artificial language, as Sager said. And artificial language is not an adequate medium or code for human communication.

Since PE revolves around efficiency, one must have tools that respond to that requirement. Since it is an attention management task, with scattered actions, the tools must be flexible enough to respond dynamically to varying demands. Interactive systems are very hard to develop, but incremental models of learning seem to have the flexible technical solutions to tackle these issues.

5.2. Contributions to the definition of Post-editing

A theory of PE must be grounded on the context of the workflows in which this task is usually performed. This section presents and discusses the definition of PE, and then discusses in more detail its implications, the definition of an editing threshold, and evidence that PE is a specialised skill.

5.2.1. Defining PE

The following proposal for a definition of PE was presented previously in section 4.2. This definition will be discussed with some detail below.

> **Post-editing** is a generic name that describes a set of tasks by which a translator modifies language content that has previously been converted from a Source Language into a Target Language by a Machine Translation system, in order to make it conform to the objectives defined for the Target Text. The set of tasks required for the modification of the machine-translated content may include translating, editing and revising.
>
> **Post-editing** may be identified as being composed only of editing, but this is only possible if the purpose of the Target Text can be achieved by performing only the four technical actions (deleting, inserting, replacing and moving) over the machine-translated content, within a defined editing effort threshold.

Over the course of this dissertation, it became clear that a designation was necessary for the micro changes applied to TL sentences in TP and revision. I suggested that the term "editing" is the appropriate designation for these tasks. Furthermore, I suggested that editing may be decomposed into the four editing actions (deleting, inserting, replacing and moving).

Based on edit distance metrics, the four editing actions appear everywhere in MT literature as the fundamental pieces that describe the PE process. The discussions on why PE is not a synonym of editing and not a form of revision have been presented in section 4.2.1. Now, all that is left to do is to explain that it has several similarities with translation, although it should be considered a separate process. The main reason why it must be considered a separate process is that it is considered as such and performed as such every day in the localisation industry.

Let us look first at a table that uses the same model that was presented for translation and revision in sections 2.4 and 2.5. This is the simplified view of PE.

Phase	**Window**	**Textual element**	**Task**	**Subtasks**
Post-editing	Source	Source text	Reading	Reading
	Target	MT target text		
	Target	MT target text	Writing	Validating / Editing
Result of the process		**Translated text**		

Table 6 – The PE process (simplified).

The descriptions of PE that assume that the translation work has been done by the MT system exclude from PE the tasks of the orientation phase and the drafting. All the post-editor does is edit the MT hypotheses. Maybe this is too radical a representation, as in the literature there are references to an Orientation phase, associated with the reading of the ST and some (although limited) room for research. But the drafting phase seems to have been totally excluded from PE – the MT hypothesis replaces the need for a first draft. Besides, it occupies the place of the contents that the translator typed, overwrote or edited. So, apparently, PE corresponds only to the last phase of the TP, the revision. But if we compare this to the tables that describe revision, in section 2.5, this is only similar to the simplified view of revision, which means that even revision has a more complex description than PE.

However, in industrial translation environments, as we have seen in section 4.2.2, PE projects are executed in CAT tools. MT content is fed into these applications, but TM takes precedence in many cases. This is the PE process in more detail.

Phase	Window	Textual element	Task	Subtasks
Orientation	Source	Source text	Reading	Reading
	Target	TM target text (full and fuzzy matches)		
		MT target text (no match)		
	Termbase	Terms	Researching	Searching
	Concordance	Words in TM		
	Browser	Words		
Drafting	Source	Source text	Reading	
	Target	TM/MT target text	Reading	
		TM/MT target text	Writing	Validating / Editing
Result of the process		**Post-edited text**		
Self-revision*	*Target*	*Post-edited text*	*Checking*	*Validating / Editing*
Result of the process		**Translated text**		

Table 7 – The PE process (detailed).

In this description, the PE process includes all the phases of the TP. However, this description needs to be clarified.

As we have seen in the previous section, there are several papers applying TPR methods that did not identify the three TP phases in PE projects. However, these projects did not involve the same mixture of TM and MT segments that appear in real PE projects. And when they do, these are treated separately, in studies that compare the two modes. The requirement of the ecological validity of process research advises that real conditions should be reproduced in the lab. So, to reproduce real situations, PE projects analysed in the labs should always include a TM and fuzzy matches side-by-side with MT segments.

In the Orientation phase of the table above, TM target text is represented by full and fuzzy matches, but the process by which fuzzy matches are processed is not discussed: if they are composed of TM fuzzy match repair, or by the input of MT. It is also not clear yet if MT will take over any of the fuzzy bands, namely the lower ones. Furthermore, in the Drafting phase, the representation of the edited content has been simplified as “TM/MT target text”. This, of course, could be detailed further, as it includes TM full matches that the users usually only validate, full matches that are edited with the usual CAT methods, and MT hypotheses, that are edited as expected in PE processes.

Self-revision is highlighted in the table because it is not clear whether or not this phase is required by the industry. As previously commented on in section 4.2.1, the existence or not of a self-revision phase in PE may determine if the output can be considered a translated text or not. Without a self-revision stage, the result of a PE process can only be considered a post-edited text. And in order to become a revised text, it needs to be revised by a second translator, after the complete TP.

It results from this discussion that PE should be studied and approached in practice as a form of translation, rather than a form of revision. This is especially true if one takes translation as a broad term, one which encompasses multiple forms of transferring content from one language to another; a term so wide that describes what happens in such different activities as HT and MT.

5.2.2. Practical implications of this description of PE

Although apparently theoretical, the systematisation of the concepts involved in PE, as presented above, may have practical implications.

In section 4.2.1, I argued that the distinction between light and full PE had not proven its usefulness. I also argued in that section that there were process instructions missing from PE guidelines.

From the point of view of the provision of translation services, it is more important to discuss and negotiate whether a post-edited text requires a revision than to establish the number of errors a translation may contain. A discussion on the basis of requirements of levels of service might not only clarify quality standards and help in claim processes, but it could force a more realistic approach to the actual effort involved in transforming a MT text into a translated text.

Ana Guerberof Arena’s work on comparing effort in fuzzy matches and TM segments (see section 4.1.2) is aiming at one of the levels at which the editing effort concentrates. As she says, it would be very useful to have compared data from extended testing across all fuzzy levels, with different MT systems, different CAT tools, different language pairs, and different domains.

However, these analyses should also include the “no match” segments, which are sometimes excluded from these studies because it is assumed that for these there is no better alternative to MT. No match segments hide specific challenges even for theoretical approaches, which are not solved by simply using MT output to help translate them. It may be counter-intuitive, because one considers that any content that may be presented in the TT is helpful, but that is not true, even for translators who belong to the “satisficer” class. One needs to accept, for example, that the decision to delete a whole MT sentence and translate it from scratch may be the most efficient one, in view of sentences that require extensive editing. Translators, being paid at a lower rate for PE than for translation, and being used to making complex decisions fast, only take this decision when they realise that this is the best, in terms of quality, but also of productivity.

One final note on the practical implications of defining what PE is and what it is not. We saw in section 4.1.1 that ISO 17100:2015 considers that when a translator copies MT content into his CAT tool he is not doing PE. However, there is no discernible difference between this practice and a project in which the MT content came with the PE assignment, in terms of the process the translator executes, or in terms of the results he will probably obtain. So, I argue that this practice is included in the definition of PE. Furthermore, PE, as a process that encompasses different dimensions of translation, should be included in an industry standard aimed at defining the processes and procedures that are approved for the provision of quality services in the localisation industry.

5.2.3. The editing threshold

Several authors comment on identifying a threshold that enables the filtering out of MT segments that reduce editing productivity. But this threshold has other implications. If a sentence requires an editing effort above the set threshold, the industry should recognise this as a segment that was translated, more than just edited. This has clear consequences on cost definitions and service level negotiations. However, in order to maintain the sustainability of the service providers and the quality of the products, this discussion should happen.

This line, which defines the point above which the complexity of editing involved in a PE project exceeds the level that is considered reasonable for an editing task, shall be called the "editing threshold".

Setting the editing threshold is not an easy target to meet. One may speak of an effort threshold, but it is not yet clear how this effort should be measured. Editing rates measure a specific dimension of technical effort, but miss others. For example, a simple replacement of a word may have implied a research effort that is not reflected in the edit distance. Michael Denkowski and Alon Lavie discuss examples that show this disconnection between edit distances and effort (see section 3.5.3).

Temporal effort is an obvious addition to this. However, there needs to be some term of comparison, some standard, for it to be possible to use time effort as a useful method to define a threshold between translating and editing. Time is a very volatile variable, as it depends on many factors, such as individual variations, interface functionalities, and others. Besides, tasks that complement the TP, but which are not performed in CATs, such as terminological and web research should also be accounted for in this definition of the threshold, but are not usually included in the analyses.

Let us consider that the threshold is placed at the usual start of fuzzy match bands: 75% fuzzy match. This fuzzy score represents the percentage of words that are similar to the reference in the TM. It may also be described by the inverse proportion, i.e. the percentage of words that are different from the reference, and that need to be edited. So, this level may also be described as showing a 25% editing rate.

Fuzzy scores and editing rates are inverted measures, which means that when one increases, the other decreases. (This is intuitive: the higher the percentage of similar words in a sentence, the lower the percentage of different words in the same sentence.)

This inverse proportionality is visible in the table below, with fuzzy scores decreasing from top to bottom, from 100% full matches at the top to 0% matches at the bottom. Editing rates increase in the same direction, from 0% editing for full matches at the top and 100% editing rate for no match segments at the bottom. The 100% editing rate at the bottom means that the last band contains, besides different levels of segments with a few matched words, segments in which there are no words matched to the TM. Even with this ample spectrum of segments in this last band, in CAT projects with no

MT, these segments are accounted for as requiring full translation, editing being the task that is performed only on fuzzy matches.

In PE projects, illustrated below, all the “no match” segments are pre-translated by the MT. The name of the task performed on the segments that have MT content is no longer considered translating, but editing. This means that there is the assumption that the translator will be able to edit all the segments across this wide spectrum (from the point of view of the TM), and that he will not be required to do any translation work. So, the effort in these segments is seen as similar to that of fuzzy matches.

	Fuzzy score	**Editing rate**	**Resource**	**Task**
Full match	100%	0%	TM	Checking
Fuzzy match	99-95%	1-5%	TM	Editing
Fuzzy match	94-85%	6-15%	TM	Editing
Fuzzy match	84-75%	16-25%	TM	Editing
No match	74-0%	26-100%	MT	Editing

Table 8 – Fuzzy scores and editing rates from TM and MT.

However, there are no fuzzy scores to assess the similarity of these segments to the TM reference, and the only measure for them is the expected editing rate. Still, following the proportion to the fuzzy score, this could include everything from segments that imply editing 26% of the words to editing all of the words in a sentence.

Furthermore, as the rate of editing that these segments will imply is unknown, one must admit that the editing may range from no need for editing, when the TM hypothesis is perfect, to 100%, when none of the words in the hypothesis may be preserved by the post-editor. Although theoretical, these lower and upper bounds must be accepted as possible, even if the representatives of the MT industry would hardly admit the need to edit 100% of the words in the hypothesis.

If the proposed threshold is set at 25%, the previous separation line in CAT is re-established. And, as in pure CAT environments, any segments that require editing of more than 25% of its words should be considered as having been translated, not edited.

	Fuzzy score	**Editing rate**	**Resource**	**Task**
MT hypotheses		0-25%	MT	Editing
MT hypotheses		26-100%	MT	Translating

Table 9 – The MT editing threshold.

The upper 0% and the lower 100% bounds are kept, but now there is a line for quantitative analyses to consider. Different effort rates may be analysed for segments above and below the threshold, the percentage of segments closer to the upper or to the lower bound may be identified, and many other studies may be conducted.

Another argument for the usefulness of the editing threshold is the fact that previous studies have shown that most editing effort is surprisingly concentrated in the medium (quality) ranges, and not closer to the upper or the lower bounds (see Turchi et al in section 3.5.4 and Krings in section 4.1.2). So, research attention could concentrate on these intermediate levels. It is in the ranges where most complicated cases are concentrated that the definition of separating lines is most helpful as a means to explain variation.

The same scale should be applied to the text level and to the project level. Even with the support of MT content, a project that requires that more than 25% of its words should be edited should be considered as a translation project, not as a PE project.

5.2.4. PE is a specialised skill

Simply dismissing the tendency for translating from scratch has a negative impact not only on training but also on the conceptualisation of the processes. Training post-editors by giving them the means to develop more efficient ways of editing is a positive advancement, but this needs to be presented in a different framework. PE is a specialised task, which requires professional translators, used to making complex decisions efficiently.

The reasonableness of retranslating a segment was discussed above, but one may mention other major changes that translators may make to MT segments. One of them is manipulation of segmentation. This means, for example, to disassemble the structure of a noun phrase in a way that disrupts the previous alignment of segmented units (e.g. instead of "<u>A tall man and a short man</u> walk into a bar." use "<u>Two men, a tall one and a short one</u>, walk into a bar."). This type of transformation may be required by the language, the sentence structure, or because of a specific formulation in a termbase that must be followed, and not because of style. In such a case, a translator may decide not to delete the MT hypothesis but to do simple editing operations, by deleting and moving words around. So, although he could not be accused of overdoing it, because he did not delete and retranslate the sentence, he would have left only four words, out of eleven, unedited. This means that the sentence would have a TER score of 0.64, or 64%, well above the editing threshold.

Finally, another reason why translators may go beyond typical and simple editing work is when they find a particular sentence with a mistake in the source, and he needs to disrespect the contents of the MT hypothesis to correct it.

This shows that there are many translation shifts that translators employ even in PE projects. There is a wide range of such examples, which emphasise that PE is a process from which translation effort and translation skills have not been removed, and that extra adaptation techniques may even be necessary.

Besides, PE is a recursive and cumulative process. Translators make several edits one on top of another, and they reopen previously edited sentences, in processes that require not only experienced and focused readers, but also efficient technical performers.

5.3. Description and support to the four editing actions

Translation tools that support PE projects should be based on a good description of the four editing actions. The following sections detail the evolution from the description of the editing actions to a proposal on how to support each action in practice.

5.3.1. Types of editing actions

This section shows how the four editing actions may be classified using different criteria, each revealing a distinct perspective over these actions. These various perspectives may give shape to different ways of studying them.

The technical analysis of the four editing actions is based on two dimensions:

- the content of the textual units the editing actions are applied to;
- the positions these units occupy in a segment.

The separate analysis of these two dimensions is important for the orientation of their automated processing. These dimensions are at the basis of the mathematical representation of data as vectors, which, as we have seen, enable many of the advances in current MT technology. Textual units include words and phrases, and their content only considers their form, or lexical features. We are always looking at an analysis of MT hypotheses vs. edited segments.

Primary and secondary editing actions

This allows us to distinguish between primary and secondary editing actions.

- **Deletion** and **Insertion** are primary editing actions. This means that the position of the edited unit is occupied in only one of the elements:
    - Deleted units only appear in the MT hypothesis;
    - Inserted units only appear in the edited segment.

For primary actions, there are no changes in content to take into account. These are pure cases of misalignment, which makes them easy to identify in an alignment check. Studies that are focused on adequacy alone assume that these actions are sufficient to model the whole editing process.

- **Replacement** and **Movement** are secondary editing actions. This means that they appear in both the MT hypothesis and the edited segment, and they represent manipulations of units with either different content in the same position, or the same content in different positions.
    - Replacement implies the substitution of a unit from the MT hypothesis by a unit with a different content, without changing its position;
    - Movement implies the change of position of a unit without any change to its content.

Both replacement and movement have been excluded from several studies on editing, because they are associated, not with adequacy, but with fluency (often seen as less important), and they can be decomposed into a sequence of a Delete action followed by an Insert action:

- **Replacement** is the deletion of a unit and the insertion of a different unit in the same position;
- **Movement** is the deletion of a unit in one position and its insertion in a different position.

Although secondary editing actions may be described by sequences of primary actions, there are advantages in considering them as specific actions, as will be demonstrated in the next section.

These actions are also more associated with fluency, and considered as less important than those related to adequacy.

Clustering editing actions according to manipulation of position or content

One of the advantages of independent analyses of positions and forms is that this enables different clusters of the actions:

- **Deletion, Insertion and Movement** imply manipulation of word positions;
  - Replacement implies no changes in position.
- **Insertion and Replacement** imply the (re)creation of a unit's content;
  - Deletion and Movement imply no changes to content.

Editing actions according to the paradigmatic and syntagmatic axes

A final classification that may be applied to these editing actions is based on the preponderance of the paradigmatic or the syntagmatic axes, mentioned before (see sections 2.3.5 and 3.5.1).

- **Deletion, Insertion and Replacement** may be estimated from alignments, in the paradigmatic axis;
- **Movement** operates on the syntagmatic axis, and it must be estimated in the syntagmatic axis.

This description of editing actions has different implications for the development of interactive interface elements.

5.3.2. Modelling editing actions

In the model that is presented in this dissertation, the editing actions are learnt and estimated as specific features of text units. Modelling these actions (i.e. representing them in a way that enables their automated processing) is not a trivial process. In fact, hard ML methods are required to adequately tackle these problems.

Let us consider that, for this analysis, we have tuples composed of three elements:

- a SL segment;
- an MT hypothesis;
- and the edited segment.

We have been looking only at the two TL elements (the MT hypothesis and the Result of the editing process).

Secondary editing actions and alignment

Approaching secondary editing actions by decomposing them as sequences of deletions and insertions is a strategy that loses important information.

Real deletions and insertions are essentially misalignments, since both inserted and deleted words align with empty positions in the other TT segment (MT hypothesis or edited segment). In fact, QE specialists have reported special difficulties in modelling the deleting action (Specia & Scarton, 2016, p. 24). However, replacements and movements are applied to properly aligned units, so considering that they include deletions complicates global alignment strategies.

Errors of alignment and missing data (like out-of-vocabulary words and sparse data) are problems that seriously affect the success of many projects (see, for example, Wisniewski et al. in section 3.5.3). The analysis and collection of all editing actions with the form and position parameters may provide invaluable information for such systems. For example, the movement of a unit is a strong indicator of the cohesion of the unit. If there was no alignment for that unit (let us imagine that this is a pentagram, and the model only trained a trigram TrM), this information should be added as a request for a future retraining or tuning of the background MT system.

A second advantage of secondary editing actions is that they enable a more precise analysis of the balance between complexity and efficiency in the use of editing actions, or of editing processes in general. Replacement and movement may be more complex to model and more cognitively demanding (as was suggested in previous sections), but their use may be associated with more efficient methods.

Actions with content creation and without content creation

Insertion and replacement are the only two actions that recreate the content of selected units. This means that they imply typing. So, these are the only two actions that may benefit from predictive writing suggestions.

On the contrary, deletion and movement are editing actions in which there is no creation of content. In these cases, after a unit is selected, it is either totally removed or moved to a different position.

As a consequence, once the user selects a unit to delete or move, there is not a lot that the system can do to help, except maybe estimate the new position for the moved unit. So, the usefulness of these two editing actions is limited to its predictive value. If a system indicates probable words to delete or to move, it will be useful. Otherwise, their value lies more on the learning side of the system than on the side of supporting users.

Actions that involve new positions

To properly estimate actions that imply changes in units' positions, it is necessary for models to consider spaces between words in the MT hypothesis as possible destinations for those actions. Each space may be the position where new units will be inserted or the destination for moved units. Besides, the system needs to consider that all words are possibly misplaced. SMT reordering models and their context features (see below in this section) are good supports to handle these two problems.

Disjoint units

Another factor of complexity is presented by disjoint units, i.e. linguistically cohesive units which occupy separate (long distance) positions. An example are units that are commanded by gender or number agreement but which have elements that occupy initial positions in a sentence, such as the subject of the sentence, for instance, and other elements that are placed at the end of the sentence, such as verb participles or inflected sentence predicates.

In an ideal editing environment, a replacement of the first part of the unit by a form that changes its inflection, for example, from singular into plural, should be propagated into the rest of the elements of the unit, no matter where they are placed. For the MT system to bring up the plural alternative for all the words in the selected unit, it would have to include non-contiguous phrases in its phrase table. To learn such phrases, the cohesion of the unit might be identified at the moment the user selects all elements simultaneously (skipping words in between them).

The challenge of aligning textual units and actions

The alignments to estimate editing actions are made on the TL side, between the MT hypotheses and the edited sentence. So, we are dealing with product data.

The textual units that are associated with any of the editing actions may, or may not, be aligned with a parallel unit in the SL segment. Usually, if the MT hypothesis is not aligned with the ST, these are considered "errors" caused by the MT system. Consequently, editing work is associated with the task of correcting these errors. This is not the view one needs to take to study the editing process.

Process data should retain, during the process, the actual editing action and the textual unit it is applied to. The corrected textual unit (in terms of content and position) should be associated with the unit it aligns with in the ST. This means that the collection of process data should ideally include a non-intrusive process for the validation or correction of the alignment of the resulting textual units with the ST units:

- In the process of deleting a text unit, the learning system registers that this unit was not aligned to any unit of the ST;
- When the translator inserts another unit, the system is informed of a missed unit in the TT for a unit in the ST;
- Replacement is the correction of the content of a correctly aligned unit;
- And movement is a validation of the alignment of the moved unit, since only its position is changed.

The recursive and cumulative nature of editing work makes this a very complex matter. Moreover, in order for this validation or correction of alignment with ST may impose a new cognitive load on the translator.

The order of the editing actions

There seems to be a natural order in the sequence of application of these actions: from primary to secondary, from those that imply empty positions to those that imply changes in content. A tentative ordering of the actions following this rationale, in a sort of procedural guideline, might be:

- First, delete superfluous words;
- then, insert missing words;
- then, move existing words to the correct positions;

- and, if any of the resulting units needs to suffer a change in form or content, replace it at the end of the editing sequence.

The advantage of this sequence is that we start with the simplest actions, by "cleaning" the hypothesis, and we leave for the end the one that is most sensitive to its contextual information for support.

Replacement is the action most often performed by translators. It is also one of the most complex, as it implies inserting new text, taking into account the context. It may include replacement of one or two characters in one word to make it conform to the agreement rules of the sentence, but it may also include substitution by synonyms or paraphrases. These edits have been tested in different systems, with the support of rules, dictionaries, and other linguistic materials. But this can also be approached by a search in the phrase table, by looking for the alternatives for the selected phrases in the phrase table. Chris Hokamp suggests this approach in (Hokamp, 2015a), for the "phrase table backed autocompleter".

Whatever the method used, if the above order is followed, the prediction method for what words a user may want to type in specific contexts, would be called in when the context it fits in is already stable.

Solutions for modelling editing actions

There are several attempts in the SMT literature to model and to process non-contiguous phrases, so this is not a ground-breaking challenge.

There are other previous experiments on modelling editing behaviours as part of SMT research. One such case is reordering models. These models attempt to anticipate the movement of words and phrases in sentences and they have been shown to improve the positioning of verbs in German sentences, for example.

One of the challenges for reordering models is the notion of "context features". See, for example, (Ni et al., 2010). These context features are words that share the same sentence. In a model that learns editing actions, context features may help estimate any actions that depend on changes in position, like movement, but also content. Defining the extension and scope of the search window for these context features is a complex challenge on itself.

As this dissertation demonstrates, there has been a growing interest in approaching editing actions as information that is useful and should be integrated into translation learning models. Some of these studies show very strong capacities to deal with these problems, in at least some of their dimensions. The work done with CATaLog by Santanu Pal and colleagues seems to be an important step in tackling the modelling issues that are presented in this section.

5.3.3. Proposal for learning each editing action

We have seen in section 3.5 how feature selection plays a fundamental role in QE. We have also seen the similarity of this application to estimation of editing. This section presents a speculative list of features inspired by QE methods that may be interesting to test as inputs for this estimation.

The purpose of this list, as a speculative exercise, is to open the door for proposals that are inspired by a focus on the transformations that are performed during translation and editing processes. It was not tested or confronted in any way with real situations or with learning systems. So, it assumes the risk of being easily dismissed by a proper test.

The following list presents only one set of features per editing action. Besides, only shallow features are presented, which require no enhancement of existing data with extra annotation, whether it is linguistic or related to the editing actions. So, these features are seen as having the advantage of being easy to train and test with existing bilingual data, in multi-task learning models, and they could, eventually, be used in online learning (OL) models too.

The only extra material that may be required is, beside the TrM and the LgM of a usual SMT system, a "TM model". This may be built from an n-gram analysis of the content of a TM. This model is a reinforcement of the data in the target LgM, and a priority validator of the units that are evaluated.

Deletion: to identify candidates for deletion, a training method might look for the following probabilistic pattern:

- Aligned units (words or phrases) with high probabilities in the MT phrase table
- BUT with low probabilities (or that do not even appear) in the target LgM and/or in the TM model.

Insertion: probable candidates for insertion could be:

- Units with high probabilities in the target LgM and/or in the TM model
- BUT which show a low probability or do not appear in the phrase table.

Movement: the existing reordering models are probably good sources of features for this editing action, but candidates for movement may be:

- Units that have high probabilities in the phrase table
- BUT, although they appear with a high frequency in the target LgM, they occupy fixed positions (relative to specific contexts) in the TM model.

Replacement: there are many different types of candidates for replacement units, but a clue may be:

- Alternative translations in the phrase table for the same TL unit, with short distances between their probabilities,
- AND with high probabilities and close probabilities both in the target LgM and in the TM model.

5.3.4. Support for each editing action

The editing model that learns the four editing actions presented above may be used to estimate editing actions. The system may then resort to these estimates to present editing suggestions to the user, especially in repeated contexts.

The following examples were collected in real professional situations, as examples of typical occurrences in a PE project. They were originally presented in (Carmo & Maia, 2016), in the ambit of this PhD project. The interface element (a context menu) was presented as a mere suggestion. The colour codes for the editing actions used in the tables below are used throughout the dissertation.

Estimating deletion

Repeated constructions may imply repeated corrections. They may imply, for example, the deletion of the same words in the same context.

In the example illustrated below, the project involved a glossary of terms. In English, the infinitive of a verb always includes the preposition “to”, but in Portuguese there is no corresponding word in the infinitive. The MT system always inserted this preposition, and the post-editor had to delete it repeatedly. Since the context is clearly repeated, and so is the editing action, the system should have a method of learning not just the new alignments, but the delete actions too.

After identifying the recurrence of the mistake and of the action, the system might highlight the unit with a discrete coloured indication of the suggested action in the MT hypothesis. The user might confirm the edit quickly. If the error was repeated enough times, the system might prompt the user to confirm the deletion of all repeated words in the same context in the text in one single action. As an initial cleaning action, even if this would have to be reverted afterwards in some specific context, this action would simplify the initial reading stage of each segment.

SOURCE	**MT HYPOTHESIS**	**POST-EDITED**
Align - to place something in an orderly position in relation to something else	Alinhar - **para** colocar algo em uma posição ordenada em relação a outra coisa	Alinhar - colocar algo em uma posição ordenada em relação a outra coisa
Allocate - to divide something between different people or projects	Alocar - **para** dividir algo entre diferentes pessoas ou projetos	Alocar - dividir algo entre diferentes pessoas ou projetos
Acquire - to obtain possession of something	Adquirir - **para** obter a posse de algo	Adquirir - obter a posse de algo

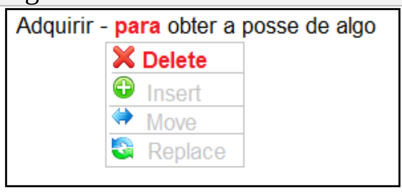


Table 10 – An example of the application of a suggestion for deletion.

Estimating insertion

Insertion behaviour is similar to deletion, with the exception that the system can only learn that there is an empty position, and which unit should fill it in, after the user makes the edits. This raises a usability challenge: after how many repetitions of the same action should the system make a suggestion? The second time the same edit is necessary, or only after a certain number of times? A model that describes the repetitions in the ST (a ST model, as described below) may help determine the answer to these questions, in view of the repetition of the "context features".

In the examples below, the same structure "Name/ID" was used in the ST, but the MT system did not insert the "/ID" part into the TT, because it included many more occurrences of "name" alone. After the set number of repeated corrections, the system may signal the position in which the word may have to be inserted, and when the user hovers over it, it presents the suggestion. A simple click may confirm the insertion.

SOURCE	MT HYPOTHESIS	POST-EDITED
Patient Name/ID	Nome do paciente	Nome/**ID** do paciente
Item Name/ID	Nome do item	Nome/**ID** do item
User Name/ID	Nome de utilizador	Nome/**ID** de utilizador

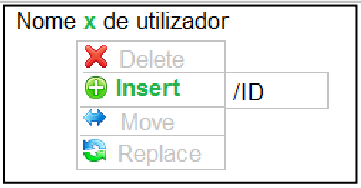


Table 11 – An example of the application of a suggestion for insertion.

Estimating movement

In section 3.4.2 above, there is a reference to a study that confirms that SMT presents very consistent results. It is not certain that this observation may be generalised for all SMT systems. As mentioned before, a specific set of challenges is presented to SMT systems by projects in which each segment is composed of isolated phrases in short segments, since each of these is processed at a separate moment in time by the SMT system. One of the challenges is the higher weight that each word has in each segment.

In the examples below, the three source segments are composed of 8 and 9 tokens (each punctuation signal being one token). Out of these, only three tokens (i.e. 30%) are common to the three. This means that, for the SMT, each segment is very different from the previous one. So, the output is very different too, and the MT hypotheses present three different translations for the repeated words.

Although these may seem like very rare translation situations, localisation projects are very frequently composed of lists of variants of similar constructions. In these cases, there is a very high disproportion between the number of words and the required editing effort. So, a mechanical approach that could learn both form and position in one single action would be an invaluable contribution towards reducing this effort.

An ST model could learn that there was a very frequent trigram in this set of segments, composed by the three first words in these three segments.

When the translator corrects the first segment, he edits not only the translation, but also the position of this 3-word phrase in the segment. Although the context (the other words in the other segments) is different, the position of the moved trigram should be the same in the three segments. This information, of the content and position of the moved unit, should be enough for the system to suggest the same editing action for the other segments in which the trigram appears. After the suggestion appears underlined, a click in the new position indicator is all it takes for the phrase to be moved there.

SOURCE	**MT HYPOTHESIS**	**POST-EDITED**
VEC 1 controller + (RD) wire	**1 Controlador VEC** + (RD)	Fio + (RD) do **Controlador VEC 1**
VEC 1 controller – (BL) wire	**VEC 1 controlador** - (BL)	Fio - (BL) do **Controlador VEC 1**
VEC 1 controller pin 7 (BK) wire	**Controlador VEC 1** fio do pino 7 (BK)	Fio do pino 7 (BK) do **Controlador VEC 1**

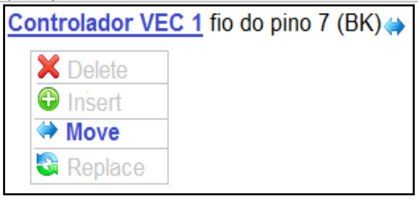


Table 12 – An example of the application of a suggestion for movement.

Estimating replacement

As described above, replacement includes character and word level changes. In the following examples, the polyssemic nature of English word forms resulted in the use of noun forms in Portuguese when verbs should have been used (in the first two cases) and in a wrongly inflected adjectival phrase (in the third example). These replacements are not indications that the MT system made a mistake that needs to be replaced in all similar occurrences of these words. These errors are merely contextual to this sentence. So, these corrections should be saved as alternatives, with a reasonable amount of context features attached to it. In order to call up alternatives, the user may select a phrase and scroll down, for the list of alternatives to appear.

SOURCE	**MT HYPOTHESIS**	**POST-EDITED**
Assess - to examine something in order to judge or evaluate it	Avaliar - examinar algo para **juiz** ou avaliar	Avaliar - examinar algo para **ajuizar** ou avaliar
Act - to do something to change a situation	**Ato** - fazer algo para mudar uma situação	**Atuar** - fazer algo para mudar uma situação
Users must be set up and maintained at the console.	Os utilizadores têm de estar **configurado e mantido** na consola.	Os utilizadores têm de estar **configurados e mantidos** na consola.

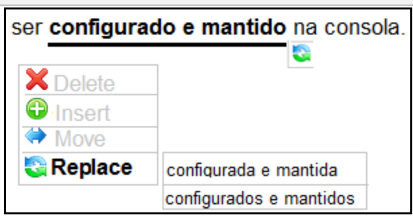


Table 13 – An example of the application of suggestions for replacement.

5.4. Contributions to the design of interactive translation systems

This section presents a proposal for the design of an interactive system that integrates MT technology into a translation editor. The MT component is seen not as a mere content provider, but as an integral contributor to the interactive process. The system is commanded by the needs of the user, but it is fed by a backend system that learns and presents contextual suggestions. These suggestions go beyond the predictive writing paradigm, by being modelled on the four editing actions.

The application of the methods presented in this section is based on the technologies described in the previous sections, and on the views of the TP and PE that have been presented. However, no claim is made on the applicability of these proposals to practice. The exercise is speculative, but it is presented as a contribution to the development of technologies that are fundamental to responding to the current challenges facing the translation industry and translation service providers.

5.4.1. Interactive tools development

There are several guidelines for the development of software adapted to users. Following is a list of a few that may help adjust such a project to its aims, mostly coming from the field of HCI – Human-Computer Interaction.

- **Augmented Intelligence** (Schmitt, 1998): this concerns the connection between AI (Artificial Intelligence) and human intelligence, which was proposed in the ambit of virtual and physical architecture design;
- **Mixed Initiative User Interfaces** (Horvitz, 1999): the focus here is on interfaces that, rather than being inspired by conceptions of automation of tasks, are guided by models of how users interact with technology, when they are solving specific problems;
- **Human Information Interaction** (Fidel, 2012): this area concerns the way users interact with information and it includes sub-areas such as “human information behaviour” and “cognitive work analysis”;
- **Component-Centric Design Framework** (Hokamp, 2015a): Chris Hokamp used this in the design of HandyCAT, an open-source translation tool adjusted to several research purposes. The idea behind this design framework is that data modification is at the centre of a

process like translation, and that interaction is based on display and inputs with that data.

Some of these areas are linked with Psychology. An area that needs to be taken into account when one analyses interaction with work instruments is the relation between conscious and unconscious processes. The study of learning curves in the use of new and complex tools may benefit from studies such as (Anzulewicz et al., 2015). Cognitive ergonomics has also been applied by researchers like Hanna Risku, in approaches like "situated learning", in which a set of environmental factors such as physical setting, task demands, and tools usability is studied (Risku, 2016), or to identify different patterns of text production in writing and translation (Risku, Milosevic, & Pein-Weber, 2016). Maureen Ehrensberger-Dow has also studied this domain, and produced detailed studies on cognitive strategies to deal with the challenges of technological interfaces of CAT tools (Ehrensberger-Dow & Heeb, 2016).

5.4.2. A structured architecture

We have seen above that most of the automated processing of bilingual data is based on sub-segment segmentation and alignment. These sub-segment alignments should be treated as a knowledge base, which contains the specialised linguistic and technical information learnt and created by the translator during translation, revision and PE projects. We have also commented on the need for a stratification of these knowledge bases, so that at each level only the knowledge that belongs to that level is managed (a personal level, a project level, a client level, and so on).

The construction of knowledge bases is part of the processes that run in the background of a translation tool, like the TM and the MT system. These backend resources may function like a black-box, as long as the user has an interface that gives him enough control over what happens at this level.

Another piece of the architecture is the editor. One of the main modules of the editor is the OL model, which dynamically learns from the user's actions and feeds suggestions, according to the contexts and the learnt models of the actions.

In this architecture, the web search is integrated into the translation environment, so that knowledge from both tasks may be used to improve the work of the other.

Finally, there are processes that are not linked directly to the production processes, but which may be fundamental for the best reuse of all the knowledge that is managed by the global system.

The following section describes a global system that manages several processes, divided by these four components: a backend learning system (section 5.4.3), an editor with OL (section 5.4.4), the web search manager (section 5.4.5) and the acquired knowledge management system (section 5.4.6). Finally, implementation and interface issues are discussed in section 5.4.7.

5.4.3. Backend learning system

The backend system is composed of different components, such as different learning models from different sets of data and mechanisms to select features to train editing estimators. The system that manages these components is described at the end of this section.

Main learning models

The typical MT models (target and source LgMs, and TrM) are included in this system. The phrase table that results from the MT training is also an important piece of the system.

Besides these, a TM model is also necessary. This results from the training techniques of a TrM applied to the contents of the TM. The TM model takes precedence over the TrM of the MT system. Together, the MT models and the TM model are called the “data resources”.

Finally, the system also needs an ST model. This is learnt as a LgM before the translation of a ST begins, and it serves two main purposes:

- as a mechanism to reduce the search in the MT phrase table;
- as the main guide for the interaction with the user.

So, at the beginning of the translation, we need to have the following models that have been trained offline:

- source LgM;
- target LgM;
- TrM;
- MT phrase table;

- TM model;
- ST model.

The content of all these models should include not just aligned units, but also patterns of editing actions, attached to these aligned units. The aligned units may be words, phrases, segments, or even texts, if one wants to account for discourse and cohesion mechanisms. Each unit should include context features, which are necessary to constrain the search during the interactive task.

The set of all the learning models is called the knowledge base. This is divided into a local model, which is active at each work session, and a global model, which saves and manages the knowledge from all sessions, texts and projects.

Expected advantages of the ST model

Some of the advantages that are to be expected from learning a ST model are explained below:

- The ST model serves as the support for the orientation or gisting phase of the TP: it results from a process that it is similar to a scan reading before translation. Eventually, it could be provided to the user to get the gist of the domain of the text, to prepare a list of terminology, or for a similar orientation activity.
- The ST model constrains the search through the search space of the phrase table, in a similar role to that of a Project TM. This reduces the memory requests to the MT system during translation.
- The presentation of suggestions to the user in a dynamic way may be guided by the ST model: the more often a phrase is repeated, the more propagation of repeated edits the system is capable of anticipating and presenting to the user. This may allow for the elimination of the frustration of "having to correct the same mistake over and over again".
- The precision of systems that learn editing actions may be improved by the combination of these factors. The disappointing results reported by the CasMaCat experiments in 2015 (see section 4.3.5) may probably be improved with the input of the ST model.

Selecting features to estimate editing actions

To provide useful feedback, the system needs a mechanism to learn the editing actions. QE methods may help at this task, since their purpose is to identify the factors that make translation difficult, so as to estimate the required editing of a text.

The estimation process used in QE is based on the selection of the features that show more accurately which editing is necessary. These features are organised into four categories: complexity of the source, adequacy of the translation (relation between ST and TT), confidence of the MT system, and fluency of the output.

A similar organisation may be applied to the component that estimates the editing actions in this system. The features on which the frequency and type of the editing actions that are to be expected depend include factors such as the degree of internal repetitiveness of the ST, but also the previously evaluated efficiency of the MT system. The following list systematises these features.

- **Complexity of the ST**: this may be measured in several ways, one of these being the internal repetitiveness of n-grams. This measure is captured when the ST model is trained.
- **Adequacy of the data resources**: this is a measure of overlap between the ST model and the data resources (the TM and MT models). This may be a measure of how much may be leveraged from them, but also of how much "in-domain" these resources are.
- **Confidence of the MT system**: this concerns the number of alternatives the system has for each phrase. This measure may help determine how informative the suggestions from a phrase table will be.
- **Fluency of the MT output**: this is a historic score from previous projects, learnt from the amounts of edits the user usually needs to make to the MT output.

A combination of features from these components may be trained and tested until an optimal combination for the specific user is found. Afterwards, this global editing estimation mechanism may be applied to estimate the editing effort that each text will imply.

The information from this evaluation may, furthermore, help score MT systems according to their relative capacity to produce a higher or a lower ratio of hypotheses that are ranked above or below the editing threshold.

Backend knowledge management system

At the backend, there are global trained models, built from the local models that are used for each work session. These models may be discriminated by text, project, domain, client, or any other level that the user believes to be useful.

Before any actual production work, the preparation stage includes the extraction of the ST model, and the selection from the knowledge base of the units and editing patterns that are likely to be used in that text. This is a stage similar to the word count analysis that should be done at the beginning of each CAT project, and ideally it should not take too long and end up wasting valuable production time.

In a scenario where the proposed system that learns editing features is active, the system may present not just a word count for the project, but also a prediction of the effort it will take. It may also show the number of segments that are above and below the editing threshold.

All learning in the backend is done offline. After the learning is done, following the ST model, the text is pre-translated, by applying the best combination of TM matches and MT suggestions that improve the editing effort expectation. The most relevant TM matches and sub-segment phrases from the MT phrase table are kept as the local model for each session, text or project, as deemed appropriate. This local model is the extract of the knowledge base that serves as a reduced search space for the OL module used during the translation stage.

5.4.4. Editor with online learning

The translation editor, no matter what task the translator is doing, translation, revision, or PE, has an active OL module. This module is fed with the local model that was extracted at the backend. Searches in the local model are constrained by the ST model that describes the complexity and regularity of phrases the translator sees during the task, and are triggered by the content of the segment in which the user is working.

As the translator proceeds with his work, the OL mechanism updates the local model for that session, text, or project. When the user validates segments and sub-segment units, the OL module increases their scores in the local model. As for the edited segments and sub-segments, the OL module looks for the new alignments, or non-alignments, and saves them to the local model too.

When the translator works on a segment that has a repeated unit, the system presents editing suggestions, which depend on the token that the user selects, or the editing action: if there is a learnt editing action that estimates that there is a high probability that a certain word is to be deleted, the OL module may suggest it beforehand, or do it only when the user clicks on it. This module may use predictive writing functionalities, but only for the insertion and replacement actions. As frequently repeated actions are performed and validated, the OL module performs those actions in the following segments. If the user wants a long-distance unit to be edited at the same time, he selects it by skipping in between words, and most likely replaces it. If the system has that unit in the local model, it may present alternatives, otherwise, it adds it to the model.

OL modules are capable of constantly updating the information used in each context. So, it may adjust the suggestions it presents according to what the translator does in each segment. If the translator starts by deleting a word, the words that constrain the search are different, and so the results of the search may be different. However, due to the recursive nature of editing work, this constantly updated search must be kept within reasonable limits.

Throughout the production stage, the user, but also the learning system, may signal segments and phrases that may need to be revisited, due to consistency or contextual reasons. Alternatively, the user may signal specific options that are intentionally different from what might be expected from an adequacy or quality expectation, like a shift from rigorous adequacy or alignment with ST content, fluency issues, or other contextual reasons, and which the translator knows that should not be transported to the knowledge base.

At the end of the production cycle, the user checks the translation according to his annotations and following the instructions for the level of quality that was defined for the assignment. The OL module may show the user the result of the OL stage, in terms of translations of more frequently repeated phrases, consistency issues, apparent misalignments (possible omissions, for example), apparent disrespect for the style guide of the project or any other issues.

5.4.5. Web search support

The article published in 2016 about the use of OL methods in CasMaCat (Ortiz-Martínez et al., 2016) reports an increase in web searches by users that receive suggestions by OL systems. On the other hand, translators are frequently faced with the frustration of time wasted online with unhelpful sites, sources which they were sure they had searched before, but now seem to present different information, and the lack of time or resources to organise and improve the efficacy of time spent searching the web. There is at least an extensive work on the relation between translation and web searches (Enríquez Raído, 2014), which shows that this is a theme that must be integrated into process research.

CAT tools currently include, either as embedded features, or as add-ins, capacities to select a few sites for one-click web searches. These are, however, no more than shortcuts to lists of favourite sites, as one may have in one’s browser. Web search engines, like Google, on the other hand, have customisation features that filter and sort searches according to user-adjustable criteria, like dates, countries and so on. Beyond this, advanced search features allow the use of search operators that improve the discriminative power of web searches. However, none of these functionalities are within easy reach, easy to setup, or adapted to the different uses that are required by translators.

Furthermore, there is no way to link the relevance of these searches for specific technical domains, legacy data for tracking previous searches on a specific project, or to easily share and improve on this knowledge in collaborative environments.

The management of data related to research stages is an invaluable asset for translators and team managers. If a translator can have the guarantee that he will easily access the most useful sites he found during a project in a new technical domain, he will probably accumulate less frustration when he has to change subjects frequently. And if a reviser can check and validate the list of references used by a translator before he starts his revision, this will probably increase his degree of confidence in the work ahead.

The proposal for the integration of web search knowledge into the translation tool is based on the idea of vertical search engines, which are specialised in specific domains. A knowledge management system, such as the one described in this chapter, should include a feature that related search history to the knowledge base of a project or a domain. Specialised searches might allow for bilingual search of a term, search for a term and its synonyms, target specific dictionaries or online references, etc.

The knowledge obtained from searches that are implemented into a translation project is useful information for time and production management too. For example, time spent on searching specific terms, in specific projects and domains, may enable the adjustment of not only the productivity expectations of a project in the same area, but also the editing threshold of similar projects.

As Pym says, nowadays it is more important to know where to find knowledge than to retain it (see section 4.2.2). If one can retain the information on where to find knowledge, maybe the best of both worlds can be secured.

5.4.6. Acquired knowledge management

After the project is delivered, the local model may be automatically incorporated into the individual knowledge base. The user may participate at this stage and check the impact that the new knowledge will have on the previously saved data. For instance, he may want to know the effect of the new data over the adequacy of the data resources (TM and MT); data on the confidence of the MT system may also present interesting data for future reuse. The user may, for example, ask for a new training of the TrM, or see a representation of the distribution of content per domain, if the knowledge base is configured by domain.

The management of the knowledge in the knowledge base may be controlled by a priority scale, adjusted by the user. Terminology may be managed outside the rigidity of current termbases, if they are assured a top priority and override status above all other units. Client-provided data may have a similar status, but the system may be sensitive to project or domain classification: a project-bound choice loses priority outside that project. Other criteria may be considered, as user-validated content takes precedence over MT content, or longer phrases take precedence over the smaller phrases they contain.

Besides, lack of alignment between small units is a common situation in very segmented and length-controlled projects, such as in software localisation and multimedia projects. When a system cannot find in a project enough aligned sub-segment units (according to the phrases in the knowledge base), it may preserve longer aligned units, from segments to sentences, or even to texts, if the system is adjusted to consider cohesive discourse features.

The information collected by editing actions may serve as a ranking system of these systems, prioritising the units that are most useful in the editing stage. This information becomes an integral part of the user’s knowledge base.

So, the system holds several levels of complexity that one cannot dismiss.

5.4.7. Implementation and interface issues

The tool that is described in this section is not just focused on translation or on revision. Yet, by addressing the needs of the orientation, the drafting and the revision phases, it serves all purposes. In this section, a final view on specific uses and implementation solutions is discussed.

Clean interfaces and workflows are a requirement for first-time use of any computer tool. This poses a design challenge for developers that are aware of the potential that their tool has: either they hide the most complex features and people end up not being aware of their existence, or the most advanced features are exposed in clickable elements, and applications end up with cluttered views and multiple panes with information. One of the solutions for this challenge is to allow for a multi-layered presentation of information, most of which is available on demand.

Another issue that must be resolved is the identification of the correct balance for the number of times an edit has to be learnt before it is presented as a suggestion to the user. For some users, the first repetition should already be helped by a suggestion, whereas for texts with a high density of repetitions, with too many regular suggestions, the user may feel that the system is always interrupting. A system that allows for flexible setting is probably the best choice.

On-demand information

Information on confidence levels of MT hypotheses, provenance of pieces of composed translations, legacy information on edits made, linguistic information, and error classification scores, have all been discussed as useful information for translation tool users. Most of these elements have been presented not as features that translation tools require, but as data collection instruments for segment-level annotation tasks. The system proposed in this chapter approaches these annotations as being provided while the translator works, through the connection between the OL module and the local model. The idea is that data collection is made from the usual work processes, dispensing extra tasks for annotation or classification of data.

For example, the priority ranking of phrase alignments that are fed to the local model may include a feature by which specific linguistic units are selected, such as “noun + preposition + noun” NPs, which is a common structure in a terminology extraction process. If the models have tagging capacities to classify these units, the user can do his work in the usual way. At the end, he may be presented with the list of selected units, and validate or reject them in a quick action. A similar process may be set up for error classification, and other annotation tasks.

But the user may also be interested in accessing a deeper level of information, while he is working. Just like when he used a concordance search to look for TM assistance in segments that require translation, the translator may want to see alternatives from the phrase table, or confidence statistics for specific segments. So, a good mixture of layers of on-demand information may be necessary.

Data collection detail

The level of detail in data collection should also be discussed. Although translators type a character at a time, and the processing unit in CAT and MT is the sentence, the action units are words and phrases. So, edit scores should be learnt at this level.

Furthermore, edits are recursive and cumulative. The translator may delete a word and insert it again, he may move a phrase only to delete it afterwards. He may confirm a segment and come back to it later to make a different amendment. So, in order to avoid noise in the system, the system should have a mechanism to discard the information that does not survive until the final version.

This will be discussed with some detail in section 6.3.3, in the context of the implementation of TER metrics to learn editing actions.

Interface elements

Two of the greatest design challenges in such an interactive system are the mode of presenting suggestions and the interface element that enables interaction. In fact, one may say that the usefulness of such an interactive tool depends on the precision and recall of the suggestions that are presented for the editing actions (recall being more important for the text typing suggestions), and the usability of the system depends on the design of the interface.

As other authors have mentioned, a suggestion popping up for every word that is typed, or a suggestion that is constantly being updated, are not perfect solutions. The interaction based on correct modelling of the editing actions may serve as a constraint to the presentation of suggestions.

All learnt editing suggestions may be signalled in the segment, by some discrete element, which shows added information when the user hovers over it with the mouse. The completion of the editing action may then be swiftly confirmed. But the user may choose to ignore these indications and ask for assistance when he finds it appropriate. After he selects each unit, the editing interaction is different for each action:

- If the user selects a unit to replace or a position to insert a new unit, he will appreciate the support of predictive writing;
- But if he selects a word for deletion, there is no need for any interaction, as the action is quickly completed;
- And as for movement, the system may suggest a new position for the selected unit, which the user confirms by clicking on that position, or rejects, by clicking on a different position.

In the proposal described above, in section 5.3.4, the interaction is done through context menus. This was the mode adopted in the tests that will be described in the next chapter, but perhaps this needs to be improved. The approach adopted for Kanjingo for touch screens, and the LFPE keyboard shortcuts approach seem to be good alternatives to this method.

Different types of users

A system with a dynamic learning interface is based on the expectation of a specialised user. So, there would need to be extended tests with users to reach a good balance between the exploration of the full potential of the tool's features and the usability requirements. Such a complex translation tool may need to have different interfaces for first-time and for experienced users.

Learners may be guided to using the editing actions separately and in a sequence, as suggested above. Even if the sequence is later abandoned, the approach to editing work through these technical concepts may be beneficial, in terms of guiding attention to specific issues, instead of looking for the whole sentence. In this case, the interface may constrain the work of translators and be more intrusive. For example, deletion and movement may be guided even after the user has selected the words to apply the action to. This constrained use of the tool may be seen as a side-by-side learning experience: the user develops his patterns of work as the system learns them.

Experienced translators, on the other hand, may already be very efficient at using traditional text editing techniques. In this case, the interface may be less intrusive, and interaction may be adjusted to the user's preference, such as keyboard shortcuts or other techniques. Besides, the system should contemplate the more advanced features, like TrMs with non-contiguous phrases, or more information levels available on-demand.

In either case, one must also consider the eventual benefits that lie at the end of a steep learning curve. Even if, before or at the end of the curve, the process is abandoned, there are gains to be made from having brought into conscious processing actions that are otherwise unconsciously applied. This unconscious repetition of actions is a cause of the association of these tasks with mechanical processing. Turning these processes into conscious efforts may transmit the notion that there are ways to regain control of the processes.

Interfaces for revision of PE

Finally, revisers may have a specific interface for the specificities of their work. Section 2.5.3 explained that revisers of PE need to have access to the hidden text that the post-editor edited. Observing this relationship may be more important for the reviser than starting from the ST, as if the reviser was a translator. The correct management of yet another layer of content should also be considered when designing a translation tool that is adjusted to all the roles played by translators.

5.5. Managing editing knowledge

Knowledge bases contain much more valuable data than TMs. TMs that were not built with knowledge from the translator may be a useful reference source, but seldom can be applied to a new project blindly. The reason for this is that translation is very sensitive to variations in contexts, the individual context being one of the most important. TMs do not transport the decision process behind each choice, or the technical way language material was processed. So, the number of data inconsistency and similar problems increases proportionally to the TM size. The simple temptation to add more data to get more leverage very often leads to an increase in these problems, with the consequent increase in the costs of maintaining the data and revising it before use.

Knowledge bases collect and manage process data, and as such, they contain data of a more individual nature. This raises issues not just in terms of privileged access to information, but also because of its value outside the individual sphere.

5.5.1. The knowledge value chain

In a service provision chain (also known as a value chain), knowledge usually flows from the producers to the major service providers, being managed only at the highest points, where the big data is collected. TAUS is an example of an institution that manages large amounts of data collected from many production outlets. Although there are ways for small users to benefit from this accumulated knowledge, the main advantages are only available to the members that have resources and tools to manage this data on a big scale. So, there is the notion that only at the top of the value chain is knowledge valuable. In such a context, knowledge bases might be just another data object that served to feed this chain. But things are not so clear.

The value of each link in a service provision chain is proportional to the knowledge it retains and manages. When a company or a person tries to manage more data than it has produced, it may fall victim to the data-dilution effect, or to the "poisoning its own drinking water" effect that was associated with Google's translation data management policies.

Personal data

Personal knowledge bases may be created around TMs, but they extend beyond that, to include terminological preferences, bilingual phrase alignments, but also the sites in which the user most frequently finds the solutions for his searches, information extracted from editing actions, and learnt parameters used to tune and improve the response of interactive systems. Each individual knowledge base is different, and not just because of the vocabulary it contains. A knowledge base built from frequent software localisation projects will most likely be very different from one built from legal documents, or from another one that mostly contains marketing content. The differences are not just in terms of vocabulary, terminology, or even "style", but also in editing patterns, context features and the different features that rank and score the results of interactive systems. The very specific structures of segments in any of these materials, and the way different MT systems handle them, make this information even more dependent on individual variation than TMs.

The contents of personal knowledge bases should not be shared, not only because of personal data and property rights protection, but also because as soon as this information leaves the personal sphere it loses its value. The data dilution effect explains it: at each level of sharing, data loses a degree of discriminative power. The conflict between data representativeness and comparability is another way to put it. This means that you cannot have a system that learns with you, a system that learns your style (as adaptive MT claims to achieve), add it to data from another user and get a collection of styles. The style learnt from each system is actually lost in the process.

Managing shared knowledge

This is not to say that these systems should create barriers to the spread of knowledge. At the level of project management, company, industry, collaborative or open-source project, specific level-bound knowledge bases should be built and managed. However, these knowledge bases will be more useful if they only import product data from other sources, and register its own process data. At the company level, for instance, more than having information on each translator's editing patterns, it is important to have data that allows project managers to measure their contribution to knowledge on the company data for a specific domain, or to the increase in fluency of a specific project. These knowledge bases may even be described as having a higher value, because they combine knowledge from a wider number of production situations.

The purpose of tools that manage knowledge is to develop a close identification between the user of a system and the knowledge he produces with the tool. Tools that manage knowledge in such a tightly integrated way reinforce the value of each link in the chain. This allows for each translator, or each company, or each project, to benefit from professional development, investment in specialisation and gained experience. In such an environment, tools must be transparent as to how this data is shared, giving control to the users over the definition of which information to share.

5.5.2. Knowledge-Assisted Translation

In the model of KAT – Knowledge-Assisted Translation that I, together with Luís Trigo and Belinda Maia, presented in 2016 (see section 2.3.4), the traditional CAT tools are surrounded by technology that manages this data as knowledge. The management of knowledge in the system described in this paper goes beyond data directly produced in the translation tool to encompass data associated with human resource and production management. The central category in this system is the technical domain: human resources, data resources, texts and systems are classified, selected and evaluated according to clusters of technical domains.

After having demonstrated that PE is a complex process and that the technical knowledge that was at the basis of the success of SMT should be put to the service of interactive modes of PE, this chapter presented a tentative proposal for such a system. The environment that surrounds an editor prepared for PE, translation and revision should be structured around the concept of knowledge management. KAT – Knowledge-Assisted Translation is just a proposal for the designation of those systems.

The focus of KAT is different from other approaches that see the MT system improvement as the ultimate goal of current technologies. This approach to knowledge management recognises the value of the individual dimension of the translation process and focuses on helping to improve it.

The development of systems like the one that is described in this chapter requires careful and extensive planning, and several competences that were not within reach during this dissertation. The next chapter describes the tests that were run to evaluate the propositions and claims that are put forward in this dissertation.

6. TESTING INTERACTIVE POST-EDITING

Chapter 2 mentioned that applied research projects should be assessed in terms of the usefulness and usability of its proposals. One of the initial proposals of this dissertation was that PE should be done with tools that understand and support specifically the actions that compose this process. The natural sequence for such a claim would be to test a tool that presented an implementation of this proposal. As demonstrated in the previous chapter, such a tool should incorporate a ML component that learnt the editing actions and presented editing suggestions, in an interface that should not be intrusive, but sensitive to word and phrase contexts. Throughout the duration of the project, it was not possible to identify a tool that complied with this requirement. Developing one that fulfilled these objectives was also beyond my technical skills. Nonetheless, it was necessary to propose the claim that editing is essentially composed of four editing actions, and to evaluate it in terms of its capacity to improve current PE work. Ideally, this claim should be tested and its relevance classified by experienced translators. This more attainable objective could be tested, as long as there was a tool that presented it in a form that allowed translators to consider it, after applying it in simulated PE assignments.

For this purpose, a specific PE working mode was implemented on a tool that was specifically adapted for a testing task (see section 6.1.1). However, while the tool that was used in the testing work did enable the evaluation of the usefulness of the software features proposed, it was not ready to be tested in terms of usability. This decision was based on two factors: due to time restraints, the tool did not include the learning and suggesting functionalities required by the PE support model, and the implementation of the editing mode based on four actions was not developed enough to be considered “ready for production”. The interface elements that composed the editing mode were intentionally intrusive, as appropriate for research purposes: it forced first-time users to pause and reflect on their decisions, thus improving the quality of their feedback. However, such intrusiveness, and the lack of suggestions for the four editing actions, made the interface unsuitable for use in full production mode.

So, the purpose of testing usability was removed from this stage. The testing procedures that were then planned focused on the acceptability and usefulness of the concepts that have been proposed in the previous Chapter.

Furthermore, as this was the first time that a PE tool based on four editing actions had been used, there were no parallel projects to learn from, and a new form of testing had to be devised. So, this stage of the project was approached in an exploratory way, as in a pilot test or a case study. These tests were seen as if they were part of an intermediate stage of a software testing process, although a proper software development process was not in place. As in a pilot test, the formulation of the concepts and the approaches to research were also under assessment, and feedback was asked from users on all those dimensions. In a possible extension to this project, the approach to the tests themselves would be improved, as presented in section 6.5 below. The project was also structured according to the characteristics of a case study, because there was no intention to generalise from the conclusions, and the project did not cover all the changes that might be required by the technology advance that was being tested (Kitchenham, Pickard, & Pfleeger, 1995).

It must also be said that this testing stage was not approached as a typical TPR project. One of the starting points was the acknowledgement that the working conditions of the tests would not represent usual working conditions. So, there were to be no claims about the ecological validity of the results, nor were the data collection methods designed to guarantee such a form of generalisation. The data collection was planned in an experimental way, as there was no guarantee of the capacity of the systems to collect and treat the data following the procedures described in the TPR literature, and the research goals were not the same of TPR. Nevertheless, this did not invalidate the research project, and very interesting conclusions could be drawn.

6.1. Post-editing tests

After some research, the option for the software application to be used in the tests was for a translation tool with an open development platform that could be adapted to projects such as this one. The chosen tool was HandyCAT, developed at DCU by Chris Hokamp (Hokamp, 2015a, 2015b; Hokamp & Liu, 2015). HandyCAT is presented as a "flexible web-based CAT tool for translation process research". This software is not a commercial tool, and it is available for download as an open-source project at https://github.com/chrishokamp/handycat. In "A Component-Centric Design

Framework for Translation Interfaces" (Hokamp, 2015a), the architecture of this application is described, including how it can be used to test different editing modes.

Chris Hokamp agreed to adapt his application to suit the needs of this research project. However, since the timeframe did not allow for extended development work, priority was given to developing an interface that allowed users to use and test the four editing actions against the tool's default editing mode.

After HandyCAT's new features were considered stable, a workshop for testing the tool and getting feedback from participants was prepared. An invitation for the workshop was launched to professional translators with experience in PE. Around 60 people enrolled, of which 50 attended the workshop.

To extend the reach of the tests, participants were asked to fill in two questionnaires, with different purposes: to collect data on their professional profiles, to gather their assessment of the usefulness of the concepts proposed by this project, and to ask for comments on the usefulness of an eventual extension of the PE tool. The next sub-sections describe the preparation and objectives of these tests.

6.1.1. The testing tool (HandyCAT)

Chris Hokamp's specific implementation of HandyCAT (version porto_v0.1) is available here: https://github.com/chrishokamp/handycat/releases/tag/porto_v0.1.

The "Quick Start Guide" included as Appendix 1 to this dissertation describes a step-by-step approach to the intended use of the software in the project workshop. After being installed on a server, HandyCAT may be accessed and used through a web browser.

HandyCAT version porto_v0.1 features two editing modes:

- **Autocomplete mode** (AC mode) – this is the default work mode in HandyCAT, which includes a predictive writing support system that presents word suggestions as translators write their translations;
- **Post-editor mode** (PE mode) – this is a purpose-built interface that conditions translators to reflect before selecting each of the four editing actions for their editing work.

In the default AC mode, translators edited the MT hypotheses in the target window, with no restrictions to their usual editing techniques: they could delete all words in the target window, use overwrite mode, or copy and paste words. However, when they started typing a word, a TL dictionary suggested a word to be auto-completed, which they could confirm by pressing a single key (Enter). This dictionary consisted of a frequency list formed by the 25,000 most used words in Portuguese (no variant) from the MT corpora in OPUS (Tiedemann, 2012). The predictive writing feature was not sensitive to the context that surrounded the typed word in the target window. However, there was a word alignment function with the ST side, but this was not tested. Auto-complete suggestions responded to each new character typed, which resulted in more accurate suggestions or alternatives as the translator typed more characters in each word.

PE mode was designed to force the users to apply one editing action at a time. The interaction element that was proposed (a right-click contextual menu) implied that all actions were performed in two-steps: first selecting a token, and then applying the editing action to that token. This selection technique was based on strict tokenization, by word, space and punctuation mark (the effect that this had on the working method is explained in detail in the guide in Appendix 1). This meant that editing actions were slow and required a good control of cursor positioning and mouse movements or keyboard actions. Furthermore, secondary editing actions (movement and replacement) lost their efficiency advantage. The counter-intuitiveness of this method was deemed appropriate for the tests, since this would force a conscious and critical process from the translators, as intended.

The version of HandyCAT that was used in the workshop did not feature the functionalities that would be required for a fully-fledged test of an interactive editing tool. However, this may be added to the application for other research projects. Some of these missing features include a TM and the related segment-matching functions, a MT engine, or a ML system to learn and estimate the editing actions. Features that depend on these, such as concordance searches or sub-segmental alignment were not implemented either. The application featured a dictionary search functionality and there was an indication of word alignments between the source and target windows, but these could not be tested and adjusted for use in this testing environment. However, although this implementation of HandyCAT could not be considered fully interactive, it provided the fundamental editing actions that we wanted to test, and it added the default predictive writing mode.

6.1.2. Sampling

The selection of every element of this research was non-probabilistic and purposive, as described by William A. K. Trochim, in the “Research Methods Knowledge Base” available online (Trochim, 2006). The initial purpose of the project was to test specific functions with specialised users, in circumstances that had been set up to allow for the observation and collection of specific details. So, it was considered appropriate to apply a sampling method that traded statistically uncertain representativeness for expert and informed responses. This sampling method encompasses the selection of the participants, the selection of the texts used in the workshop and the preparation of the data collection methods.

In the case of the participants, an expert and snowball methodology was applied. Taking into account the focus of the research, a panel of experts was seen as the best way to reach an adequate level of response. Since professional translators usually work in networks, the invitations to participate were launched by both targeted messages and social media, with several of the first participants then inviting other professional translators to enrol as well. The decision to not analyse the results in terms of their representativeness of the population is also justified by the fact that there are no clear data on the composition of this population. There are, for instance, no official statistics on the numbers of professional translators who accept PE jobs, as opposed to those who do not accept them. Besides, usual factors like age and gender are also not studied in terms of the representativeness and of their effects on the results, so these were not included in the collection of data either. The only factors that might bring up some interesting information to the data collected were related to the experience and initial attitudes of the translators towards translation technology and PE. These, instead of being included in the sampling stage, were considered as variables for analysis.

For the selection of the texts to be post-edited, a heterogeneity sampling methodology was applied. The purpose in this case was to have a good variety of texts in the small number that would be possible to handle in a brief workshop. The focus was on technical texts that were part of real translation projects, provided by anonymous sources. The texts were briefly analysed in terms of their textual structure and technical domain, so as to guarantee that the sample was varied. These texts are included as Appendix 2, and their processing is described in more detail in section 6.1.4.

For the data collection methods, the decision was taken not to identify the subjects of the workshop, so as to elicit unconditioned responses from them. It is important to stress that professional translators are under constant pressure to show that they are capable of accompanying all technological developments, while, at the same time, they feel that they should defend linguistic quality against any attempts to diminish its centrality in their work. These pressures gain a special importance in the context of hired translators working in translation companies, and in training contexts. Since several participants were working in these conditions, and the workshop was held at the Faculty of Arts and Humanities of the University of Porto (an important translator training school to which several participants were connected), it was deemed important to give the reassurance to participants that there would be no attempt to link their responses to their identities. So, an anonymity system was devised to permit the association of the responses in the different data collection methods to each user, while keeping the user's identification anonymous. In addition, the different texts and editing modes were randomly distributed to all participants, so that the results showed a balanced representation of each.

6.1.3. Data collection methodology

A mixed methods approach was selected for data collection, in which both qualitative (open questioning and behaviour observation) and quantitative (scaled survey and quantified actions) data was retrieved. The complementarity of the data collection methods was intended as a reinforcement of the information retrieved from each method, in view of the exploratory nature of the tests. Although there are several points of contact between the methods employed and the different mixed methods strategies described, for example, by Creswell (2003), the design and planning of this research did not follow specifically any such strategies. This is due not only to constraints in terms of planning, but also to the open nature of the questions which were being put forward and the unknown relative importance of the data to be collected.

As stated above, this mixed method approach did not envisage answering the typical TPR questions presented in previous chapters of this dissertation. The aim of the questionnaires, for example, was not for translators to describe their working processes retrospectively, but to register their impressions on the features they had tested. The activity logs of the software were not used either to accurately measure productivity, to identify professional profiles or to assess quality and classify production errors.

This was not a straightforward project, since it demanded a flexible approach and constant learning. Nevertheless, the data collection methodology fulfils the principles of scientific research, presenting valid, accurate and reliable data, which may hold interesting research questions for further investigation. As in a pilot test, the data classification categories had to be adjusted for each set of data collected. And, as in a case study, the data analysis processes described in sections 6.2 to 6.4 below, although hard to interpret, allows for the identification of trends and features that provide a good input about the research questions and proposals.

6.1.4. Setting up the tests

One of the first stages for preparing the tests was to choose the texts to be post-edited. Four texts were used, all extracts from longer PE assignments. All texts were anonymised, by replacing any references to commercial names with <tag>. Then, they were pre-translated with MateCAT (Federico et al., 2014), and converted to XLIFF, so that they could be post-edited in HandyCAT. The table below presents a brief description of these texts.

ID	Description	Total words	Words in fuzzy matches	Words in repetitions
Text A	Mobile phone instructions	733	42 (6%)	0
Text B	Marketing questionnaire	839	61 (7%)	19 (2%)
Text C	Product catalogue	974	45 (5%)	32 (3%)
Text D	Instructions manual	503	66 (13%)	10 (2%)
Total		3049	214 (7%)	61 (2%)

Table 14 – Source texts used in the workshop

The contents of these texts, the full analysis made by MateCAT and the MT hypotheses created by MateCAT are all presented in Appendix 2. Table 14 above presents values for full and fuzzy matches against a public TM, as measured by MateCAT. In Appendix 2, besides the analysis, there is also an explanation of the way MateCAT estimates the productivity gains the tool provides, which shows that fuzzies are only considered above the 75% similarity score.

The percentage of fuzzy matches and repetitions in these texts is not very high, but this is not an important feature for this project, since there was no TM in any stage of the process. An analysis of the complexity and lexical density of the texts was made with Text Inspector (an online tool available at: http://textinspector.com/), but the results obtained were not relevant to produce a reliable classification of the texts. The small size of the sample and a relative similarity between the texts was the most probable cause of these results.

Data collection questionnaires

The second step was preparing the questionnaires, using "Google Docs forms" functionality. The first questionnaire was distributed to everyone who enrolled in the workshop, for completion before the workshop. The aim of the questionnaire was not only to collect data from the participants, but also to give a taste of the work that was planned for the workshop. The questionnaire was accompanied by an explanation of the purpose of the workshop, of the data collection that was planned and of the fact that all participation was anonymous and voluntary. It was also explained that the workshop was targeted at professional translators with PE experience. This questionnaire is presented below as Appendix 3.

Based on this first stage of information, several users asked for confirmation on whether they were eligible to participate in the workshop, due to their lack of experience in PE. At this stage, a few users cancelled their enrolment, but others, though they did not really correspond to the intended profile, still showed a big interest in participating. Some of these were translation company owners interested in the evolution of PE, translation technology teachers and freelance professional translators with no or little experience in PE. By accepting these users, a new adaptation of some of the contents and purposes of the workshop was required. The full characterisation of the participants in the workshop, based on all data collection methods, is presented in the next sections.

Groups of users and work sessions

A plan to divide the participants into 3 groups during the workshop was prepared. Each group would be assigned a different sequence of texts for four different sessions, in which they would apply the two editing modes provided by HandyCAT.

Before the proper testing stage, in which all activity data were collected, there was a familiarisation stage, described in the next section. Text A was used for this familiarisation stage only. Since it was clear that an adequate amount of time should be given for this familiarisation, the total testing time had to be reduced, from the planned 60 minutes (with a few breaks) to ca. 35 minutes.

This was the plan for the four work sessions:

- Session A: first, they would edit a text using AC mode for 10 minutes;
- Session B: next, they would edit another text in PE mode for 15 minutes.

Then, they would edit the third text, in two sequential sessions:

- Session C: first, the text would be edited in AC mode for 5 minutes;
- Session D: and then the same text in PE mode, for another 5 minutes.

This would take a total of 35 minutes, with PE mode being used for more time (20 minutes, and AC 15 minutes), because that was the new method that was being tested. For the proper testing stage, there were only three texts (B, C and D), which meant that all groups processed all texts, and all texts were processed in all sessions and modes. The table below describes the planned distribution of users and texts, based on an expectation of 51 users.

	Session A		Session B		Session C		Session D	
	Time	00:10:00	Time	00:15:00	Time	00:05:00	Time	00:05:00
	Mode	AC	Mode	PE	Mode	AC	Mode	PE
	Group	Nº users	Group	Nº users	Group	Nº users	Group	Nº users
B - Questionnaire	1	17	3	17	2	17	2	17
C - Catalogue	2	17	1	17	3	17	3	17
D - Manual	3	17	2	17	1	17	1	17

Table 15 – Planned distribution of texts and users for the workshop.

This distribution meant that all texts should be edited in four sessions by 17 users, totalling 34 users per pair of text and mode. The total time, from all users, in AC mode should be over 12 hours: 12:45:00. For PE mode, there should be 17 hours of editing. The following table represents this distribution and totals.

	AC		PE		
	Users	Time	Users	Time	Total time
Questionnaire	34	04:15:00	34	05:40:00	09:55:00
Catalogue	34	04:15:00	34	05:40:00	09:55:00
Manual	34	04:15:00	34	05:40:00	09:55:00
Total times		12:45:00		17:00:00	29:45:00

Table 16 – Totals distribution of texts and users for the workshop.

Next, we may see the distribution of times and texts as planned.

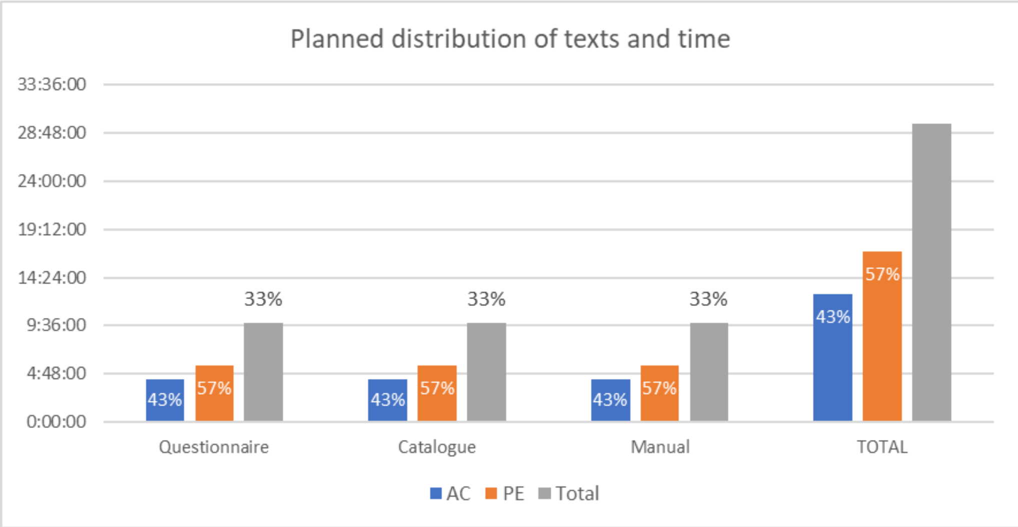


Figure 9 – Planned distribution of texts and total times.

This shows how AC mode would account for 43% of the total time and PE 57% of the time, for each text, and in total. Each text would be edited by the same number of users in an equivalent number of sessions, thus representing 33% of all texts. So, in the end, only the difference between the two editing modes would remain.

6.1.5. Description of the workshop

The workshop was held in two different sessions, each lasting about 3 hours, on the same day (7 January 2017). 26 people attended the morning session and 24 the afternoon session, in a total 50 people.

Workshop presentation

The first part of each session of the workshop was a presentation on “Interactive Machine Translation and Post-editing”, from the point of view of research on translation tools (Carmo, 2017). The presentation is included in this dissertation as Appendix 4. The presentation included periods of discussion on themes such as methods of work in PE with the available tools and effects of PE on the translation industry.

The proposal for the redefinition of PE as being composed of four editing actions was introduced at the workshop, as presented below. This definition was later refined, until it reached the form that is now presented in this dissertation.

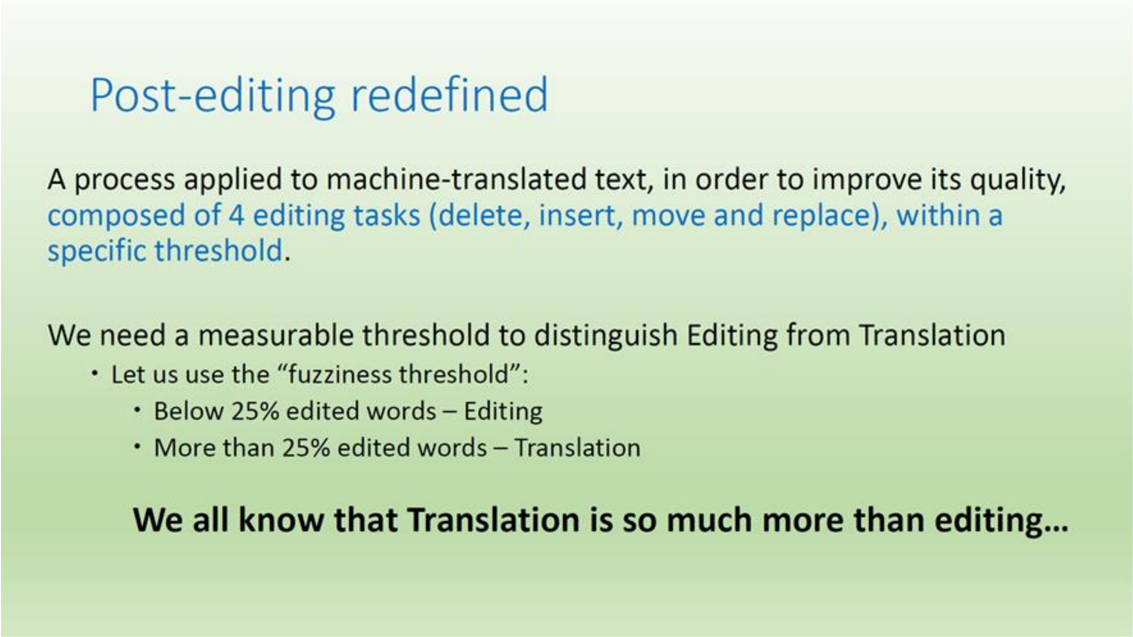


Figure 10 – The redefinition of PE, as presented at the workshop.

During the presentation, participants accessed the second questionnaire, and in different pauses during the presentation, they were asked to answer questions concerning their impressions on the contents of the presentation and on the different views that were discussed. The presentation included a description of the project for a complete interactive PE tool, and described the tests that were to be done at the workshop. It also explained the type of data that HandyCAT would collect from the participant's actions, and that there would be no measurements or assessments of translation quality or individual productivity.

Familiarisation session

After this presentation, participants filled in the first part of the second questionnaire. After a short break, there was an introduction to HandyCAT, following, step-by-step, the Quick Start Guide that had been distributed. Then, the participants had about 30 minutes to familiarise themselves with the two editing modes. For this familiarisation, users worked with Text A, in two different work sessions, one in each editing mode.

The focus during the familiarisation with HandyCAT was on the selection of tokens and the different methods for the four editing actions in PE mode. The explanation of the four editing actions followed the division between primary and secondary actions, as described below:

- **Delete** should be used when the translator identifies words that do not belong to that sentence, in a position that is not going to be occupied by another word;
- **Insert** should only be used to insert words in positions where they are missing;

This meant that translators should not delete a word in one position and then insert another in the same position, as this should be done with the Replace action. And they should not use Delete to delete one word in one position and then use Insert to insert it in a new position, as Move should be used for that.

- **Move** is useful when the right word is in the wrong position. In this case, instead of deleting and inserting, the translators should select the word, choose the Move option and then click on the new position;
- **Replace** is used to make changes in words in the same position. These may be small changes, like capitalisation, or number, gender, tense or other type of inflection, or to replace the whole word by a synonym or even a totally different word.

These guidelines are important to enable the application of efficient methods like moving and replacing, which join two actions into one. However, it was clear that it might be difficult to make sure translators followed them. This would need to be further explored, namely in the context of a tool that presented suggestions for these edits. In this context, it was decided not to introduce the proposal to test the order of the editing actions, suggested in section 5.3.2.

Work sessions

After the familiarisation stage, users participated in four work sessions using the two editing modes (AC and PE) alternately. Before the actual work, participants received all instructions concerning purposes and the data collection methods again. It was explained that there were no targets set in terms of productivity, but that there was a timeframe for each work session. It was clarified that the purpose was for translators to test as many situations as possible during that timeframe, so as to build a solid impression on the editing modes being used, and to be able to give their impressions on the editing modes in the final part of the questionnaire. No instructions were given in terms of the extent of PE, concerning grammar, style or any other features.

During the work sessions, one of the servers in which HandyCAT was installed stopped and a few users had to change server. Some of them also changed groups because of this, and others did not choose their respective texts correctly. In this process, some of the activity logs were lost and other logs did not include the expected sessions and edited texts. The effects of this are reported below in section 6.3.2.

After the four work sessions, participants registered their impressions and the assessment of the workshop in the final part of the questionnaire. At the end of the workshop, participants closed the online questionnaire and delivered the activity logs. The second questionnaire is included in this dissertation as Appendix 5, and a sample of the activity logs in JSON format is presented as Appendix 6.

6.1.6. Data analysis methodology

After the workshop, the data analysis stage began. The methods employed were very much influenced by the available tools, and they will be described in detail in the sections below. Most of the data was converted, cleaned and processed in Microsoft Excel. Within this tool, simple data analysis processes were performed, like estimations of totals and averages for the main variables, clustering and identification of sets of cases according to variable distribution and conversion to simple data visualization formats.

Section 6.2 describes the preparation of the data and the results of the answers given to the two questionnaires, and section 6.3 describes the different tasks performed on the data collected from the activity logs and describes the results from these.

After this detailed description of the results, section 6.4 describes statistical tests that were applied to the data, in order to identify the main correlations between the many dimensions that were identified both in the data collected from the questionnaires and from the activity logs. On the final section of this chapter (6.5), the main conclusions from this work are presented.

6.2. Presentation of results from questionnaires

The participation in the workshop was very high, which allowed the collection of plenty of data, both in close quantitative data and in open-ended qualitative data. All data collection was anonymised to guarantee that all participants gave their answers freely, without concerns about representing their professional roles or statuses. This section explains how the questionnaires were prepared and presented, and describes the results collected with this method, aggregating the results by themes.

6.2.1. Preparation of data for analysis

Questionnaire 1 was distributed and filled in by each participant online before the workshop. Questionnaire 2 was filled in at different moments throughout the workshop. The answers to both questionnaires were collected in Google Docs forms. They were then exported and treated in Microsoft Excel. Both questionnaires included several qualitative questions, open to free transmission of impressions by participants. After collection, these were codified according to the most relevant categories for the research questions.

In sections 6.2.2 to 6.2.5 of this Chapter, answers from questionnaires are presented in the form of charts. In these charts, questions are identified by the number of the questionnaire followed by the number of the question in that questionnaire (e.g. Q1.2 represents question 2 of questionnaire 1). The order of the questions in the questionnaires is not kept, in order to enable the clustering of the themes in a descriptive way. The text of the questions and options is presented in a simplified form. The full text of the questionnaires is presented in Appendices 3 (Questionnaire 1) and 5 (Questionnaire 2), and the results from both questionnaires are presented in Appendix 7.

6.2.2. Characterisation of participants

The first purpose of the collection of data in questionnaires was to analyse patterns of professional profiles and attitudes towards technology.

Professional profiles

The next three charts present data that characterise the professional profiles of the participants. The first chart shows that there was a good mixture of experience, from translators with less than a year to others who had worked for more than 20. The second chart shows that most of the participants were freelance translators (ca. 58%), with the next biggest group of users working at translation agencies (36%). Only 6% are hired at companies or institutions other than translation agencies. This means that 42% of the participants have an employment work relationship, against 58% who are freelancers.

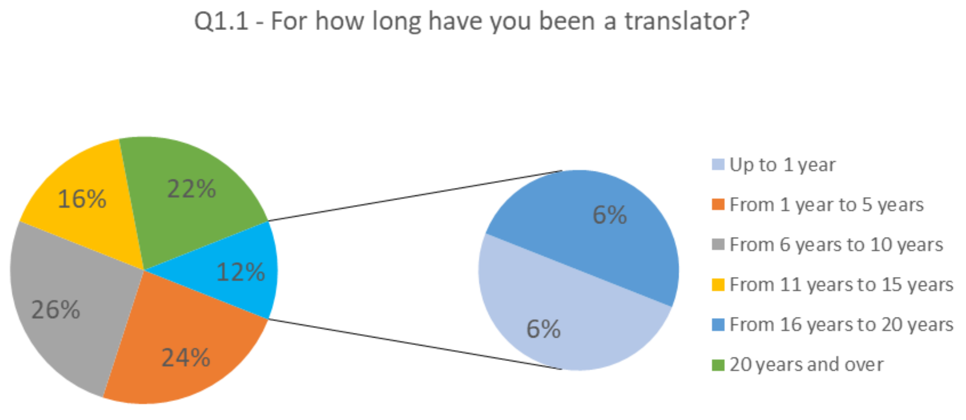


Figure 11 – Distribution of experience in the sample (Q1.1).

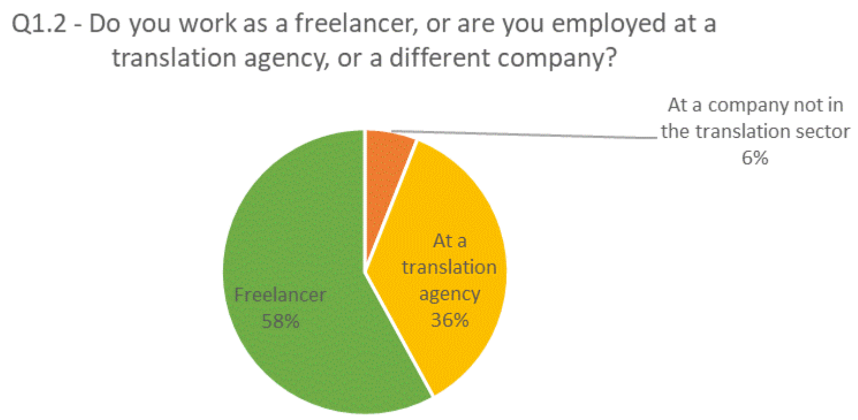


Figure 12 – Professional status of respondents (Q1.2).

The next chart shows that translation into the mother tongue (L1) is the most frequent task for virtually all participants, revision coming in second place. PE is the fourth most frequent task, coming behind "other language tasks", and the least frequent is translation into a second language (L2).

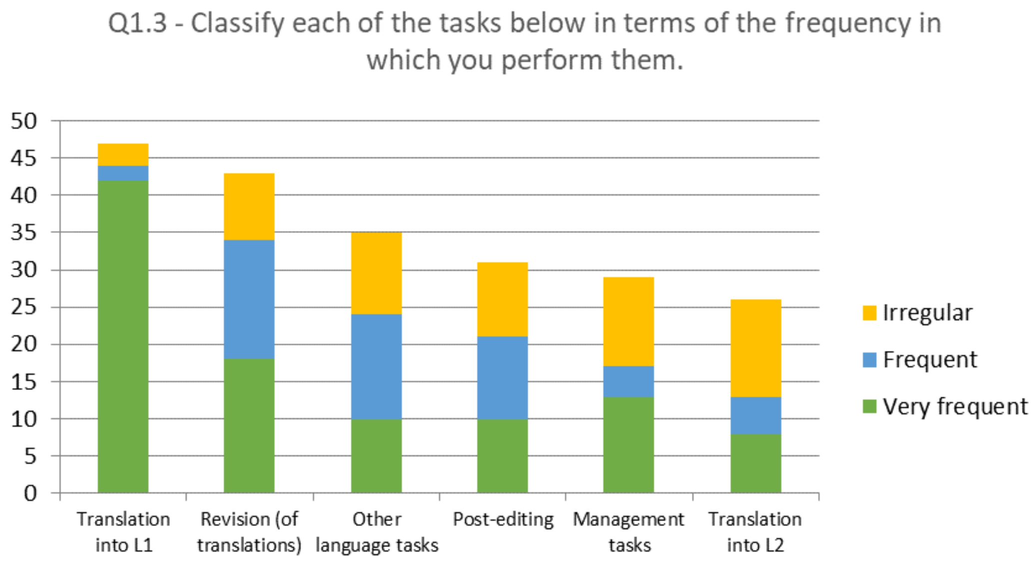


Figure 13 – Frequency of tasks performed by respondents (Q1.3).

Use of translation technologies

The first question in this set concerned the relationship with translation technologies in general. One can conclude that the vast majority of participants had a good relationship with technology:

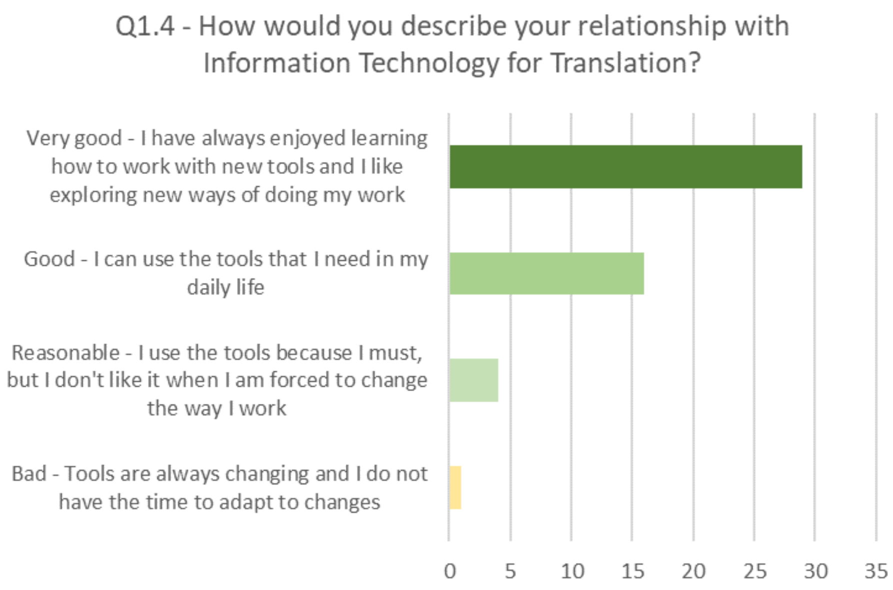


Figure 14 – Relation with IT for Translation (Q1.4).

In section 4.2, it was claimed that the separation between interactive and pre-translation mode does no longer describe the complexity of methods that are used nowadays by translators. The next questions address different dimensions of how translators work with tools that offer them a wealth of aids, and which may alter their working modes. Question 5 takes the simplified view of those two modes and the results were somewhat surprising. The next questions detail this a bit more.

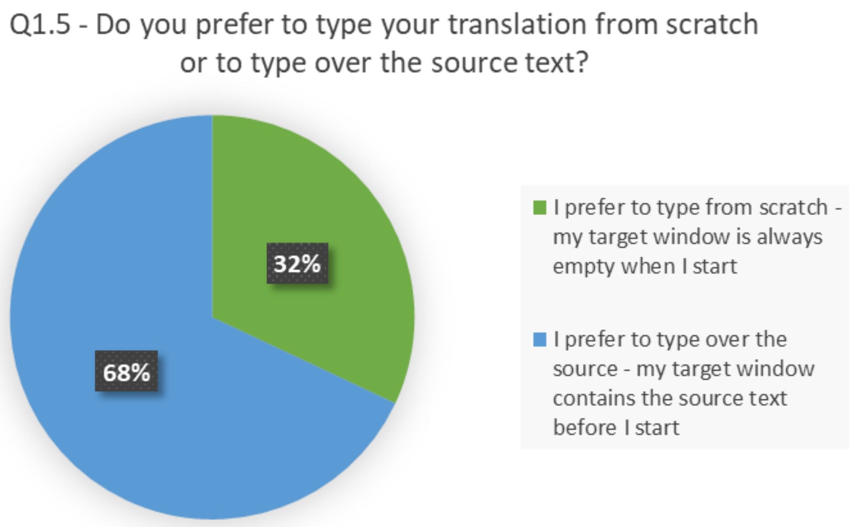


Figure 15 – Interactive vs. pre-translation mode in CAT tools (Q1.5).

As we can see in Figure 6 above, 32% of the participants prefer to work in interactive mode, with no ST to type over. Taking into account all the available features offered by CAT tools, and the way CAT projects are commonly set up, it was not expected that the number of translators typing the whole translation was so big. Nevertheless, the vast majority of users are more adept at the pre-translation mode, overwriting the ST segments.

Either way, most translators in this sample take advantage of the suggestions presented by CAT tools, with most of them choosing the option that considers that sometimes these features are intrusive, as may be seen in the chart below.

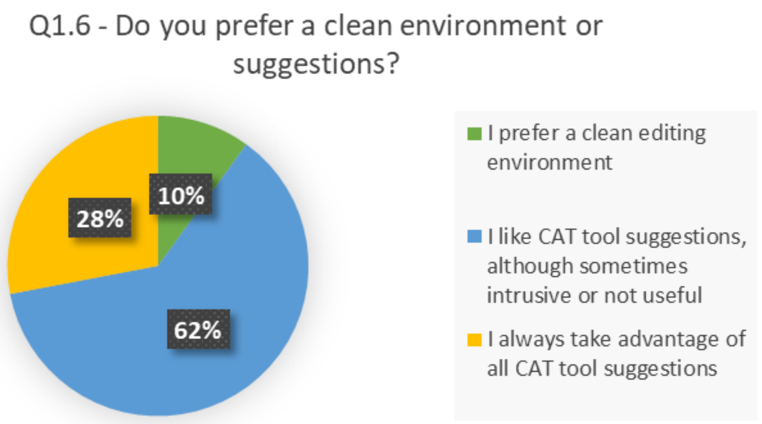


Figure 16 – Taking advantage of suggestions in CAT tools (Q1.6).

When asked about features that have been presented as belonging to "second generation" CAT tools, nearly half of the participants said that they only use them in specific situations or not usually. The chart below includes the two questions that focused on these features (Q1.7 and Q1.8), one about predictive writing features, which may be seen as implementations of sub-segment alignments, and the other on the use of MT suggestions.

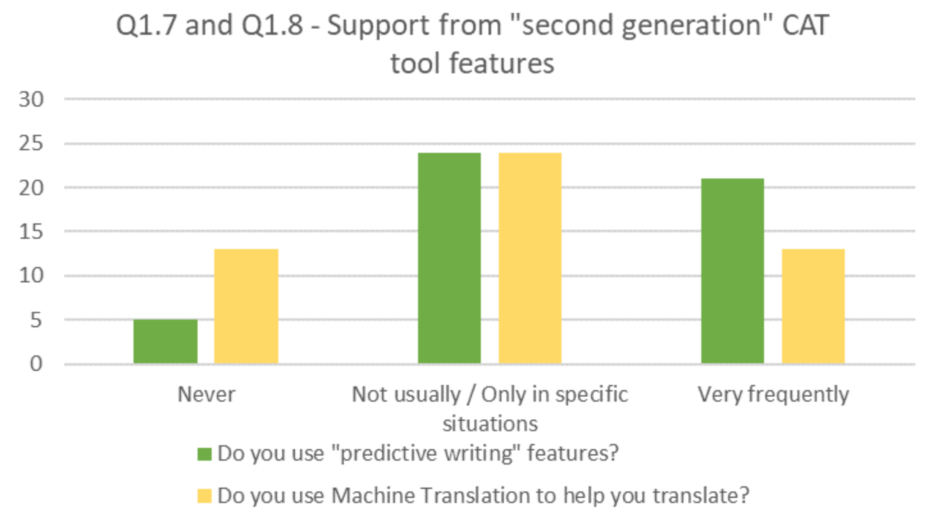


Figure 17 – Support of second generation features (Q1.7 and Q1.8).

The number of users reporting the use of predictive writing supports and MT suggestions as support forms a majority over those that say that they never use such features. This is clearer for predictive writing than for MT suggestions. Besides, as may be seen next, the answers to the MT use by translators can only be read as a tendency, because part of the answers do not match the question.

Use of MT as an aid and in PE

Question Q1.8 focused on the use of MT as an aid to regular translation work, not as part of a PE project. Still, several users understood it in this second sense, as expressed in Q1.8b. This was an open question directed to participants who had replied that they used MT as a support to their work. For the analysis of all replies, these were codified and clustered around the central concepts presented by participants as the reasons why they use MT as an aid. 36 participants replied to this question, since the other 13 said they never used MT in this way. Of these 36 answers, 10 involved using MT in PE projects, and 4 mentioned a constant use of MT. These were considered invalid replies (in dark orange below).

The factors that affected the decisions to use MT (in blue below) included the type of text (technical texts as the main type that calls for the support of MT), the number of words and repetitiveness of ST (long and repetitive texts are seen as a good reason to use MT) and only one user said that whenever the quality of the MT hypotheses justified it, he used MT. The reasons for using MT (in yellow) were mainly to increase productivity and to look for words. When looking for words, some said that they looked for specialised words, while others said that they never used MT to look for specialised vocabulary. A few said that they used MT to look for the translation of Named Entities, such as names of countries. In terms of the TP, there were replies that related the use of MT to an aid to the interpretation of the ST and to writing the TT, the latter mainly associated with writing in a foreign language (L2).

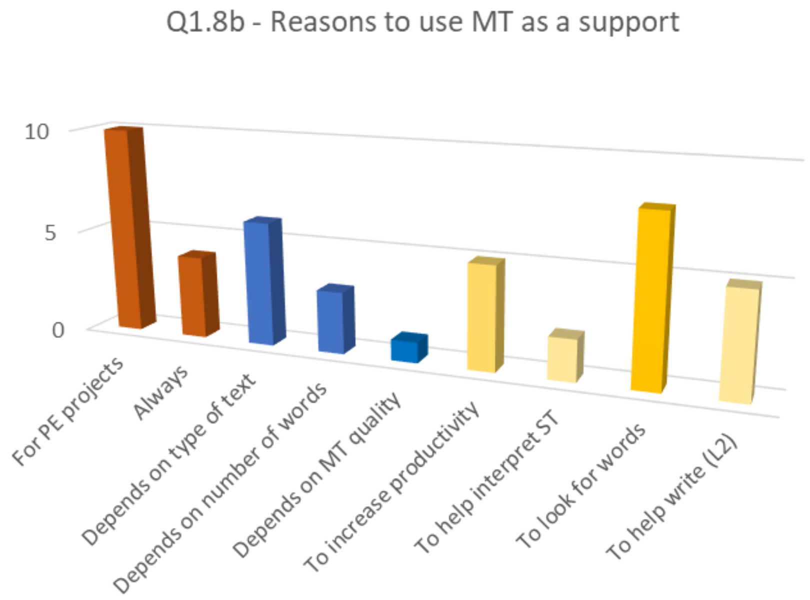


Figure 18 – Reasons to use MT (Q1.8b).

The next question queried participants about their experience in PE. As may be seen below, more than 50% considered that they had "some" experience in PE, while an extra 23% said they had a lot of experience.

From this point onwards, the questions focused on details of the participants' experience and impressions of PE. So, it was decided to divide the group between those with experience and those without experience. Still, the issues that were focused by the questions were the same.

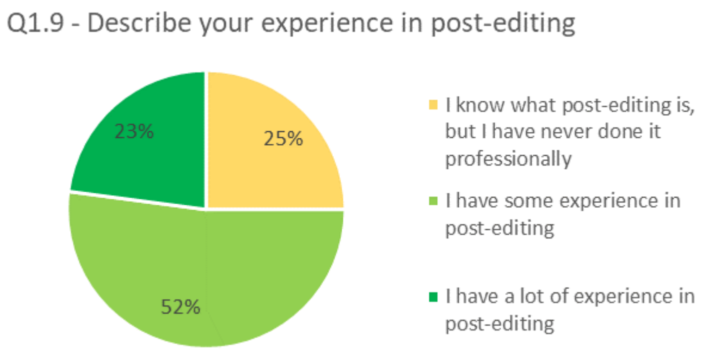


Figure 19 – Different levels of experience in PE (Q1.9).

After being asked whether they had experience in PE, the participants were split into two groups. The two groups replied to the same questions, but they chose between options in terms of experience-based opinions or expectation-wise impressions. The first two questions focused on productivity and regularity of PE projects.

Apart from the first bar below (Figure 12), which refers to the users that do not do PE, the majority of participants reported productivity between 100 and 1000 words per hour. As for the last bars, with respondents claiming that their productivity per hour is above 2000 words, this can only be associated with a too hasty reading of the question, which may have been interpreted as concerning productivity per day. A closer look at the two highest bars is detailed in Figure 13, showing that there is a normal distribution of responses for each level between 100 and 900 words, with most numbers between 400 and 700 words per hour.

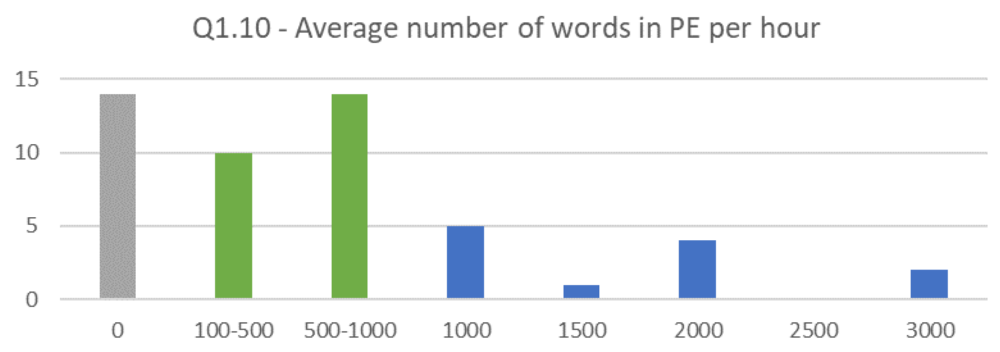


Figure 20 – Hourly productivity in PE (Q1.10).

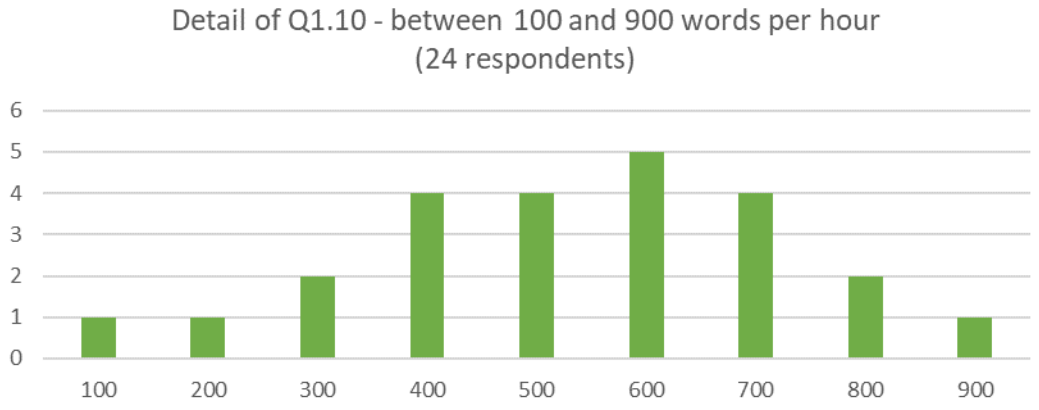


Figure 21 – Central distribution of productivity in PE (Q1.10).

The next question projected the timeframe into a month, not in the perspective of productivity, but of frequency and regularity of PE work. In this case, most respondents report that they produce on average between 1000 and 25,000 words, in different PE projects per month.

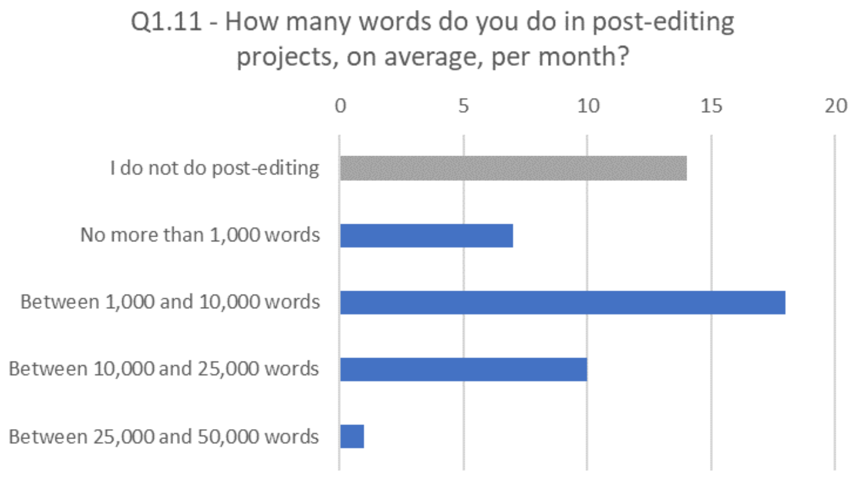


Figure 22 – Regularity of PE projects (Q1.11).

6.2.3. How translators view post-editing and the industry

The next section of questionnaire 1 focused on the views of participants on PE and the translation industry.

Introducing discussion themes

The next three charts show the results of questions that not only intended to capture impressions from the participants, but which also aimed at introducing topics that might inform their positions and responses during the workshop. This led to a few participants reporting that some of the questions conditioned their responses too much, but this was compatible with the purpose of enlarging the content of the answers that were expected later in the workshop. As seen above in Q1.8b, none of the respondents associated cognitive effort as a reason to use MT as an aid. This was one of the items that needed to be elicited in the responses by an instrument such as this questionnaire.

For the first question, participants only had two options, according to the experience they had reported. Translators with no experience only replied in terms of "I think I would like… / I think I would not like…" (these are represented by the central bars in the 3 charts below), whereas the translators with PE experience chose whether or not they liked PE. In both groups, the majority stated that they liked or would like to do PE.

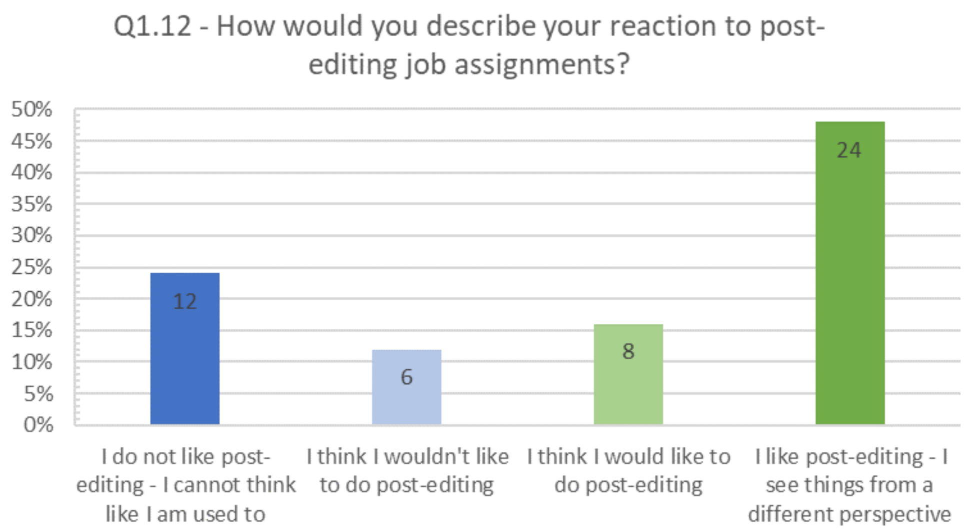


Figure 23 – Reactions to PE job assignments (Q1.12).

The options available in this question were intentionally conditioned to criteria related to how PE affects the way translators think while doing PE, so as to encourage reflections about the cognitive effects of PE. The same effect was sought in the next question, which focused on the perception of the comparison between PE and translation.

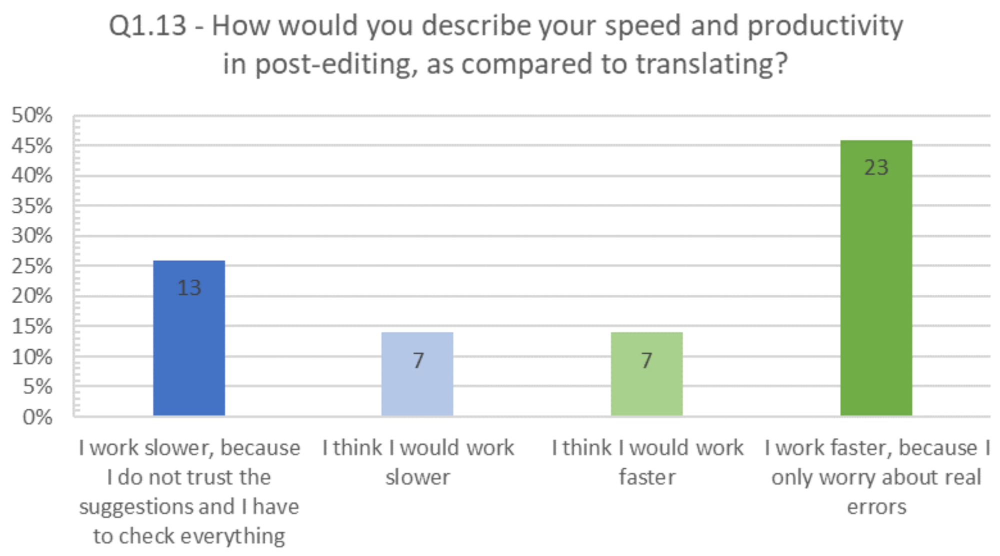


Figure 24 – Comparison between PE and translation (Q1.12).

In this answer, the numbers were very similar to the previous one, and translators chose more often the options associated with an increase in speed and productivity. The suggested options related this with how much attention was paid to errors and the extent of the checks translators did.

The next question focused on the impact of remuneration on the dedication to PE projects. The distribution was again very similar, with translators choosing more often the options that show that low remuneration does not affect the dedication to PE projects.

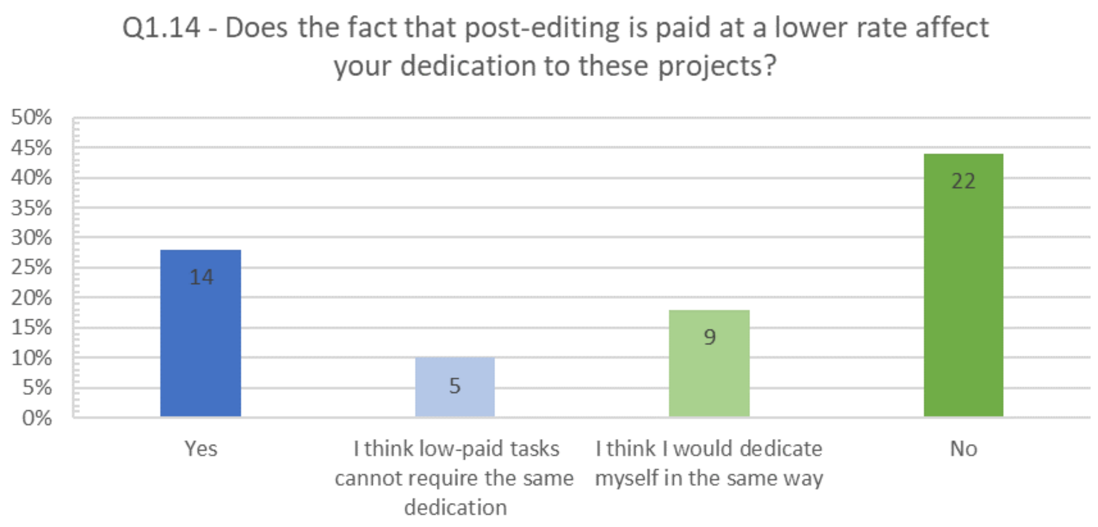


Figure 25 – The impact of remuneration on dedication to PE (Q1.14).

The answers to these three questions show a positive attitude towards PE: translators who like PE, those that do PE faster than they translate, and those that dedicate in the same way to PE and translation projects, even when PE is a lower paid task, are virtually the double of those that do not like PE, are slower doing PE and who admit that low pay affects their dedication to PE assignments. The percentages of these negative answers in this set of questions are steadily around 25% of the whole sample, which reflects that these concerns are still very important for the global evaluation of the impact of PE on professional translators.

The first part of questionnaire 1 raised several questions that were worth further study, if the aim of the project was to fully characterise and classify post-editors' behaviour and attitudes. Some of these include:

- the interactive vs. pre-translation modes should be reassessed in view of the new tools that translators use;
- the understanding of the reasons why MT may be used as a support, instead of as part of a PE project;
- or the way translators think during PE, the attention they give to errors and the impact of remuneration.

In this project, these questions were mainly used to raise awareness for some of the items that inform the post-editors' position in relation to PE and how it is done, so as to elicit more informed feedback on the editing modes they were going to use during the tests.

PE, revision and translation

The next section of questionnaire 1 posed the issue of the relation between PE and revision, and between PE and translation.

Question 15 was presented in the following terms: *Q1.15 - According to industry standards, translation should always be followed by a revision. What about post-editing? Does it produce "enough" quality to dispense with revision?* (The text of the question and of the optional responses is presented in the chart below in a shorter version.) The results were very evenly distributed, between options that clearly stated that PE always requires revision to others that admitted that sometimes PE dispenses with revision. One of the options was a good refuge for respondents who did not want to commit to a view of PE requiring revision: *Post-editing should always produce the same quality as translation, so it does not depend on the process whether the target text* *dispenses with revision or not.* Even so, this option had a similar number of responses to the other ones. More impressive is the fact that the option which identifies PE as revision (*Post-editing is already a type of revision, so it does not require another revision)* did not receive any answer.

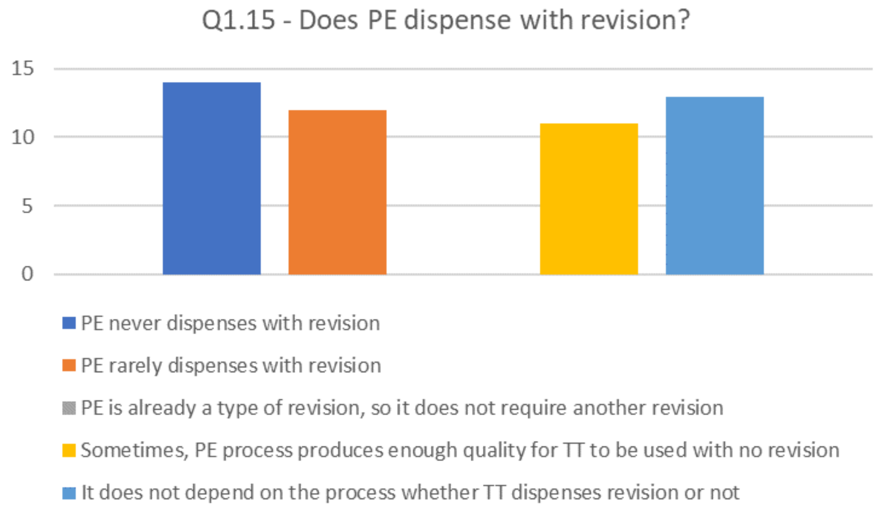


Figure 26 – The relation between PE and revision (Q1.15).

Question Q1.16 asked respondents to present a description of the differences between translation and PE. This was an open answer, with no suggestions or options conditioning the response from participants. The answers collected were treated as qualitative data and codified according to the most important elements the users decided to highlight. Here is the result of this analysis, in which the same answer may be represented in different bars, because it included several of these elements:

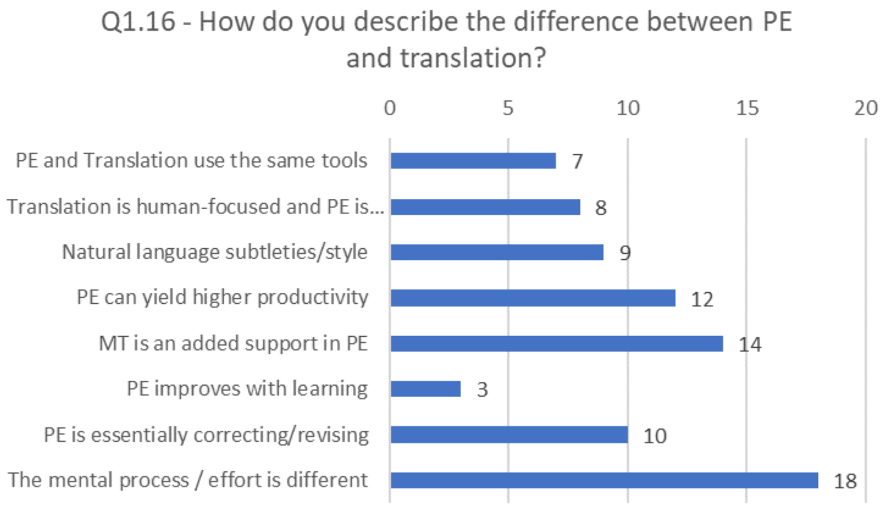


Figure 27 – The differences between PE and translation (Q1.16).

Some of the respondents referred to the fact that the tools used for these processes are the same (7 answers), and one even said that there was no need for specific tools for each process, admitting, however, that the approach and the techniques were different. Several answers had references to the focus on human processes in translation and on machine processes in PE (8). Others referred to the difficulties posed by natural languages and style, specifically in certain types of text for machine processing (9). In a bigger number of answers, there were references to the usefulness of PE in terms of increased productivity (12), and as an added support to features provided by CAT tools (14). Only 3 respondents mentioned that the quality of results of MT improves over time, one of them stating that this was one of the criteria for him to limit the scope of changes that he made to MT hypotheses. Although several answers stated that one of the differences between PE and translation is that PE is more of a revision process (10), we have seen in the question before that no one admitted that PE was a process that dispensed with revision. But there were other descriptions of the distinction between these processes in interesting terms. One user said that PE is a *language service, instead of a translation job,* and that it should be performed by translators. Another answer called PE, in opposition to translation as a creative and critical thinking process, a *guided exercise of informative writing*.

The items most frequently referred to when distinguishing PE from translation were editing effort and processes (18). One user wrote that PE required more effort, and another one mentioned the need to pay more attention to what he read, in order to identify all mistakes. One of the respondents mentioned differences in how sentences need to be structured, how to edit text, and even how to organise ideas, adding that this might involve a higher risk. This risk is referred to by another answer as *"shared responsibility"* (while translation involves *"full freedom and total responsibility"*. Another respondent said that this risk was relying on the suggestion and only adapting the wording of the text. In Q2.9 (see section 6.2.5 below), a user refers to "*feeling lost in the chaos of a translation to be post-edited".* Another user wrote: "*Post-editing can be misleading and more unreliable than translation. As there is already a translation suggestion available, it is very easy to leave that suggestion as it is and not to perform research to confirm if the suggestion is correct".* Finally, a translator wrote that PE may imply a two-stage process, with an initial mental translation and then the comparison and correction of the MT suggestion. This would create a feeling of detachment, or "distance" from the authorship of the translation, in a way that he related with the revision and classified as a useful approach, taking into account what is at stake in such a process.

Evaluating the technological environment

In the questionnaire that was open to answers after the initial presentation in the workshop, a few questions concerned the contents of the presentation (these questions are identified here as Q.2.A). Some of the slides in the presentation described the future of translation, in the context of the technological evolution caused by MT, contrasting dark and bright perspectives. Question Q2.A.3 asked what participants thought about the impact this technological environment might have on the profession.

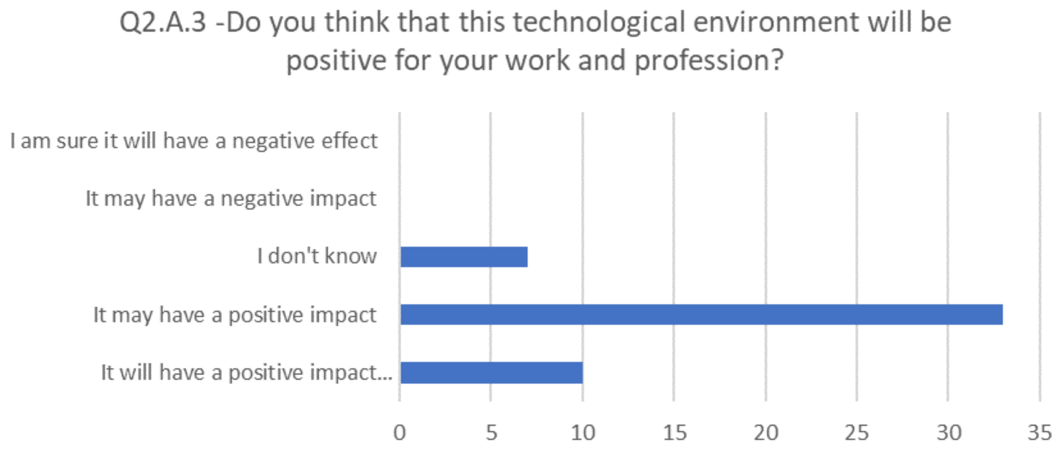


Figure 28 – The impact of the technological environment (Q2.A.3).

This chart reveals that the vast majority considers that these technological advancements will not have a negative impact on the profession, to the point that there were no negative answers. But this view was somewhat balanced by some of the comments left in response to Q2.A.4, which allowed for free comments on the presentation. Two participants expressed concerns that this technological context was used by clients to request PE instead of translation services, even when the quality of the MT hypotheses did not guarantee increases in productivity, as a way to force a global reduction in prices and rates. Another participant showed a more positive view, which included not only increases in productivity produced by MT, but also the raising of awareness from clients about the complexity and higher reliability of HT.

6.2.4. Assessment of texts, editing modes and actions

Apart from these initial questions, most of questionnaire 2 was filled in after the practical experiments with HandyCAT. In this final part, the questions moved away from the participants' professional profiles and concerns, to focus on their impressions on the features that they had the opportunity to test with HandyCAT.

Assessment of texts

The type of text is one of the factors that is associated with the quality of the results of MT. So, questionnaire 2 started by asking whether the texts that had been selected were seen as appropriate for MT.

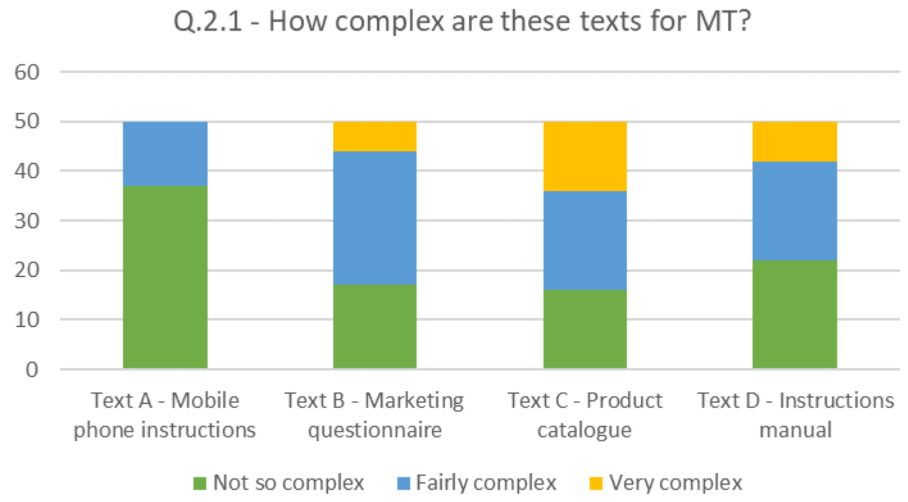


Figure 29 – The complexity of texts in the sample for MT (Q2.1).

Most participants considered that the texts were not very complex, especially text A, which was used for the familiarisation stage. Text C, an extract of a catalogue of office supplies, was the text most often classified as "very complex", which is justified because if it is composed of lists of features, with a very reduced number of full sentences. Still, this is the text with most repetitions, in terms of TM matches, and also seemed to be the one with the highest number of repeated phrases and expressions. This classification of texts based on users' impressions was complemented with the evaluation of the quality of the MT hypotheses created by MateCAT for each text.

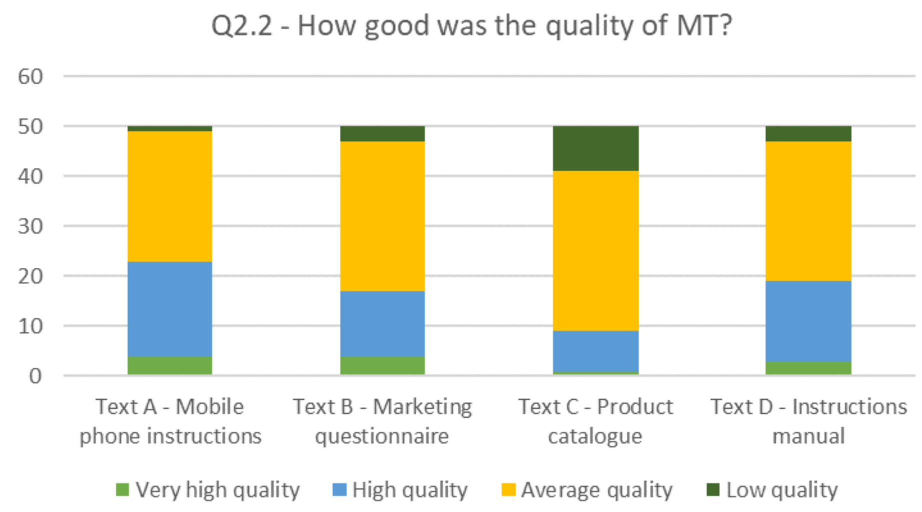


Figure 30 – The quality of MT hypotheses per text (Q2.2).

Most respondents considered that MT had an average level of quality, followed by those who considered that quality was high. Text C was considered as the one that presented most problems, which might confirm that this might be the text that presented more challenges to MT.

Comparing the two editing modes

The following charts sum up the answers to a question that gathered several criteria of comparison between the two editing modes enabled by HandyCAT.

The first two charts below show the preference for the AC mode (predictive writing), both in terms of speed and user-friendliness.

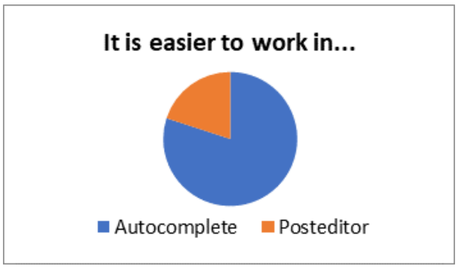


Figure 31 – AC vs. PE modes: speed and usability (Q2.3).

The next two options got more balanced replies, with participants preferring AC mode for translation and PE mode for post-editing.

Figure 32 – AC vs. PE modes: translation and PE (Q2.3).

The same division of votes appears in the next set of charts, which describe the suitability of each mode according to the extension of characters edited, with AC mode being considered as better for longer sequences and PE mode for shorter sequences.

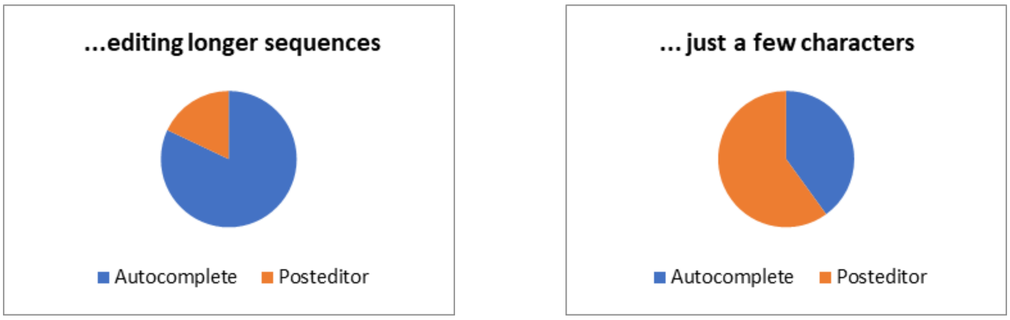


Figure 33 – AC vs. PE modes: short and long segments (Q2.3).

The next two charts reveal participants' evaluation of PE mode as a more intrusive method of work, but which allows translators to prepare editing work better.

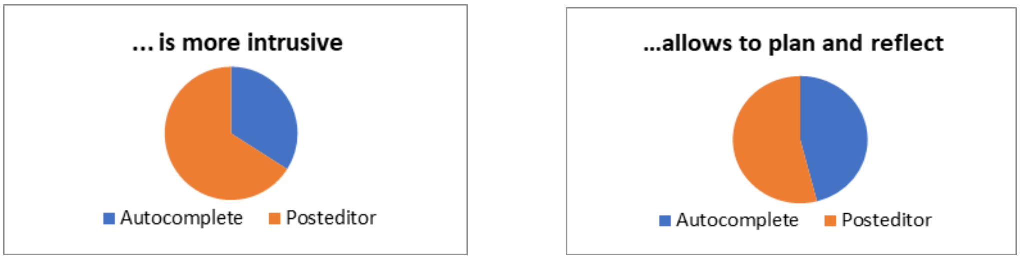


Figure 34 – AC vs. PE modes: intrusiveness and planning (Q2.3).

These results are in line with the expectations. PE mode was designed on purpose to make translators think before they edited, in order to devise more economical ways to edit the MT hypotheses. The trade-off in intrusiveness and some negative reactions was thus expected.

Evaluating the four editing actions

The next two questions involved the participants' perception on the actions that were performed during the four work sessions.

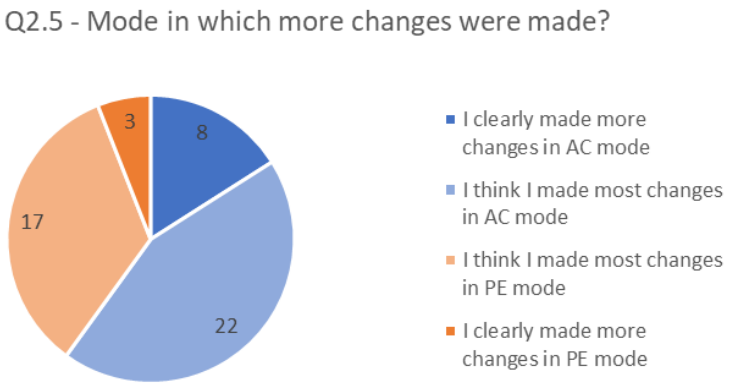


Figure 35 – Editing mode with most changes (Q2.5).

The perceptions of participants in relation to the compared number of changes between editing modes are not very clear, since most users chose one of the options which present some doubt (*I think I made...*) in both modes. However, the numbers of participants that associate the AC mode with more changes are clearly superior to those that consider that they made more changes in PE mode. These answers seem to confirm the expectation that there is some connection between extent of editing and work method used. However, these results are not enough to draw any conclusions.

In the next question, the focus was on the action that was most performed, in both editing modes.

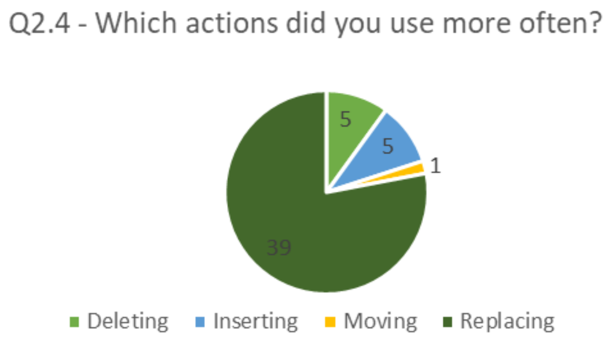


Figure 36 – Actions most performed in both editing modes (Q2.4).

For the large majority of participants, replacing was the action most used. This challenges the simplified views of PE as composed essentially of deleting and inserting. This result is consistent with the results of other studies reported in chapter 4. Replacing seems to be the most visible action again in the next question.

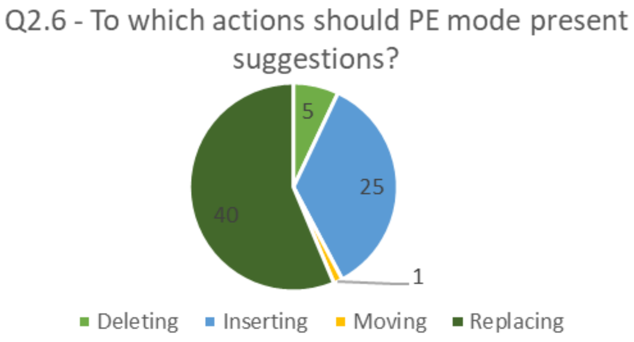


Figure 37 – Actions that should receive suggestions in PE mode (Q2.6).

Virtually all users considered that replacing should be supported by suggestions from the learning system. Only half considered that inserting also required some assistance, but the number of participants that considered that deleting and moving required suggestions was residual. This may be linked to the implementation difficulties of these two features, as commented above in section 5.3.2. In fact, deleting and moving words and phrases are usually very simple actions to perform in word processors and CAT tools, using cut and paste or drag and drop features, but in HandyCAT they became complex actions in comparison.

Assessment of usefulness

Nevertheless, more than on usability, these tests focus on usefulness, and the next section called the users' attention to that criterion of classification of the editing actions.

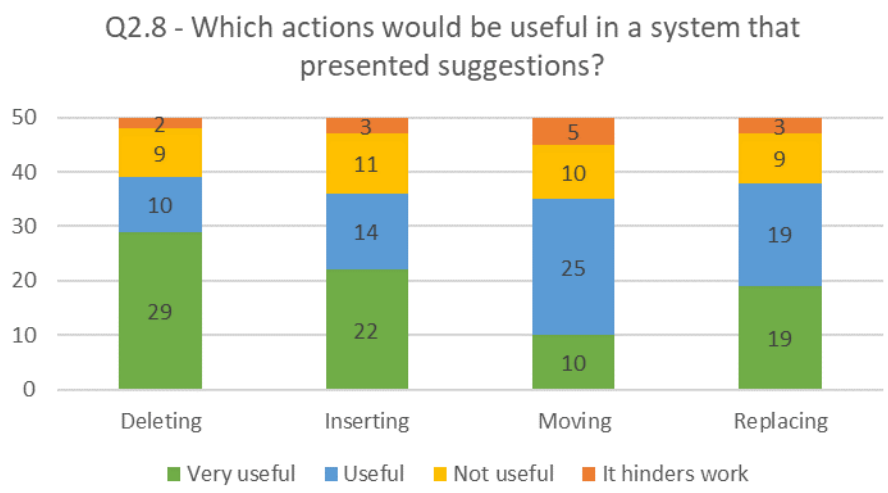


Figure 38 – Actions' usefulness in a system that presents suggestions (Q2.8).

When participants were asked to imagine a system that presented suggestions for each action, the classification of usefulness also encompassed the actions that were considered not to require support in the previous answer. Deleting, for example, was the action that received most votes as "very useful", with replacing occupying the second position. Moving remains the action that seems to be struggling to prove its worth, as 15 participants classified it as "not useful" or "it hinders work". Still, even for this, the number of participants classifying editing action support as "very useful" or "useful" is above 70%.

When asked to complete the sentence *"Post-editor mode may be very useful if…",* participants presented varied open answers. The answers were classified according to the central concepts that were being tested, and grouped around those concepts. The graph below describes the results of this organisation of the answers.

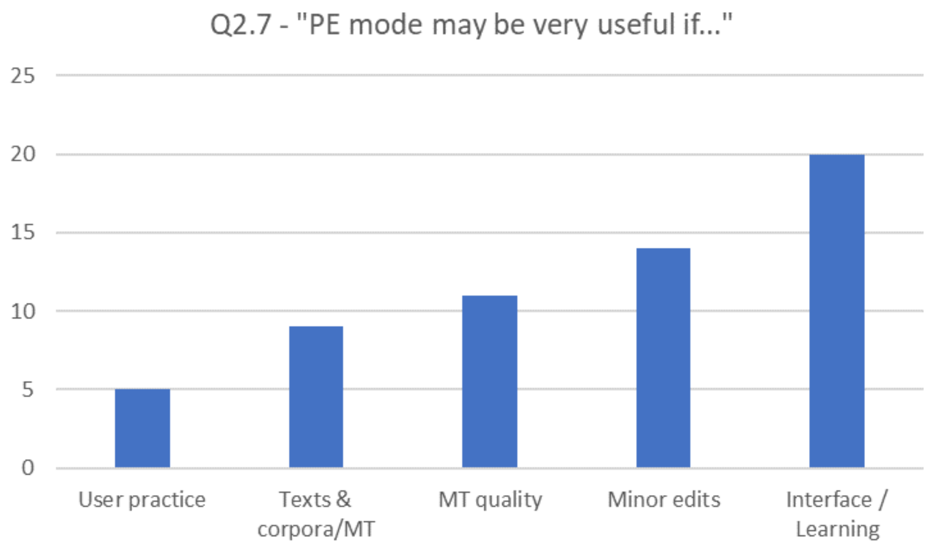


Figure 39 – Factors for the usefulness of PE mode (Q2.7).

A few participants related that there was a correlation between the usefulness of the tool and the translators' capacity to learn and practice using it. Others included in their responses factors associated with the type of ST, namely in that it should be composed of short sentences and be technical for PE mode to be useful, and also with the contents of the corpora on which the TM and MT were built. The third type of factor most frequently mentioned was related to the quality of the MT hypotheses, this factor being several times associated with the next factor, the short extension of the required edits: the better the quality of the MT hypotheses, the less editing would be required, and the more useful PE mode would be. When commenting on the type of minor edits required, most participants mentioned replacing a few characters in words. But nearly half of the participants reported the need to improve the interface and the usability of the software to make it more useful. Several answers mentioned the need for the tool to learn and reapply the same action (such as deleting often repeated words and correcting capitals). Other users made suggestions related to the selection of tokens, using keyboard actions instead of mouse clicks, and making this more flexible, by allowing for the selection and editing of parts of words, or offering a drag and drop interface element for moving.

A different question asked which features, of the ones presented at the workshop or any others, should be present in a PE tool. The replies from participants were varied.

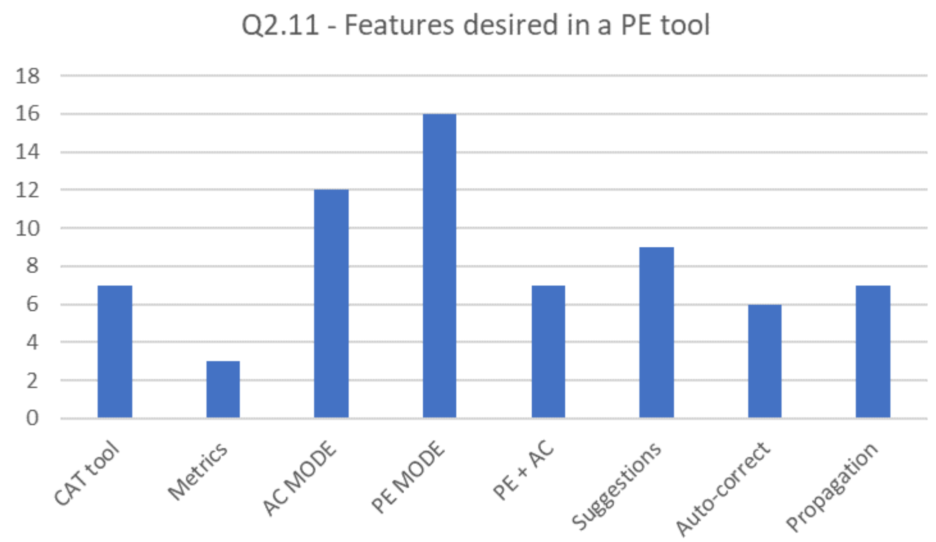


Figure 40 – Features desired in a PE tool (Q2.11).

Several answers included references to regular CAT tool features, such as terminology, grammar or QA checking, but other participants reported interest in functionalities to measure the changes made and time spent during editing. A fair number of responses referred directly to predictive writing, used by AC mode, and to the four editing actions of PE mode, as features that are desirable in a PE tool. Furthermore, several participants defended the blending of features from both modes, namely by adding an autocomplete capability to the insert action, or as a correction feature for the replacing action. The terms "(auto)suggestions", "(auto)corrections" were also used several times to describe support features that provide the user with aids for writing. Finally, users stressed several times that regular edits should be propagated, in view of their repetitiveness.

To finish this section, question Q2.10 raised the subject of the interest in using the four editing actions in a pedagogical context, by asking whether the user agreed with the sentence: *"The four editing actions are an interesting form to learn or teach post-editing."* The majority of participants considered that the four editing actions, as they were presented at the workshop, have a high interest for learning or teaching how to post-edit. No participants chose the option that showed that they disagreed with the pedagogical interest of a tool with the 4 editing actions. Only 8% were inclined to disagree with the sentence, the same number that did not have an opinion on this. So, only a total 16% showed some hesitation in supporting the claim that the 4 editing actions were useful as pedagogical instruments.

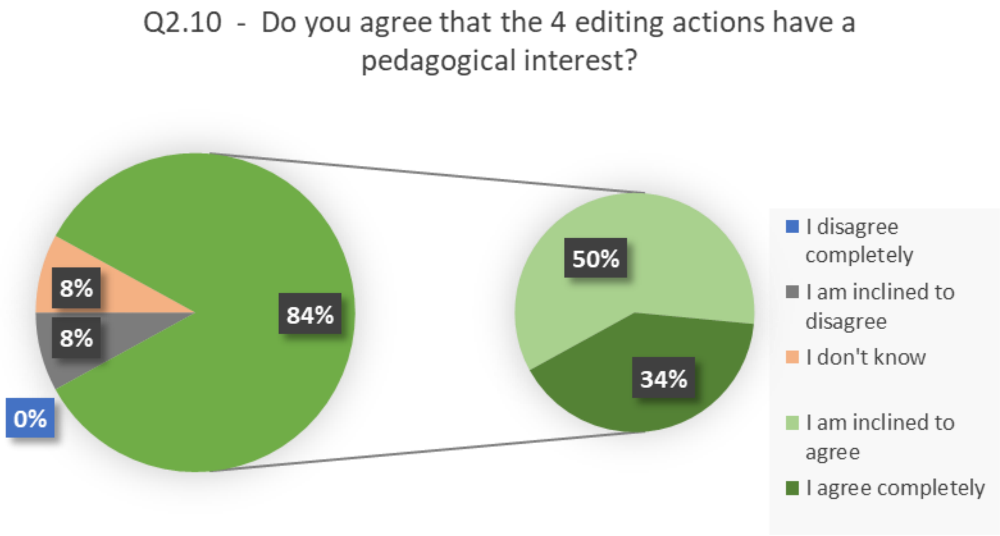


Figure 41 – Pedagogical interest of the four editing actions (Q2.10).

6.2.5. Assessment of the workshop

Both questionnaires included questions aimed at assessing the data collection through questionnaires, the workshop and the theoretical background that assisted it.

Questionnaire 1 ended with Q1.17 asking for feedback on this data collection element. Some of the respondents took the opportunity to add their impressions on MT and PE, namely in terms of how much the evaluation of their usefulness depends on the quality of the results. One user referred to a very common problem when working with MT into Portuguese, which is the fact that most of the data in MT engines has been collected from sources with no variant discrimination. This means that most of the data will be in Brazilian Portuguese, which increases a lot the editing that is necessary when applied to European Portuguese projects.

The first question asking for feedback in Questionnaire 2 focused on the proposal of a redefinition of PE, presented above, in section 6.1.4. Question Q.2.A.1 was: “*Do you think that the redefinition of post-editing is clear and useful?*”

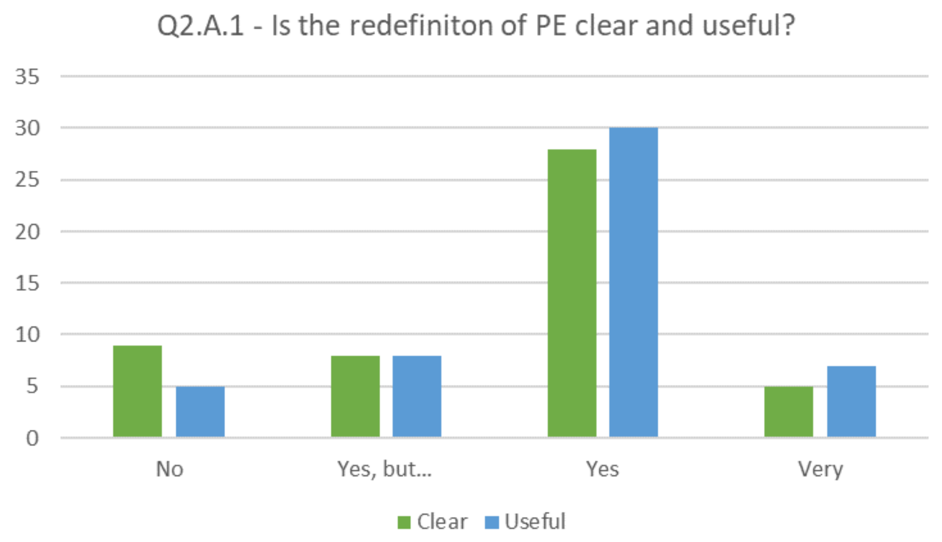


Figure 42 – Clarity and usefulness of the redefinition of PE (Q.2.A.1)

This was an open question, which participants could reply to in any way they wanted, so the chart above is a summary of the results of a coding process. A large majority of the results are positive, but participants considered the redefinition more often useful than clear. The replies to this question were complemented by comments that users made on question Q.2.A.4. This feedback will be discussed below, after the presentation of the results of the next question, which was dedicated to the proposal that 25% editing effort could be defined as a threshold between editing and translating.

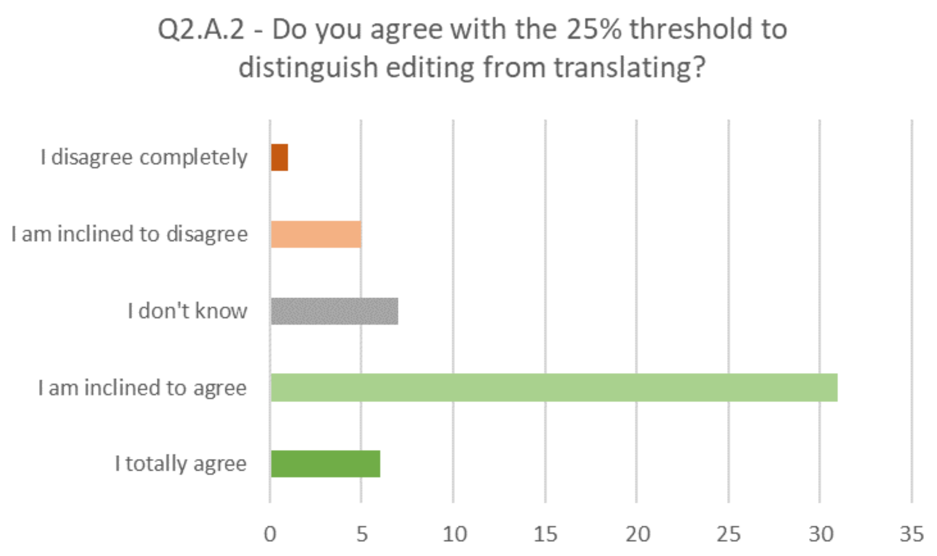


Figure 43 – Opinions on the 25% threshold (Q.2.A.2).

As Figure 43 illustrates, most participants were receptive to this threshold, even if only a few chose to affirm that they "totally agreed" with this proposal. Since this admittedly requires further testing, this type of response was not surprising.

One of the positive comments received on 2.A.1 highlighted the possibility that such a threshold established a metric to distinguish between translating and editing as a positive contribution for the profession. However, another participant claimed that this quantitative threshold was not enough for this distinction. Several respondents stressed that beyond mere quantification of editing actions, some measurement of effort, time and the type of changes made was required. Other participants related this to the external negotiation, in terms of discussions on intended quality levels and budgeting (foreseen costs). One answer referred to the current lack of clarity in distinguishing between light and full PE, and several others mentioned that this definition does not clarify the distinction between PE and revision.

At the end of the workshop, after the practical experiments with the two editing modes in HandyCAT, feedback was again requested from participants. The first question was whether this experience had somehow changed their perspective on PE. This was again an open question, which was coded as described in the chart below.

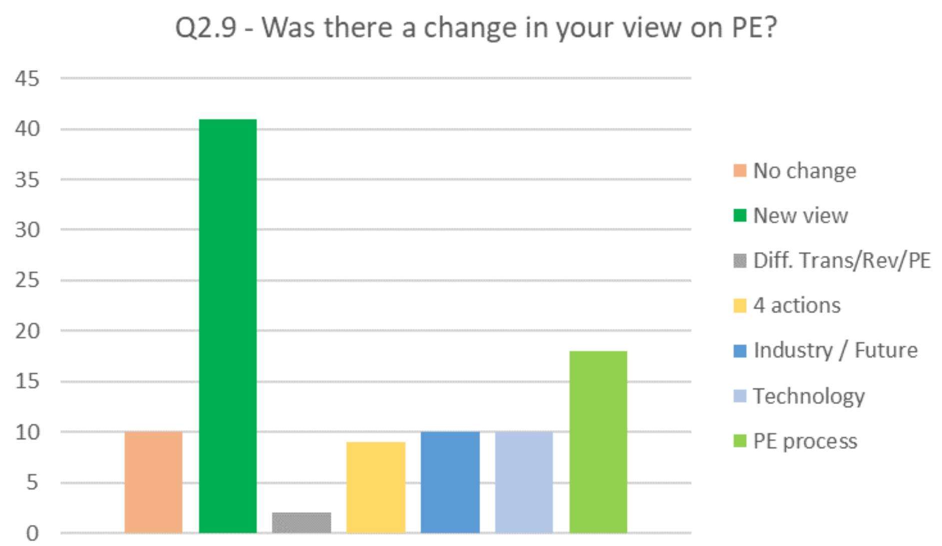


Figure 44 – Changes in perspectives on PE (Q.2.9).

A fifth of the participants considered that their perspective on PE did not change with the workshop (first column above), although most did not clarify this position. On the opposite side, the majority of the participants (41, to be exact) mentioned different ways in which the workshop had changed their perspective, namely because it had given them new information on PE or by causing some new reflections on this process (second column). The next columns present the details of all the answers. Some of the answers clearly mention a change in perspective, while most of the others use terms that reveal this, whether by just referring to the new information received, by mentioning that the participant had become more conscious of certain processes, that he reflected on new details, or that he realised that PE has certain characteristics which were not evident before. Only two answers referred to effects on the comparisons between translation, revision and PE (third column), and only nine mentioned the effects caused by using the 4 editing actions (fourth column). A fair number of participants (10) reported changes to the way they think about the future and the industrial context in which they work (columns five and six). A high number of responses (18) included items related to the description of the PE process. Some answers referred directly to the "decision-making process"; others mentioned that the workshop made them see the process as more "cerebral", and others as more "mechanical". In fact, the connection between human and machine supports was present in several answers, in comments on how tools may help organise the "chaos" of editing decisions to make before an MT hypothesis, and as aids in a process that is more complex than initially considered. A few participants mentioned that the methods used at the workshop helped them reflect and decide on the need to make certain changes/edits or whether these were superfluous. So, it seems that a tool that forces and brings to light unconscious processes may prove its usefulness in a context in which the threshold of necessary/unnecessary decisions is a fundamental requirement.

Although it was made clear at the presentation and the rest of the workshop that HandyCAT was not a fully-developed tool, and that it was to be expected that usability problems could affect work during the experiments, an evaluation of the tool was still requested at the end of the workshop. As may be seen below, in spite of all the problems users faced, there were only six negative classifications of the tool, in contrast to nearly half the users giving it a clear positive evaluation.

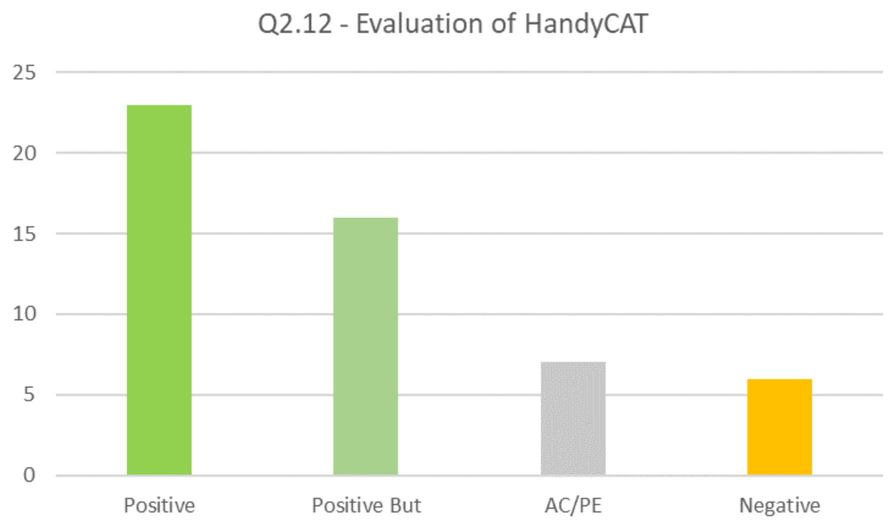


Figure 45 – Evaluation of HandyCAT (Q.2.12).

A few users commented on the differences between AC and PE mode (generally, evaluating AC mode more positively), and most users considered that the initial impression of a difficult tool to work with disappeared with practice. Several participants, namely some of those with negative impressions, highlighted the fact that this was a tool aimed at being used in research, and that it was not ready for professional use.

The final question in Questionnaire 2 was a request for a global evaluation of the workshop. All users gave a very positive feedback, with some admitting that they would like to see the concepts that were presented and the tool that was tested further developed. A few also commented on the need to expand the time available for this type of event, so as to allow for a better learning and reflection experience.

6.3. Presentation of results from activity logs

The details of the activity produced during the workshop have been recorded by HandyCAT in activity logs (see sample in Appendix 6). This extensive volume of data was later converted into Microsoft Excel for analysis. The amount of data collected and produced at this stage is such that it is impossible to present everything in printed form. In the next section, all this data is presented in charts, tables and other visual summarisation instruments, as considered appropriate for each detail studied.

This section describes the processing of this data and the results achieved. The methods applied to the results are also explained and discussed in this section. In the description of results, five main variables are analysed: texts, segments, users, the two editing modes and the four editing actions in PE mode. For each variable, different analysis methods were chosen.

6.3.1. Preparation of data for analysis

Each time an action was performed by the user, a new event was recorded. For example, opening a segment was recorded in an event called "segment-change". This designation was later replaced by "open-segment" for clarification. (Some of the other names of events and variables were also replaced. The new names are used throughout this dissertation.)

In AC mode, only two events were registered for each segment: "open-segment" and "close-segment". In PE mode, a new event was recorded, not only each time a segment was opened and when it was closed, but also when the user chose each editing action, by selecting an option from the context menu.

For each event, the log included variables such as the name of the event, the work session it belonged to, the editing mode, the time stamp, the content of the target segment before editing, and the result after editing.

HandyCAT exported one activity log per user, with the sequence of all work sessions, in JSON format. Since the data analysis tools that were available did not handle this format, these logs were converted into tables, with each event transformed into a line (or a case) and the parameters of each event turned into columns, or variables.

The activity logs were converted to CSV format with the trial version of a format converting tool available at https://json-csv.com/. They were then processed by several Excel macros to clean each log and join all logs into a single table. Extra calculations were then added, such as the duration or processing time of each event, the length (number of words) of each MT hypothesis, the editing distance of each event, and the count of editing actions, as identified by Post-Edit Compare (PEC), a tool that will be described in section 6.3.3.

The variables extracted from the activity logs for each event are:

1) USER: The anonymised code for each user in the workshop
2) TEXT: The name of each of the 3 different texts:
    a. Questionnaire (Text B);
    b. Catalogue (Text C);

c. Manual (text D).

3) SEGMENT: The segment number (in each text);
4) MODE: One of two editing modes (AC or PE)
5) EVENT: The type of event, according to the editing mode:
   a. **AC mode**: two events per segment:
      i. "open-segment"
      ii. "close-segment".
   b. **PE mode**: several events per segment, depending on the user's actions, chosen in HandyCAT's contextual menus:
      i. "open-segment"
      ii. "PE.delete"
      iii. "PE.insert"
      iv. "PE.replace"
      v. "PE.move"
      vi. "close-segment";
6) SESSION: The identification of each work session:
   a. Session A: first text, 10 minutes in AC mode;
   b. Session B: second text, 15 minutes in PE mode;
   c. Session C: third text, 5 minutes in AC mode;
   d. Session D: third text, 5 minutes in PE mode.
7) SOURCE: The source segment (English);
8) REFERENCE: The MT hypothesis or the previous state of the segment (after each editing action);
9) RESULT The result of the event, be it the end of the editing sequence of a segment (in close-segment) or the result of the application of each editing action in PE mode;
10) TIME: The time stamp for each event.

For each line, additional information was added, either for organisation of the data in the table, or for its characterisation. Some of this additional data implied calculations, some of which were performed within Microsoft Excel, and others implied the use of external tools:

11) No.LINE: A sequence number to sort the lines;

12) START: A starting time for each event, to estimate its duration (see below);
13) DURATION: The duration of the event, as estimated in Excel, using a method that is described in the section 6.3.3;
14) Delete, Insert, Replace and Move: Four columns imported from PEC, which account for the estimation of the edits performed in each event. This process is described in section 6.3.3;
15) EDITS: The total number of edits per event estimated by PEC, and which is basically the sum of the 4 previous columns;
16) LENGTH: The number of words of the Reference, as estimated by PEC;
17) TER: The Translation Edit Rate described in section 3.5.1, which results from the estimates of the distance between the Reference and the Result performed by PEC.

In this table, each case/line is characterised by 20 variables. Each case is a specific combination of five variables, which composes a single tuple. The five main variables that define each tuple are: User, Text, Segment, Mode and Event. All other variables in the table (like Duration, Edits, and others) are features of these tuples.

After the first cleaning stage, in which all data from the familiarisation sessions was deleted, the full table contained 8916 cases. After the cleaning stages described in the next section (6.3.2), the full table contained 8565 cases.

It is important to note again that the focus of the analysis of the logs does not involve any linguistic analysis. The purpose of the analysis of edit or performance rates (such as TER and Speed), for example, is to see how these metrics relate to the other technical features, such as alignment of words and phrases, the editing actions applied to sentences and the use of the tools. There was no analysis on the recurring nature of the contents being edited, or on the type of edits in terms of content (spelling, lexical replacement, syntactic correction, or any other). There should neither be inferences related with the decision processes, based on editing time or any other parameter, such as the complexity of the contents being edited, or concerns about the extent of editing, at least not in terms of obligatory vs. optional language features, or editing styles.

6.3.2. Improving the reliability of the data

The data collected from the activity logs suffered from several problems, which were solved by different strategies.

Differences in allocation of time to texts and modes

For different technical reasons (like the need to replace one of the servers in which HandyCAT was running), the distribution of users per groups did not go according to the plan, and it was not possible to achieve an equal number of texts and users per session. The intended distribution implied that there should be 17 users per session, for a total of 51 participants. This distribution guaranteed that the same time was allocated for each text (33% of the total for each), and that the intended proportion of extra time for the PE mode was kept until the end (43% of the time for AC mode and 57% for PE mode). At the end of the work sessions, these were the numbers for the distribution of the 49 users that participated, in the activity logs of the work sessions.

	Nº participants	Session	Time (min)	Text	Mode
Group 1	17	Session A	10:00	Text B – Questionnaire	AC
	17	Session B	15:00	Text C – Catalogue	PE
	17	Session C	05:00	Text D – Manual	AC
	18	Session D	05:00	Text D – Manual	PE
Group 2	16	Session A	10:00	Text C – Catalogue	AC
	17	Session B	15:00	Text D – Manual	PE
	18	Session C	05:00	Text B - Questionnaire	AC
	17	Session D	05:00	Text B – Questionnaire	PE
Group 3	10	Session A	10:00	Text D – Manual	AC
	10	Session B	15:00	Text B – Questionnaire	PE
	13	Session C	05:00	Text C – Catalogue	AC
	13	Session D	05:00	Text C – Catalogue	PE

Table 17 – Distribution of participants in sessions, texts and modes.

Group 3 was the one most affected by problems, which led to several logs being missed. Moreover, a few users swapped groups during the work sessions, which also contributed to this result. This could have caused an ample distortion in the data, which might invalidate the analyses of the different variables. Or, since users edited the texts in differently timed sessions, this effect could be more apparent than real. The table and the chart below describe the actual total numbers of users and times in which the three texts were edited, in all sessions in the two editing modes.

	AC		PE		
	User sessions	Time	User sessions	Time	Total time
B–Questionnaire	35	04:20:00	27	03:55:00	08:15:00
C–Catalogue	29	03:45:00	30	05:20:00	09:05:00
D–Manual	27	03:05:00	35	05:45:00	08:50:00
Total times	91	11:10:00	92	15:00:00	26:10:00

Table 18 – Total times and users allocated, per texts and modes.

One may see in the table above that the planned 29 hours in total of editing work that were planned had been reduced to little more than 26 hours, which is a loss of only ca. 10% of the total time. Besides, the total number of user sessions in each mode were also pretty balanced, with 91 sessions for AC mode and 92 for PE mode.

So, the problems in the distribution could affect two main variables: modes and texts, in terms of time allocated to each of these variables.

The analysis of the chart below shows that the intended ratio between the modes was not affected in global terms. As we may see below, in the total, AC mode was used in 43% of the total time, and PE mode in 57%, which were the actual values proposed. However, these totals were obtained from unbalanced sets of texts.

The equal proportion of times dedicated to each text was disrupted, but for a short margin. Instead of 33% of the time for each text, text C-Catalogue was edited in 35% per cent of the total time, and text B-Questionnaire in only 32%.

The major effect of the distribution is in the balance of the two modes within each text. The text most affected was text B-Questionnaire. The translators that worked with this text ended up spending more time in AC mode than in PE mode.

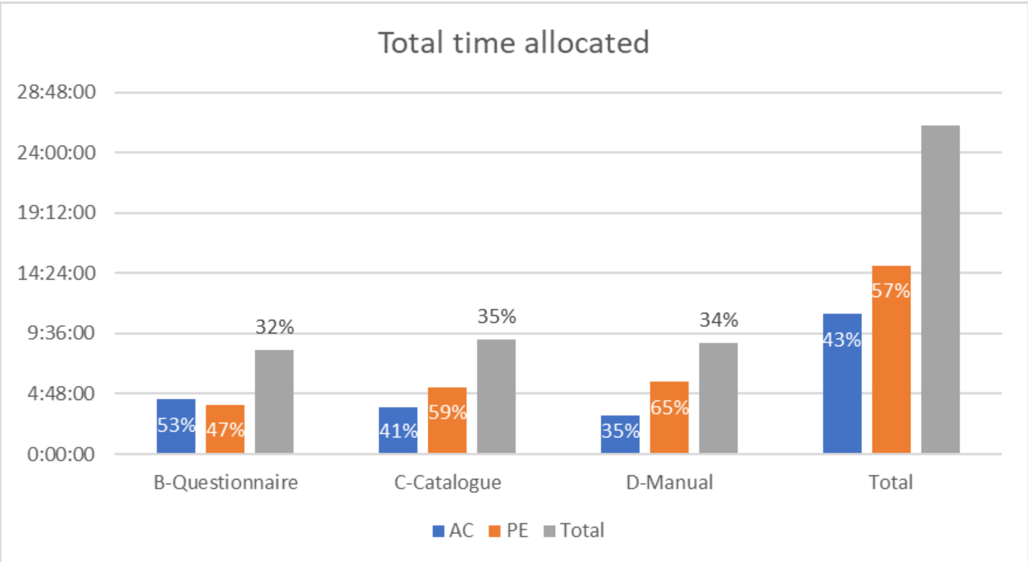


Figure 46 – Total time allocated for each text and mode.

The strategy to deal with this inconsistency was two-fold. In the first part of the analysis, presented in this section (6.3), each main variable is described separately. In this section, all calculations apply the absolute values, disregarding the variations identified, or indicating them when they may affect the data presented. This will be highlighted especially in the case of the analysis of modes per text. In the second part of the analysis (section 6.4), the instruments that were used were not sensitive to these distortions in the data.

Different processing of AC and PE data

Besides this factor of eventual distortion of the data, another feature of the data could have affected the consistency and reliability of the results: the fact that PE mode creates many more events than AC mode. From the total 8565 events, 40% are editing actions registered only in PE mode. This fact had to be taken into consideration in all calculations, especially those that involved summing up or averaging different values across variables, like duration and edit scores. So, the methods for data analysis were adapted to avoid this affecting the results.

Data that had to be cleaned

In several stages of analysis, it was apparent that there were discrepancies and inconsistencies that had not been anticipated. When this noise was identified as having been caused by some of the work methods, these were adjusted. This section describes some of the stages in which it was decided to clean the data, in order to achieve more consistency, while not jeopardising reliability.

The initial analyses of duration showed considerable discrepancy between several users in terms of this variable. A closer look at open-segment and close-segment events provided a good insight into this and provided appropriate solutions. The decisions to clean the tables were based on the different impacts and the uses of the different cases and variables.

After a brief analysis of duration, it was soon evident that for one of the users there were only records of a very short work session (a 5-minute session in AC mode). Since this did not allow any comparisons in terms of texts, modes, or any other variable, it was decided to remove this user from the table.

For other users (mainly from group 3) there were only records of the two last 5-minute sessions C and D, in which the same text was edited using the two editing modes. Besides, the records of the first session (10 minutes in AC mode) of one of the users were also missing. Since a strategy had already been established to deal with global analysis for text and mode level, the data concerning these users was kept in the table.

Another effect that was visible was that several users had overrun the total duration proposed for each work session. However, the variations observed were not very big, and most of the analyses would not affect total productivity. There was, however, an outlying case that requested investigation: a user had recorded 20 minutes of work in a 10-minute work session. After checking the whole log, it was concluded that, instead of using Text A in the familiarisation stage, he had used the same text that he later used for the first session in PE mode. Thus, both these sessions were being summed up in one single production session. It was decided to remove the data concerning the familiarisation session of this user from the table.

Most open-segment events are very short, and most last no time at all. In AC mode, for example, they only record the status of the segment when it is open – everything that was done during the editing is recorded in the close-segment event, including the time lapse since the segment was open. This might mean that most of the calculations might be done excluding open-segment events, which represented 33% of all events. However, the assumption is not always correct, since, in some cases, these events even have a longer duration than close-segments. By studying these cases, it became apparent that some users tended to have several segments open at the same time and only closed and confirmed them after a long time. So, open-segments included all the editing time, and close-segments were instantaneous. In one such case, if the duration of confirm-segment events was the only time considered in one of the 5-minute sessions, a mere 2 seconds would had been studied. Because of these issues, the process by which duration was estimated was improved, by making sure that all segments were clearly contained between an open-segment and a close-segment, and that any calculation of duration might grant the inclusion of everything that happened between those two events. This process is described in the next section.

Two types of segments were deleted from all activity logs. Some users had to reset the editing sequence of a segment. When they deleted the full segment, and then closed it without saving it, HandyCAT replaced the reference text with the text *“Modify this text to translate.”* A few users closed and saved segments with this text. Since these segments did not include properly edited text and they would add noise to the edit rates, they were deleted from the tables. But since they were in the middle of a sequence of edited segments, and duration is estimated by the differences from past events, the time spent doing this was not lost from the records, as it was added to the duration of the following event. Another type of segment that was deleted was segments that had not been closed. Most of these were very short ones that had been opened right at the end of each session, and left unedited. But, in some cases, a few lines of editing action events also had to be deleted. The reason for this is that it would not be possible to compare the edit scores of these segments without a close-segment line. So, in these cases, although the user's session may lose some information on duration and editing, the global reliability of the data is improved.

Repeatedly edited segments

Another effect that had to be avoided in the final stages of cleaning the table was the repetition of "close-segment" events in the same segment. This was caused by users that reopened a segment, sometimes leaving another one unfinished, to come back and re-edit it later. This was detected because a comparison showed that the number of close-segment events was higher than that of edited segments. If the result was the opposite (more segments edited than closed), that would have meant that the cleaning had not been completed and there were still segments opened but left unclosed. But this result meant that some users had closed the same segment more than once. Since closing a segment in a CAT tool is called "confirming", the decision was made to call "confirm-segment" only to the last event in each segment, and to keep all intermediate events of closing a segment simply as "close-segment".

This differentiation between a close-segment event (which means that the segment was reopened later) and a confirm-segment event (which means that it is in its final form), guaranteed the proportionality of the analyses that depended on confirmed lines, such as duration totals and edit scores. Besides, it provided a method for the analysis of this action of reopening a segment and re-editing it. After this, a check was made to make sure that each segment had only one confirmation event.

Calculating averages in AC mode and PE mode

The results obtained by the use of tools that efficiently perform many calculations also needs to be carefully analysed, to avoid unreliable conclusions. Microsoft Excel calculates totals and averages by selecting the same level of cases. For example, total duration of a segment is calculated by adding all events in that segment. Averages are then calculated by dividing the total duration by the total number of events in that segment. Although perfectly reasonable, this created a surprising result, in which all segments edited in PE mode had much lower average durations than those edited in AC mode. The reason for this was the fact that there is a much higher number of events in PE mode segments, due to each editing action being a new event. In one such case – segment 2 of the text known as Questionnaire – AC mode editing had produced 76 events in total for all users, while PE had produced 534 (!) events. So, although total durations are reliable, average durations cannot be calculated directly, and they have to be estimated in two different stages.

Inconsistencies that affect TER calculations

The next section describes how duration and TER were calculated for each event. This section explains an adjustment of the data that was necessary to guarantee that TER estimation was consistent and comparable.

For comparison of editing scores between both editing modes, only close-segment events are used. Usually, a direct comparison between close-segment events in the two modes is possible, because these segments show the initial reference sentence (the MT hypothesis) against the final edited version, in both modes.

However, due to the way PE mode behaves in intermediate events, and to the fact that some segments were reopened by users, this is not always the case.

In AC mode, no editing information is registered in other events apart from close-segment events. So, we know we have the comparison between the initial MT hypothesis and the final edited version as the total editing distance for all segments in this mode. In PE mode, one needs to dig a little deeper to understand why some editing scores were not comparable. The symptom that revealed this was the fact that the same segment appeared with varied lengths. For example, a segment might be considered as having 4 words in one case and in another the same segment could appear as having 12 words. As length is the base value for the editing scores, these values were distorted.

A detailed analysis showed that whenever a segment was reopened, the reference was no longer the initial MT hypothesis, but the status of the segment the last time it had been edited.

After having reclassified closed segments that were later reopened as confirm-segment events, the references of these segments had to be replaced by the initial MT hypotheses, and then all edit scores had to be recalculated. This guaranteed that all close-segments had the same length, allowing for a direct comparison between editing scores of both modes for all users and all texts.

6.3.3. Estimating duration and edit scores

Since each event in HandyCAT logs a timestamp, in order to measure duration, it had to be decided whether this marked the beginning or the end of each event. By choosing this as the stamp of the end of the event, duration was measured as the lapse of time from the previous timestamp to the timestamp of the current event. This considers that all pauses between one timestamp and the next belong to the duration of the second event.

A method was devised to control the fair distribution of time, even when several segments were open at the same time. To restrain any eventual errors in this process, it was decided to make sure that at the beginning of each user's session time was restarted, so that any calculation from previous lines did not interfere with the calculation within each session. By doing this, times recorded from the previous sessions and any pauses before opening the first segment did not affect the calculation of duration of each session.

With the method applied, pauses that occur between the first event ("open-segment") and the last ("confirm-segment") are included in the durations of each segment. This also excludes the time attributed to other segments that could be edited at the same time. Besides, the time between each close-segment and the next open-segment, instead of being separated as a pause between segments, is included as the duration of the open-segment event. So, all the time from the being to the end of each session is included as duration of the session.

It is important to emphasise at this point that there was no way to segment pauses from action events. Furthermore, if pauses had been discriminated, there would be no additional data to interpret them, in terms of, for example, if a segment with a long duration was associated with a long time typing or thinking. This is a shortcoming of this type of data collection, but it was not considered relevant, since this study does not analyse cognitive effort and its manifestation through pauses.

In AC mode, most segments only have two events. In PE mode, the total duration of a sequence of editing a segment includes the sum of the durations of all editing action events between the opening and closing events. This means that the number of lines accounted for is not regular in this editing mode. The processing described above and the cleaning stage described in the previous section assured that all calculations could be done accurately or in good conditions of comparability.

The estimation of the edit scores was more complex. The metric that seems to be most adjusted to the purposes of this study is TER, since it presents the results in terms of the four editing actions that have been proposed as the defining elements of editing. After several tests with different ways of measuring the edit distance between the MT hypotheses and the results of the PE, and since it was not possible to incorporate such a measurement from the outset of the experiments within HandyCAT, the chosen strategy was to analyse the data available in the HandyCAT logs with the help of an external tool. The chosen application provided a very good insight into the mechanics of this metric.

Measuring TER with Post-Edit Compare

The tool used to measure TER was Post-Edit Compare (PEC), an add-in to SDL Trados Studio 2015. Documentation about this tool is only available online (Filkin, 2013b; Hartnett & SDL Community, 2014). This application provides several reports that are very rich, not only in terms of content, but also in presentation formats: there are XML reports which register each word and character modification, charts for all metrics at text and project level, track change visualisations for each segment, and also a visual description of the edits. This level of detail made this tool adjusted to the purposes of this dissertation.

PEC measures the number of words in the reference segment, estimates the number of edits that transforms the reference into the edited version, and presents the values in different reports. One of the parameters is the total number of edits in the segment and the other is TER, which is the division of the number of edits per length. These two parameters will be referred together as “edit scores”. The “No. of edits” parameter gives a more detailed insight into what translators did, and TER weighs the number of edits against the length of each sentence.

PEC analyses two versions of a document and estimates the edit distance between them, segment by segment. This comparison uses two different methods:

Post-Edit Modification percentage (PEM%) – this is a method that follows the Damerau-Levenshtein distance calculation, by counting the minimum number of edits (delete, insert, substitution of a character and transposition of two adjacent characters) necessary to modify the reference of a segment into its edited version, and then estimates the weight of these changes as a percentage of the length (number of characters) of the reference;

Translation Edit Rate plus (TERp) – this method seems to follow closely usual implementations of TER, but it is not documented how it is calculated. However, the reports provided by the tool allow a detailed understanding of how this metric was implemented. This method identifies the number of deletions, insertions, replacements and movements of words performed on the reference, and then divides the sum of these actions per the number of words of the reference.

PEM% was considered inappropriate for this study, mainly because this method is character-based. The effects of such a method will be discussed below. From this point onwards, the analysis focuses on the effects of using TER.

As described in section 3.5.1, TERp is different from TER because of the possibility for replacements to be identified and weighted not just by full lexical similarity, but also by synonymy, stemming and paraphrases. In a full TERp implementation, a replacement by a synonym does not have the same cost as a replacement by a different word. However, this feature depends on a dictionary of word forms and relations, and these are usually only available for English. Besides, TERp has a limit at 100% (or 1.0), which TER does not. Finally, and more importantly, TERp supposes a hill-climbing search algorithm, which adjusts the weights of each edit, so to better correlate with human quality assessment.

None of these features was implemented in PEC. So, it is misleading to call this feature "TERp". In this dissertation this will always be called TER. However, in some of the figures that illustrate PEC's reports, the name TERp may also appear.

Understanding TER in PEC

The implementation of TER in PEC, although an add-in to a CAT tool, SDL Trados Studio, does not record the actual actions. Instead, it does a posterior estimate to measure the edit distance. Before applying this edit metric to the 3 texts used in the workshop, the metric was tested with Text A (used at the familiarisation stage).

The purpose of this analysis was not only to check whether TER could identify the actions performed by users and the text units they worked with, but also to improve the capacity to interpret the results from the metric.

A sample of one of the most detailed reports presented by PEC may be seen in this dissertation as Appendix 8.

The terminology used in PEC needs some explanation. In this report, the first version of the sentence being compared is called “Original Reference”. In the dissertation, this is simply the “Reference”, and it includes the “MT Hypothesis” and the initial state for each editing action. The second version of the sentence being compared is called in this report “Original Hypothesis”. In this dissertation, this is called the “Result”. The two terms used in the PEC report make sense if one considers that this terminology is grounded in MT Evaluation and edit distances. The “original reference” was the gold standard and the “Hypothesis” was the MT segment that was evaluated against that standard. But the terminology in PEC changes according to the report one looks at, and the report that may be seen in Appendix 9, one may find the term “error” instead of “edit”.

In this report, one may see that the application compares the two versions of a segment, word by word. It aligns all words, and logs an edit whenever that alignment is not perfect. The identification of each action or edit is made as follows:

1. Whenever the application does not a find a word in the same position in the result (all other words being the same), it marks this as a “deletion” (see below):

Segment ID	[hyp][[Text A - Source text for MT.txt][0001]][][] (back to top)											
Number of Edits	1.0											
Number of Words	11.0											
TERp Score	0.09090909090909091											
Original Reference	Você **pode** usar o teclado para digitar caracteres alfanuméricos e símbolos.											
Original Hypothesis	Você usar o teclado para digitar caracteres alfanuméricos e símbolos.											
Alignment	**Reference**	você	pode	usar	o	teclado	para	digitar	caracteres	alfanuméricos	e	símbolos.
			D									
	Hyp After Shifts	você		usar	o	teclado	para	digitar	caracteres	alfanuméricos	e	símbolos.
Shifts												
Phrase Substitutions												

Figure 47 – A report from PEC for a deleted word.

2. When there is a new word in an "empty" position in the first version (all other words being the same), it considers that there was an "insertion":

Segment ID	[hyp][[Text A - Source text for MT.txt][0002]][][] (back to top)	
Number of Edits	1.0	
Number of Words	16.0	
TERp Score	0.0625	
Original Reference	Por exemplo, você pode adicionar entradas para os contatos [...]	
Original Hypothesis	Por exemplo, você pode **ainda** adicionar entradas para os contatos [...]	
Alignment	**Reference**	por exemplo, você pode adicionar entradas para os contatos,
		I
	Hyp After Shifts	por exemplo, você pode **ainda** adicionar entradas para os contatos,
Shifts		
Phrase Substitutions		

Figure 48 – A report from PEC for an inserted word.

3. When all words are the same except one, it considers that there was a "replacement" (segment 3);

Segment ID	[hyp][[Text A - Source text for MT.txt][0003]][][] (back to top)	
Number of Edits	1.0	
Number of Words	10.0	
TERp Score	0.1	
Original Reference	Você pode percorrer esta lista e **selecione** o caractere desejado.	
Original Hypothesis	Você pode percorrer esta lista e **seleccionar** o caractere desejado.	
Alignment	**Reference**	você pode percorrer esta lista e **selecione** o caractere desejado.
		S
	Hyp After Shifts	você pode percorrer esta lista e **seleccionar** o caractere desejado.
Shifts		
Phrase Substitutions		

Figure 49 – A report from PEC for a replaced word.

4. When all words are the same, but they do not occupy the same positions, the application tests movements of sequences of words into a new position, as a means to estimate the best path to reorder the segment. If any change in position of a word transforms the reference into the result, then it considers that there was a shift.

Segment ID	[hyp][[Text A - Source text for MT.txt][0004]][][] (back to top)									
Number of Edits	1.0									
Number of Words	8.0									
TERp Score	0.125									
Original Reference	Pare quando **você** vê um "h" na tela.									
Original Hypothesis	Pare **você** quando vê um "h" na tela.									
Alignment	**Reference**	pare	quando	você	vê	um	"h"	na	tela.	
	Hyp After Shifts	pare	quando	você	vê	um	"h"	na	tela.	
Shifts	**Start Pos**	**End Pos**	**Moved To**	**Cost**	**Hyp String**	**Ref String**				
Phrase Substitutions	1	1	2	1.0	você	você				

Figure 50 – A report from PEC for a moved word.

This simple description of one of the reports from PEC does not clarify how this tool measures the number of edits in sentences with more complex editing. In order to bring the effects of this metric to light, the decision was made to test PEC in a systematic way. For this test, text A was first translated by MateCAT and then edited by applying a sequence of increasingly complex editing actions. This simulation was done with no attention to the meaning and linguistic impact of the edits, but by following the logic of a step-by-step demonstration.

Step-by-step demonstration of edit scores for complex editing

The result of this analysis may be seen as Appendix 8 and 9. A table was created from extracts from the two reports and it is presented below. The columns under the title “Expected” show what was expected if the metric showed the actual actions performed. The columns under “Post-Edit Compare Report” compile the results from different reports from PEC.

In the first group of segments, only an edit action was performed on each segment. This is described in the first table below.

As can be seen in the table, for this group of segments, PEC identified all actions accurately. Beside an example of a deletion, an insertion, and a replacement in each segment (segments 1 to 3), movement was also tested, but in this case in different conditions and segments. The results of these experiments were not always accurate: the direction of the movement (forwards or backwards) and the number of positions the word was moved (one or two) affected the capacity of the tool to identify this action (segments 4 to 7).

The "track changes" column shows that, although all segments were correctly classified as having been edited by one action, in segment 5 the word that was moved (*pode* was moved one position back, from right to left) was not correctly identified. This is quite understandable: moving the word in position 4 to position 3 is exactly the same as moving the other word, which was in position 3, to position 4. This is an example of a movement forward that was estimated as a movement backward. However, if the purpose of the tool is to learn to validate alignments based on the edits, this type of error has an impact on the capacity to learn such an alignment. As we will see below, this effect is amplified when edits are more complex.

					EXPECTED								POST-EDIT COMPARE REPORT						
No. Seg.	REFERENCE	LENGTH	ACTIONS	EDITED	Delete	Insert	Replace	Move	EDITing actions	Edits / length	No. Words edited	Edited words / length	Delete	Insert	Subs	Shift	Track changes	Nº of errors	TERP
1	Você **pode** usar o teclado para digitar caracteres alfanuméricos e símbolos.	**11**	**Delete 1 word**	Você usar o teclado para digitar caracteres alfanuméricos e símbolos.	1	0	0	0	1	9.09%	1	9.09%	1	0	0	0	Você ~~pode~~ usar o teclado para digitar caracteres alfanuméricos e símbolos.	1	9.09%
2	Por exemplo, você pode adicionar entradas para os contatos, escrever mensagens ou agendar eventos no calendário.	**16**	**Insert 1 word**	Por exemplo, você pode **ainda** adicionar entradas para os contatos, escrever mensagens ou agendar eventos no calendário.	0	1	0	0	1	6.25%	1	6.25%	0	1	0	0	Por exemplo, você pode ainda adicionar entradas para os contatos, escrever mensagens ou agendar eventos no calendário.	1	6.25%
3	Você pode percorrer esta lista e **selecione** o caractere desejado.	**10**	**Replace 1 word**	Você pode percorrer esta lista e **seleccionar** o caractere desejado.	0	0	1	0	1	10.00%	1	10.00%	0	0	1	0	Você pode percorrer esta lista e ~~selecione~~seleccionar o caractere desejado.	1	10.00%
4	Pare **quando** você vê um "h" na tela.	**8**	**Move 1 word 1 position forward**	Pare você **quando** vê um "h" na tela.	0	0	0	1	1	12.50%	1	12.50%	0	0	0	1	Pare ~~quando~~ você quando vê um "h" na tela.	1	12.50%
5	Depois de inserir a primeira letra, você **pode** pressionar diretamente outra tecla (exceto Enviar) para inserir a próxima letra.	**19**	**Move 1 word 1 position back**	Depois de inserir a primeira letra, **pode** você pressionar diretamente outra tecla (exceto Enviar) para inserir a próxima letra.	0	0	0	1	1	5.26%	1	5.26%	0	0	0	1	Depois de inserir a primeira letra, ~~você~~ pode você pressionar diretamente outra tecla (exceto Enviar) para inserir a próxima letra.	1	5.26%
6	Se a próxima letra desejada estiver na **mesma** tecla que a actual, aguarde até que o cursor aparece à direita da letra atual e, em seguida, você pode digitar o próximo.	**31**	**Move 1 word 2 positions forward**	Se a próxima letra desejada estiver na tecla que **mesma** a actual, aguarde até que o cursor aparece à direita da letra atual e, em seguida, você pode digitar o próximo.	0	0	0	1	1	3.23%	1	3.23%	0	0	0	1	Se a próxima letra desejada estiver na ~~mesma~~ tecla que mesma a actual, aguarde até que o cursor aparece à direita da letra atual e, em seguida, você pode digitar o próximo.	1	3.23%
7	Você pode acessar as configurações de **tela** selecionando Configurações a partir do menu principal.	**14**	**Move 1 word 2 positions back**	Você pode acessar as **tela** configurações de selecionando Configurações a partir do menu principal.	0	0	0	1	1	7.14%	1	7.14%	0	0	0	1	Você pode acessar as tela configurações de ~~tela~~ selecionando Configurações a partir do menu principal.	1	7.14%

Figure 51 – Extract of edit scores report - single edit actions.

In the second group of segments (see below), these positive results were not repeated. In this case, the same single actions were performed, but on phrases of 2 or 3 words. Again, the alignment perspective is an important one, due to the learning purpose. If the user edits two words together, and the system does not identify this unit correctly, it may lose all the useful information – not only that the two words form a unit, but also which action is likely to be applied to those two words when they appear again. The number of words edited was also measured, in order to check whether the score was based not on the number of actions, but on the number of words edited. Segments 9 and 10 confirm that each word counts as an edit, but other factors seem to affect the capacity of the tool to identify the actual alignment, the editing action, and the number of words edited.

					EXPECTED								POST-EDIT COMPARE REPORT						
No. Seg.	REFERENCE	LENGTH	ACTIONS	EDITED	Delete	Insert	Replace	Move	EDITing actions	Edits / length	No. Words edited	Edited words / length	Delete	Insert	Subs	Shift	Track changes	Nº of errors	TERP
8	Definições do telefone - Pode alterar as definições **do telefone**.	**10**	**Delete 1 phrase (2 words)**	Definições do telefone - Pode alterar as definições.	1	0	0	0	1	10.00%	2	20.00%	2	0	1	0	Definições do telefone - Pode alterar as definições ~~do telefone~~.	3	30.00%
9	Data e Hora - Para definir o formato de data e **hora do sistema** do telefone.	**16**	**Delete 1 phrase (3 words)**	Data e Hora - Para definir o formato de data e do telefone.	1	0	0	0	1	6.25%	3	18.75%	3	0	0	0	Data e Hora - Para definir o formato de data e ~~hora do sistema~~ do telefone.	3	18.75%
10	Perfis - Para selecionar diferentes perfis para o telefone para atender diferentes situações ambientais.	**14**	**Insert 1 phrase (2 words)**	Perfis - Para selecionar diferentes perfis para o telefone **do vizinho** para atender diferentes situações ambientais.	0	1	0	0	1	7.14%	2	14.29%	0	2	0	0	Perfis - Para selecionar diferentes perfis para o telefone do vizinho para atender diferentes situações ambientais.	2	14.29%
11	Idioma do telefone - Para definir o idioma do telefone.	**10**	**Insert 1 phrase (3 words)**	Idioma do telefone - Para definir o idioma do telefone **como o seu**.	0	1	0	0	1	10.00%	3	30.00%	0	3	1	0	Idioma do telefone - Para definir o idioma do telefone como o seu.	4	40.00%
12	Esta opção é utilizada para definir o idioma **do telefone** como sendo o mesmo utilizado no cartão SIM.	**18**	**Replace 1 phrase (2 words)**	Esta opção é utilizada para definir o idioma **da chamada** como sendo o mesmo utilizado no cartão SIM.	0	0	1	0	1	5.56%	2	11.11%	0	0	2	0	Esta opção é utilizada para definir o idioma ~~do~~da ~~telefone~~chamada como sendo o mesmo utilizado no cartão SIM.	2	11.11%
13	Você também pode selecionar Configurações do telefone / perfis para definir os sons **para o telefone**.	**16**	**Replace 1 phrase (3 words)**	Você também pode selecionar Configurações do telefone / perfis para definir os sons **com a chamada**.	0	0	1	0	1	6.25%	3	18.75%	0	0	3	0	Você também pode selecionar Configurações do telefone / perfis para definir os sons ~~para~~com ~~o~~a ~~telefone~~chamada.	3	18.75%

Figure 52 – Extract of edit scores report - edits applied to phrases.

If we go back to the report in Appendix 8, we may see why segment 8 was not correctly identified as having been edited by deleting the two final words.

Segment ID	[hyp][[Text A - Source text for MT.txt][0008]][][] (back to top)				
Number of Edits	3.0				
Number of Words	10.0				
TERp Score	0.3				
Original Reference	Definições do telefone - Pode alterar as definições **do telefone**.				
Original Hypothesis	Definições do telefone - Pode alterar as definições.				
Alignment	**Reference**	definições do telefone - pode alterar as	**definições**	**do**	**telefone.**
			D	**D**	**S**
	Hyp After Shifts	definições do telefone - pode alterar as			**definições.**
Shifts					
Phrase Substitutions					

Figure 53 – Deletion of two words at the end of the sentence.

The application always tries to align the final words in a sentence. This comes from the Damerau edit distance: there must be an end of sequence marker to constrain the estimates (see sections 3.5.1). If the words in that ending position are not the same, it considers that there is a replacement. This invalidated the capacity to align the word in position 7 in the reference *(definições)* with itself, in position 9 in the "hypothesis". Then, it could have identified the actual deletion of the two final words in the reference. Besides this error in identifying the correct editing action, TER registered 3 edits, because it considers that all 3 words were edited, instead of only the 2 words that were deleted. This is a serious issue and it may be seen again in segment 11.

In segment 9, the deletion of three words was correctly identified, but the number of edits was considered as having been 3. To allow for the identification of those 3 words as a phrase, a single unit, the tool should have marked this as just one deletion. In segment 10, the insertion was identified correctly, but the number of edits shows again that the unit was not identified. This effect was also visible in segments 12 and 13, with the insertion and replace actions, but it did not happen with the move action, as will be described next.

					EXPECTED								POST-EDIT COMPARE REPORT						
No. Seg.	REFERENCE	LENGTH	ACTIONS	EDITED	Delete	Insert	Replace	Move	EDITing actions	Edits / length	No. Words edited	Edited words / length	Delete	Insert	Subs	Shift	Track changes	Nº of errors	TERP
14	Período de bloqueio do teclado - Pode definir o **período de bloqueio** do teclado.	**14**	**Move 1 phrase (2 words) 1 position forward**	Período de bloqueio do teclado - Pode definir o bloqueio **período de** do teclado.	0	0	0	1	1	7.14%	2	14.29%	0	0	0	1	Período de bloqueio do teclado - Pode definir o bloqueio período de ~~bloqueio~~ do teclado.	1	7.14%
15	Se não houver nenhuma operação **no telefone** dentro de um período de tempo predefinido, o teclado é bloqueado automaticamente.	**19**	**Move 1 phrase (2 words) 1 position back**	Se não houver nenhuma **no telefone** operação dentro de um período de tempo predefinido, o teclado é bloqueado automaticamente.	0	0	0	1	1	5.26%	2	10.53%	0	0	0	1	Se não houver nenhuma ~~operação~~ no telefone operação dentro de um período de tempo predefinido, o teclado é bloqueado automaticamente.	1	5.26%
16	Você pode pressionar a **tecla de função** esquerda e a tecla de função direita, por sua vez, para desbloquear o teclado.	**21**	**Move 1 phrase (3 words) 1 position forward**	Você pode pressionar a esquerda **tecla de função** e a tecla de função direita, por sua vez, para desbloquear o teclado.	0	0	0	1	1	4.76%	3	14.29%	0	0	0	1	Você pode pressionar a esquerda tecla de função ~~esquerda~~ e a tecla de função direita, por sua vez, para desbloquear o teclado.	1	4.76%
17	Atalhos - Para atribuir as funções **usadas com freqüência** às teclas de rolagem como teclas de atalho.	**17**	**Move 1 phrase (3 words) 1 position back**	Atalhos - Para atribuir as **usadas com freqüência** funções às teclas de rolagem como teclas de atalho.	0	0	0	1	1	5.88%	3	17.65%	0	0	0	1	Atalhos - Para atribuir as ~~funções~~ usadas com freqüência funções às teclas de rolagem como teclas de atalho.	1	5.88%
18	Auto On e OFF - Para configurar **o telefone** para ligar ou desligar automaticamente.	**14**	**Move 1 phrase (2 words) 2 positions forward**	Auto On e OFF - Para configurar para ligar **o telefone** ou desligar automaticamente.	0	0	0	1	1	7.14%	2	14.29%	0	0	0	1	Auto On e OFF - Para configurar ~~o telefone~~ para ligar o telefone ou desligar automaticamente.	1	7.14%
19	No modo de espera, você pode pressionar **e segurar** para mudar o perfil atual para o perfil silencioso.	**18**	**Move 1 phrase (2 words) 2 positions back**	No modo de espera, você **e segurar** pode pressionar para mudar o perfil atual para o perfil silencioso.	0	0	0	1	1	5.56%	2	11.11%	0	0	0	1	No modo de espera, você ~~pode pressionar~~ e segurar pode pressionar para mudar o perfil atual para o perfil silencioso.	1	5.56%
20	Se o perfil atual estiver offline, você **não poderá usar** essa função.	**12**	**Move 1 phrase (3 words) 2 positions forward**	Se o perfil atual estiver offline, você essa função **não poderá usar**.	0	0	0	1	1	8.33%	3	25.00%	0	0	2	2	Se o perfil atual estiver offline, você essa função não poderá usar ~~essa função~~.	4	33.33%
21	Configurações de exibição - Em seguida, você pode alterar a **senha de restrição** de chamadas.	**15**	**Move 1 phrase (3 words) 2 positions back**	Configurações de exibição - Em seguida, você pode **senha de restrição** alterar a de chamadas.	0	0	0	1	1	6.67%	3	20.00%	0	0	0	1	Configurações de exibição - Em seguida, você pode ~~alterar a~~ senha de restrição alterar a de chamadas.	1	6.67%

Figure 54 – Movement of phrases.

In this case, in segments 14 to 21, although all except one were correctly classified as one edit (movement), the alignments were not correctly made. So, let us look more closely at this set of segments.

In segments 14 to 17, the choice of the application is acceptable, in view of the search for the minimum effort. In the first two segments, although the movements involved 2-word phrases, since these were only moved one position, the application identified this as a single word movement in two positions, in one segment backwards and in the other forwards. In the second pair of segments, the process was the same, but with 3-word phrases. The negative effect of this result is missing the 2-word and 3-word phrase alignments in all these segments. Segment 18 was the only case in this sequence in which both alignment and edit action are correct. In segment 19, the movement of a 2-word phrase two positions back was identified as that of a 2-word phrase movement forward, and the same happened in segment 21, in which a 3-word phrase movement, two positions backwards, was also identified as a 2-word phrase moved forward three positions. So, it seems that movement from right to left (or back) is never correctly identified, being always replaced by movements forward of the words and phrases it takes over.

The only segment that did not have the correct "one movement" score in this group was segment 20 (see below), in which the application identified a total four edits. The effect of the alignment of final words may be observed again in this segment. Still, it is surprising to find this result in a tool that aims at identifying the shortest edit paths. It is indeed difficult to explain how a single action (moving the 3-word phrase *não poderá usar* two positions forward) was identified as being composed of two replacements and two movements, one which even identifies correctly a sub-unit composed of two of the words that were moved (*não poderá).*

Segment ID	[hyp][[Text A - Source text for MT.txt][0020]][][] (back to top)
Number of Edits	4.0
Number of Words	12.0
TERp Score	0.3333333333333333
Original Reference	Se o perfil atual estiver offline, você **não poderá usar** essa função.
Original Hypothesis	Se o perfil atual estiver offline, você essa função **não poderá usar**.

Alignment	**Reference**	se	o	perfil	atual	estiver	offline	você	não	poderá	**usar**	essa	**função.**
											S		**S**
	Hyp After Shifts	se	o	perfil	atual	estiver	offline	você	não	poderá	**função**	essa	**usar.**

Shifts	**Start Pos**	**End Pos**	**Moved To**	**Cost**	**Hyp String**	**Ref String**
	9	10	6	1.0	não poderá	não poderá
	9	9	10	1.0	essa	essa
Phrase Substitutions						

Figure 55 – Movement of phrase to the end of the sentence.

But this segment shows another feature of how estimates of movement are done. One can see at the last lines of the table that, as was described in section 3.5.1, shift privileges phrase-level edits over word-level ones. TER chooses sequences of two words and estimate the effect of moving them. If the number of edits is reduced, i.e. there are less deletions and insertions needed, the shift is recorded. Then, word movements are tested. This shows that TER is setup to look for sequences of only two words. Otherwise, it would have identified the 3-word phrase movement.

The next groups of segments to which more complex actions were applied are not illustrated here, but these may be checked in Appendix 9.

In group 3 (from segments 22 to 37), the tested situations involved two editing actions, in different combinations and extensions. In the first four segments, in which each action was applied to only one word, but twice in the same segment, delete, insert and replace were correctly identified, but movement was not. From this point forward, all movements were from left to right (forward). The next 3 segments, which combined a delete action with either an insert, a replace or a move, all with one-word extension, were correctly classified. From that point onwards, the combinations of actions were not accurately identified, and these always involved 2 or 3-word phrases. Even the few examples of a correct identification of the number of edits (segments 31 and 34), involved the wrong identification of the moved phrases.

In group 4, three actions were tested in the same segment, each involving only one word. The combinations of delete, insert and replace, besides the combination of delete, insert and move, were correctly identified. However, the last example, with a combination of delete, replace and move, was affected by the "last word" effect. Otherwise, the application would have identified the three actions correctly, as it may be seen in the "track changes" column.

This analysis might have been extended with more combinations of edit actions, longer phrases and longer movements, but the main effects seem to have been identified in these few preliminary tests.

A review of edit metrics as process data

The purpose of using an edit metric in this part of this study was two-fold:

- To analyse the editing production during the workshop;
- To assess whether the metric accurately recorded the editing operations performed by the translators, as they selected the editing actions in PE mode.

It is important to emphasise the fact that these metrics are applied *a posteriori*, as instruments of analysis of products, not as logs of the actual processes, while they are being performed. For example, when a metric presents an edit rate based on a deletion of a word, that may not have been the action performed, but the action that the metric considers as the minimum required to transform the Reference into the Result. However, if that was not the actual action, or if the tool does not identify correctly which text unit was edited, this may show that TER is not capable of informing the learning module of an interactive PE tool with accurate data.

Character-based metrics and edit distances may be very informative about the editing processes, but these are not suitable to a system that needs to learn word and phrase alignments and the editing actions applied to these textual units. However, there is one situation in which a below-word analysis might be called for. This is the case of replacements of only one or a few characters in one word, because, for instance, of errors of inflection or capitalization. Although in the examples analysed above there were no edits concerning changing of capitalization, one can see in Appendix 8 that all words are un-capitalized at the tokenization stage, before the alignment is made. Besides, punctuation marks (except dashes) are associated with the words that precede them. Edits with such detail as changing the capitalization of a word or punctuation are only counted by character-based metrics.

A system such as the one that is described in Chapter 5 favours the classification of any character-editing in a word as a word replacement. This stresses that these are contextual errors and that the words are closely linked together, being likely alternatives to one another, although inappropriate in the current segment. So, the information saved by a system in terms of full word replacement compensates for the loss of detail at character level. Moreover, there needs to be some balance between descriptive detail and capacity to interpret that detail. Any collection of data during the process must be weighted in view of the final result. Character-level editing includes a lot of interrupted actions, double and back corrections, which have no effect on the final result.

PEC's implementation of TER estimates the edit actions from an alignment of the reference and the edited results. Besides deletions and insertions, it is able to identify replacements (when the two words that are not aligned occupy the same position) and movements, but with some limitations. Furthermore, it counts each word edited as one edit, even when the surrounding words were edited at the same time in one single action. The only exception to this is with movement, which is first estimated as movements of two words.

Word-level logging of edits is a reasonable method, but it does not capture phrases that are edited simultaneously, as when the user deletes a sequence of two words, or moves a phrase with three words.

We have also seen how alignment-based edits (deletion, insertion and replacement) are easier to estimate, and are estimated before movement. This action implies a search process that is recognised as particularly complex. The strategies used to reduce the complexity of this estimation also reduced its capacity to handle movements of sequences longer than two words. Its capacity to handle backwards and forwards movements is not clear, but it also seems to be affected by the estimation method. Besides, the end of the sentence constraint adds up the limitations of this type of estimate to actually model which editing actions were performed by the user over which textual units.

TER seems to be a good approach to measuring editing, but it does not seem to fulfil the second purpose of this analysis: identifying correctly the four editing actions as performed by translators. Still, TER fulfils the first purpose, being useful for the comparison of the rate of editing between the different variables in the collected data. That is how it will be used for the rest of the dissertation.

Total edit scores per segment

When PEC calculated the edit scores in this data, it did it per event, comparing each reference and result in each line. So, in PE mode, it estimated edit scores for each line that contains an editing action. As users select one editing action at a time, these lines should always show an edit score of 1. But that is not the case, as will be seen in section 6.3.10.

Besides, it does not sum up these lines and present the total of edits in editing actions in a segment. Instead, it recalculates the edit scores for the values in the confirm segment line, in which the initial MT hypothesis is compared against the final editing result. This means that one may have a sequence of action events that would total 3 deletions and 2 moves (5 edits) in the same segment, but then, in the confirm-segment line of that segment, PEC only estimates 3 edits, such as 2 replacements and 1 move. The effect of this is that the editing action events must be analysed separately from the analyses of the other variables.

6.3.4. Combining duration and edit scores

Besides duration and edit scores, the available data allows the analysis of the relation between these parameters. If one has a good measure of the number of edits applied in a full PE work, and a fair register of the duration of each editing sequence in a segment, it may be interesting to know how many seconds on average did each edit require.

There are several ways to measure this relation in the literature. In "Advances in Computer Aided Translation Beyond Post-Editing", Phillip Koehn presents two measures of "speed": seconds/word and words/hour (Koehn, 2015). Both are presented as measures of "productivity" and used as measures of "translator variability".

In this study, "speed" is measured by dividing the duration of each case by the number of edits in that case. So, it presents values of seconds per edit. This metric is more meaningful if approached not at the event level, but as the speed of each sequence of edits in each segment or text. Speed may be a useful metric to compare interface improvements or upgrades in working methods and resources provided by editing tools. This does not assume any correlation to effort, but simply a combination of duration for editing actions.

For the global analysis of results presented in section 6.4, it was required to have a measurement of speed that had a relation of direct proportionality with the other variables, i.e., that the better values were always in higher positions in the scales. For this reason, for these analyses, the inverted speed measure was used: number of edits per minute.

6.3.5. Introduction to the description of the main variables

In the next sections, the results from the five main variables in the activity data will be presented. The main variables are: texts, segments, users, editing modes and editing actions. These are descriptions of results only, and they do not include any statistical analyses yet. After this description, section 6.4 will present a selection of global conclusions on all variables, after a statistical analysis of correlations between all variables, both from questionnaires and from activity logs.

Each main variable is described below in terms of their values in different measures:

- number of cases (events or segments);
- duration (editing time);
- length (number of words of reference);
- number of edits (estimated by PEC);
- TER (estimated by PEC);
- 25% editing threshold (TER score);
- and speed (seconds per edit).

For this description of results, totals, averages and percentages will be presented. Not all calculations are presented for all variables, as they are not always relevant, or they would show distorted results if they were retrieved directly from the data. Throughout the analyses, the two editing modes will always be compared, since these are the two methods of work being tested in this dissertation.

6.3.6. Description of main variables – Texts

This section will present the main features identified in the three texts that were edited in the workshop. The next section is an extension of this, since it looks at the segments that compose each text. The first result to be discussed is the number of lines, or cases in the tables, which reveal how far participants in the workshop went in editing each text.

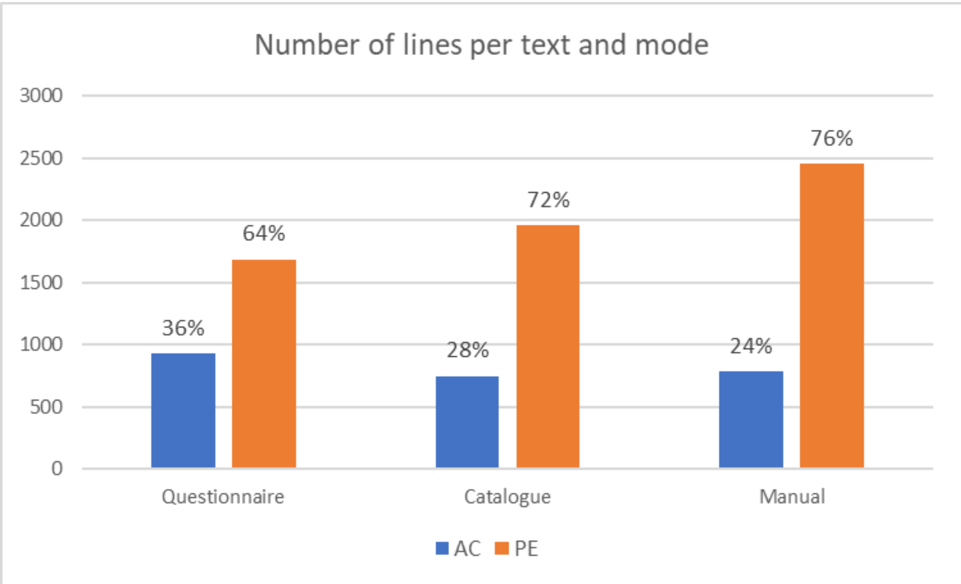


Figure 56 – Number of lines for each text and mode.

The chart above shows the relative number of lines in the table for each text. The percentages are the weights that each mode has in each text. PE mode is clearly the mode that accounts for the majority of events registered in each text. Since the last element that defines the tuples is events, this comes as no surprise: PE has many more cases, due to the editing actions being recorded as events. Globally, the text identified as "Manual" is the one with the most events (3240 in total), while "Questionnaire" is the one with the lowest number of cases (2616 lines in total). Users logged a total of 2709 events in "Catalogue".

The type of events in each text is the theme of the next chart.

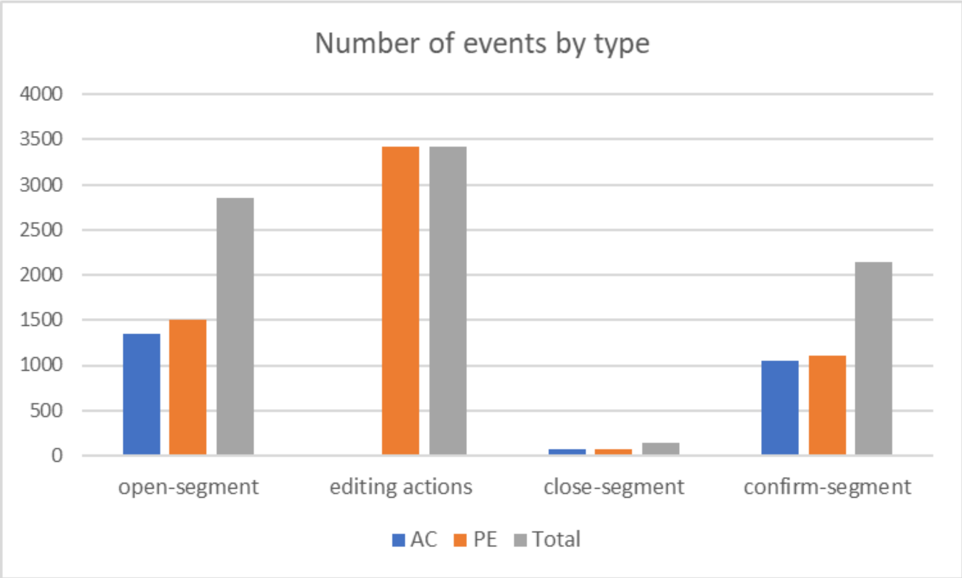


Figure 57 – Number of events by type for each mode.

The highest numbers of events are editing actions, even when these events are only registered in PE mode. Next, are open-segment events, then confirm-segments and finally, with virtually no impact, close-segment events. The latter type of events only occurs when a user decides to reopen a segment, so it is natural that these are not very frequent. Of all these events, the only ones that occur once per segment are confirm-segments, but this is only true after the cleaning process described above.

"Duration" is the time that users dedicated to editing, within the time allocated for each session. Due to the lack of proportion in the time assigned to the different texts, as described in section 6.3.2, the comparative analysis of duration between texts cannot be made directly from the data collected. Besides, since the number and types of events is also so varied, an estimation of average duration of editing per text is not very relevant either. So, only a global analysis of total duration of texts, compared to the total time available for each session, is presented.

The sum of all editing times (duration), by all users in all texts, totals nearly 24 hours. This accounts for ca. 90% of the total time (about 26 hours) made available for all work sessions. The percentage of available time used in each text and mode was not equal, with most users logging more of the available time in PE mode than in AC mode (95% vs. 83%). Questionnaire in PE mode was the session which users edited for a higher percentage (98%) of the available time, and Catalogue in AC mode was the session in which a lower percentage of time (81%) was used.

	AC	PE	Total
Questionnaire	85%	98%	91%
Catalogue	81%	96%	89%
Manual	85%	92%	89%
Average	83%	95%	90%

Table 19 – Percentage of available time used per text and mode.

This unbalanced use of available time also had an effect on the proportion between the data collected for each text and mode, as commented above.

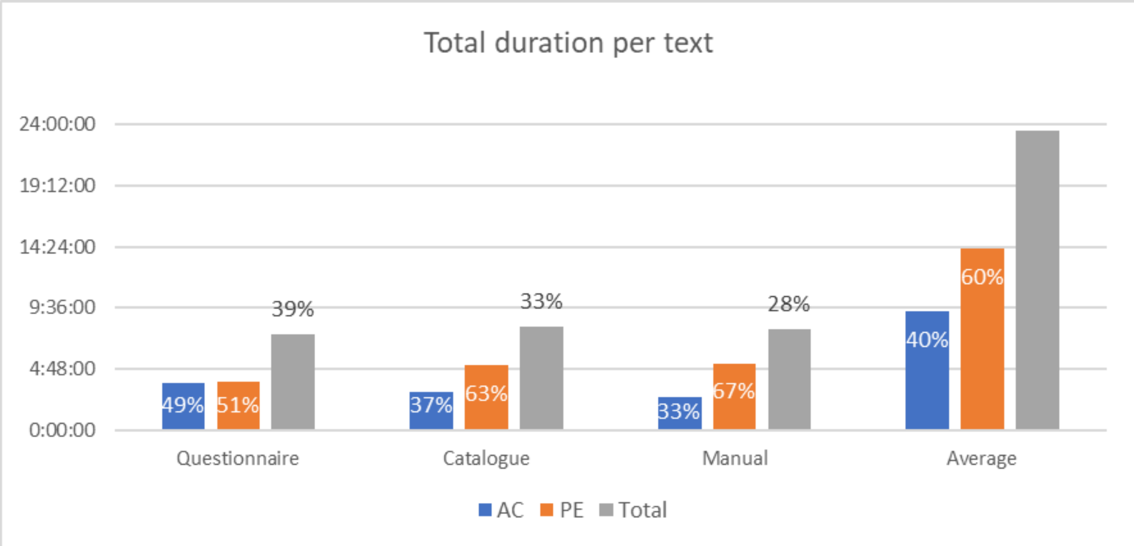


Figure 58 – Total duration per text.

After registering durations for all texts, it was observed that Questionnaire was the text for which the most time was logged (ca. 40% of the total time for all texts), while Manual was the text with less time logged (28%).

Within the sessions for each text, the ratio between AC and PE modes was also varied. For Questionnaire, the ratio between AC and PE is virtually the same in all logged sessions (49/51). In Catalogue, this ratio is 37/63 and for Manual it is 33/67. Globally, the planned ratio of 43/57 for the two editing modes suffered a slight change in the direction of PE mode, being now at 40/60.

The analysis of length (number of words) cannot be used to compare texts, since the logs only include the number of edited segments (and words) and these values depend on all other variables. For example, one single user edited up to segment 40 in Questionnaire in AC mode, but another user, who processed more segments in PE mode, only went as far as segment 31 in the same text. So, totals and average lengths at the text level are not relevant.

The number of edits is the sum of all the edits that PEC has identified in each event in confirm-segment events only. As explained in 6.3.3, these events register the editing scores between the initial MT hypothesis and the final result in each segment. The following table describes figures for total number of edits, and the percentage of contribution of these values to the total.

	AC	PE	Total		AC	PE	Total
Questionnaire	1251	916	2167		58%	42%	33%
Catalogue	1164	1088	2252		52%	48%	35%
Manual	936	1137	2073		45%	55%	32%
Total	**3351**	**3141**	**6492**		**52%**	**48%**	**100%**

Table 20 – Distribution of number of edits per text and mode.

In total, the translators produced about 6500 edits, a little above 2000 edits per text, which represents a quasi-perfect distribution of 33% of all edits for each text (last column). The internal distribution of edits between each mode is not so balanced. For Questionnaire and Catalogue, most edits were done in AC mode (58% in Questionnaire and 52% in Catalogue), whereas, for Manual, most edits were done in PE mode (55%). Since the number of edited segments was different for all texts and modes, averages are not conclusive, so this will be analysed in more detail below.

TER scores were calculated by PEC for all edited segments, as the ratio between the number of edits on each segment and its length. Since each segment was edited in different sessions by different users, total TER per segment is not relevant, but average TER scores may be analysed per text.

	AC	PE	Total
Questionnaire	17%	18%	18%
Catalogue	34%	29%	31%
Manual	29%	24%	26%
Total average	**26%**	**24%**	**25%**

Table 21 – Average TER scores per text and mode.

Each text presents an average global TER score (last column) in a different range relative to the suggested editing threshold of 25% (see section 5.2.3). In Questionnaire, the global TER average is below the threshold, at 18%, in Manual it is virtually at the threshold (26%), and Catalogue is above the threshold, with 31% average global TER score. The distances to the threshold seem to be equal, which makes the global average TER score for the three texts exactly 25%, at the editing threshold.

Catalogue shows the highest average TER scores in both modes. AC mode seems to produce a higher average TER score than PE mode for two of the three texts.

Speed measures the number of seconds it takes, on average, to make an edit. To analyse this at text level, only average values are applicable. Total duration was calculated as a sum of all events per text. Then, the total number of edits, in confirmed segments only per text, was retrieved, as this retains all edits between MT hypotheses and final versions. The average speed per text is the result of the division of the total duration divided by the total number of edits.

As the table below shows, global average speed in these texts is 13 seconds per edit. (These values are close to those that were identified in the study presented in Popovic, Lommel, Burchardt, Avramidis, & Uszkoreit, 2014). Users were generally faster making edits in Questionnaire than they were in the other texts, and they were faster in AC mode than in PE mode. The differences in speed seem to be larger between editing modes than between texts.

	AC	**PE**	**Average**
Questionnaire	00:00:11	00:00:15	00:00:12
Catalogue	00:00:09	00:00:17	00:00:13
Manual	00:00:10	00:00:17	00:00:14
Average	**00:00:10**	**00:00:16**	**00:00:13**

Table 22 – Average speeds in texts and modes.

6.3.7. Description of main variables – Segments

The first three charts in this section show the number of confirmed segments in each text (blue and orange curves), together with the smaller count of segments that were reopened and re-edited (grey and yellow curves), per mode.

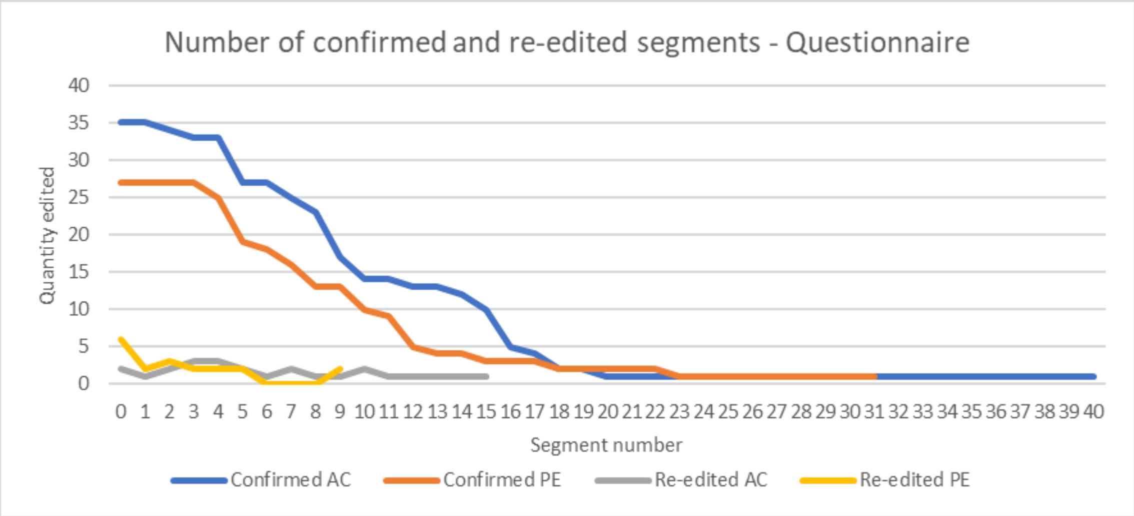


Figure 59 – Confirmed and re-edited segments in Questionnaire.

This chart shows two curves that begin at their highest peak, and then come down, with a long sequence of very low values. As each user only confirmed each segment once, the number of times each segment was edited represents the number of users that edited each segment. Thirty-five users edited this text in AC mode, but two of them only edited segments 0 and 1. So, the curve falls down naturally, as we move along the texts and the number of users editing segments decreases. From segment 20 onwards, only one user edited the segments in AC mode, but he reached segment 40. In PE mode, the highest number of users editing the text was 27 and this number remained the same until segment 3 of the text. After segment 23, there was only one user editing the text in PE mode, and this user went up to segment 31. The fact that only one user extended the editing down a text is a common characteristic in all these analyses.

In this text, AC mode was clearly the method that allowed for more segments to be edited. However, in view of the imbalance in terms of time allocated for each mode in this text, this is not conclusive about the correlation between text and mode, in terms of productivity.

Re-edited segments are obviously a smaller number than those that are edited only once, but there are still 44 segments in this text that were re-edited. This operation only occurs at the beginning of the text – only up to segment 15 in AC mode. Segment "0" is the segment with more re-editing events, but only in PE mode. Segments 2 to 5 have been re-edited in both modes, and then 7, 9 and 10 in only one of the modes. The number of re-editing events per segment is quite low, below 5.

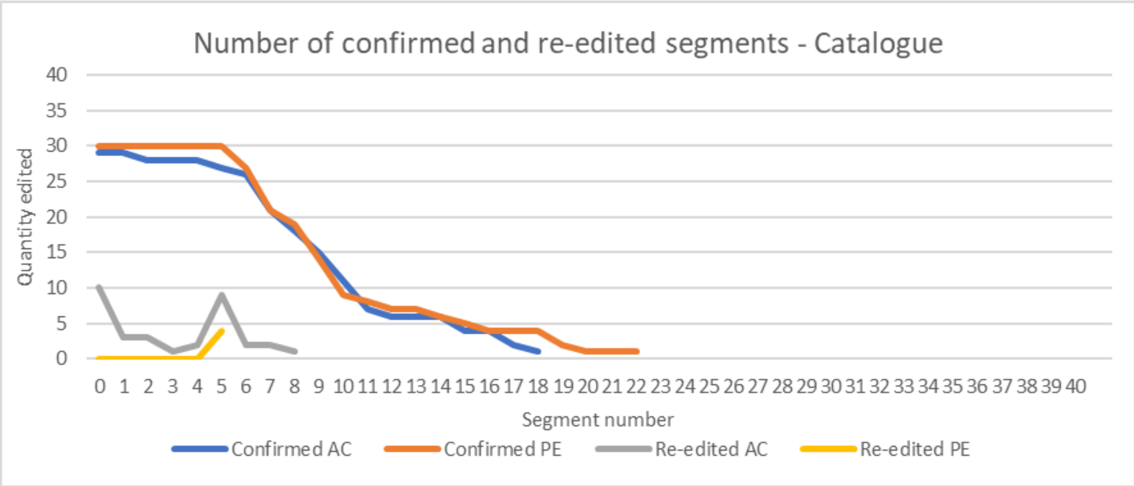


Figure 60 – Confirmed and re-edited segments in Catalogue.

In Catalogue, users did not edit more than 23 segments, in PE mode. Most users did not edit beyond segment 18 in PE mode and segment 15 in AC mode. In this text, PE mode seems to be associated with a higher production, as more users edited segments in this mode, but, again, the lack of proportion between the two modes hinders this analysis.

As for re-edited segments, there are a total of 37 re-editing events (the lowest total in the three texts), but the numbers are concentrated in a list of a few segments, with many such actions each. Up to 10 re-editing events in segments 0 and 5, for example. In PE mode, only one segment was re-edited (segment 5), but it was re-edited 4 times in total, by different users.

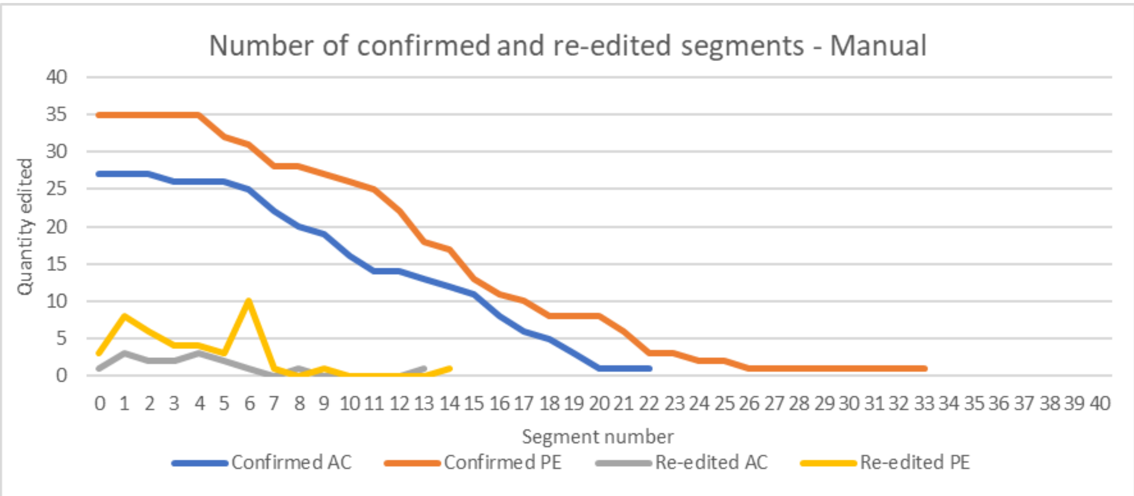


Figure 61 – Confirmed and re-edited segments in Manual.

The Manual is the text to which most time was allocated, in total, in the distribution, especially in PE mode. In AC mode, it was the text to which least time was allocated. This may account for such a long distance between the two modes in terms of users that edited each segment, as in PE mode all users edited more segments than in AC mode. The highest number of segments edited in this text was 33 segments, in PE mode, and only 22 in AC mode.

Re-edited segments are very frequent in this text, 57 in total, with more cases in PE mode (41). Segments 1 and 6 concentrate most of this effort, but the list of segments that required re-editing is longer than the list in other texts.

"Length" is the number of words per sentence, not in the source English text, but in the MT hypotheses in Portuguese, created by MateCAT. The values for length were calculated by PEC and, since this is usually proportional to the length of the source sentence, it may be used to characterise all texts. The following chart describes the length of each segment in each text that was edited during the workshop.

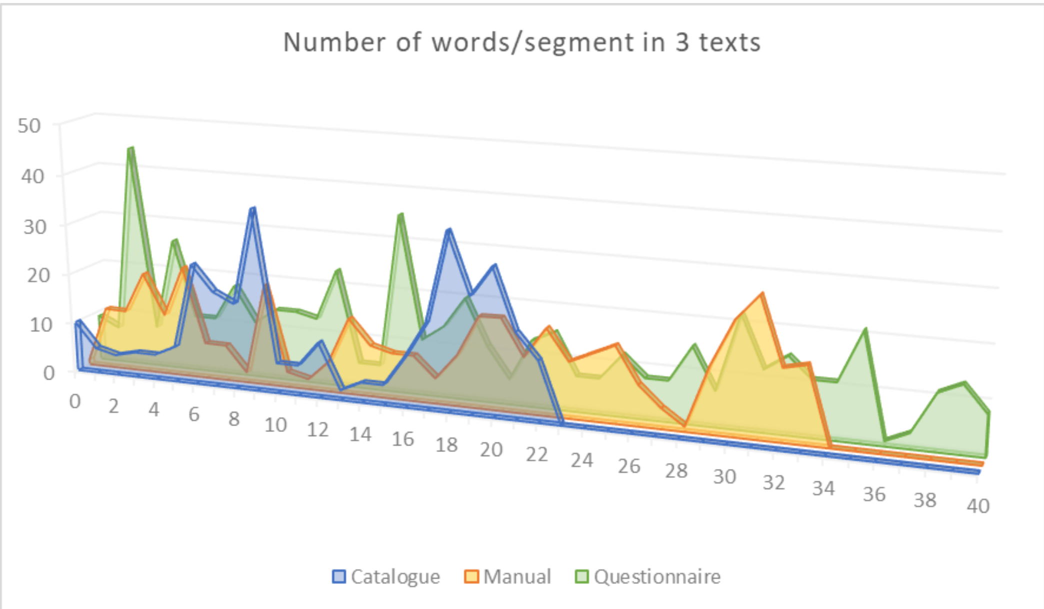


Figure 62 – Words per segment in 3 texts.

The curves in the graph describe the edited segments only. That is the reason why the curve for the "Questionnaire" (green) is longer than those for the other two texts. This text also features the longest segment, with 44 words, and the fourth longest, with 35 words. The other text with segments above 30 words is the "Catalogue", showing two sentences with 36 words. This text is, at the same time, the one with the highest percentage of short sentences, below 10 words (45%), whereas "Manual" has

40% of such sentences and “Questionnaire” has 35%. The distribution chart below illustrates how the texts diverge in this feature.

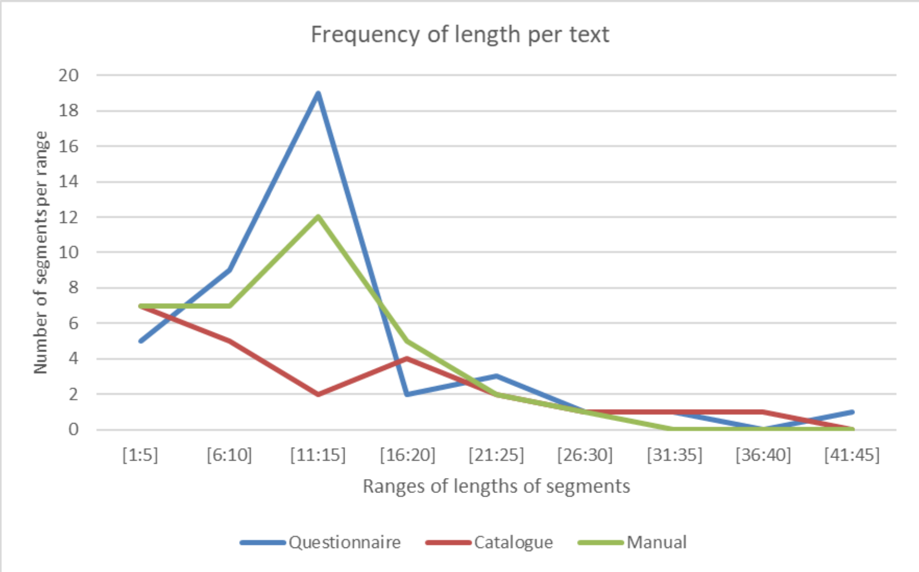


Figure 63 – Frequency of length in all texts.

The horizontal axis (categories), at the bottom, represents different ranges of lengths, at intervals of 5 words, starting from segments with between 1 to 5 words on the left to segments with between 41 to 45 words on the right. The values in the vertical axis are the numbers of edited segments in each range.

Questionnaire has a high dispersion of lengths, from five sentences that have from 1 to 5 words (on the left) to one sentence that has more than 40 words (on the right). Most cases (19) are in medium length sentences, between 11 and 15 words. This is the text with the highest concentration of cases in a well-defined peak.

The dispersion is much lower in Catalogue. The highest number of sentences (7) in one range are very short sentences, with no more than 5 words. This text is also the one with the least number of sentences within the most common average of between 11 and 15 words. However, it stretches all the way to the up to 40-word range, with one sentence in each of the three ranges above 26 words.

In Manual, the dispersion is not very big, as there are not so many cases at each extreme of the curve (the longest sentence is below 30 words) and the highest number of segments in the same range is only 12, in the range of lengths between 11 and 15 words.

“Duration” measures the time each user dedicated to editing each segment. Since there is a wide range of users editing the same segment, from segments which were edited by 35 users to others that were edited only by one, total durations are not a valid factor of comparison between texts. So, the analysis on duration in this section only involves average durations per segment on each text, as seen in the three charts below. The duration of each segment includes all events associated with that segment, including the time spent reopening and re-editing the same segment, and the averages were calculated for the full editing of each segment. Let us first look at the distribution of the average duration in the edited segments of the three texts.

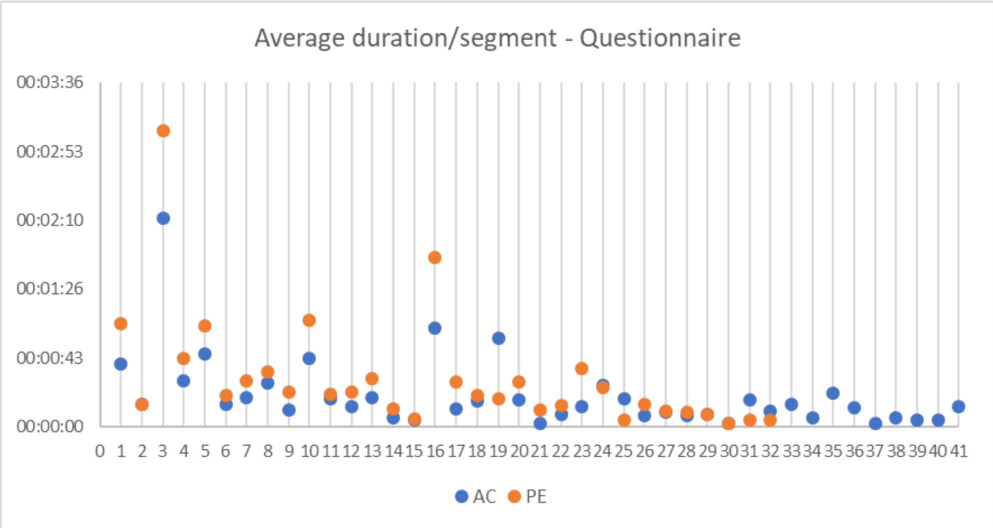


Figure 64 – Average duration/segment in Questionnaire.

This text has very long segments, such as segment 2 and 15 (note that numbers in the categories line are incorrect, as these segments appear as 3 and 16). These segments are the ones that show the longest editing duration. In this chart, one can also see how PE average durations are always above AC durations, on each segment (vertical line), with only a few exceptions. Most segments have low average durations, below 00:43:00 (43 seconds). Most segments with longest durations are at the beginning of the text, and the edited segments at the opposite side have very low average durations.

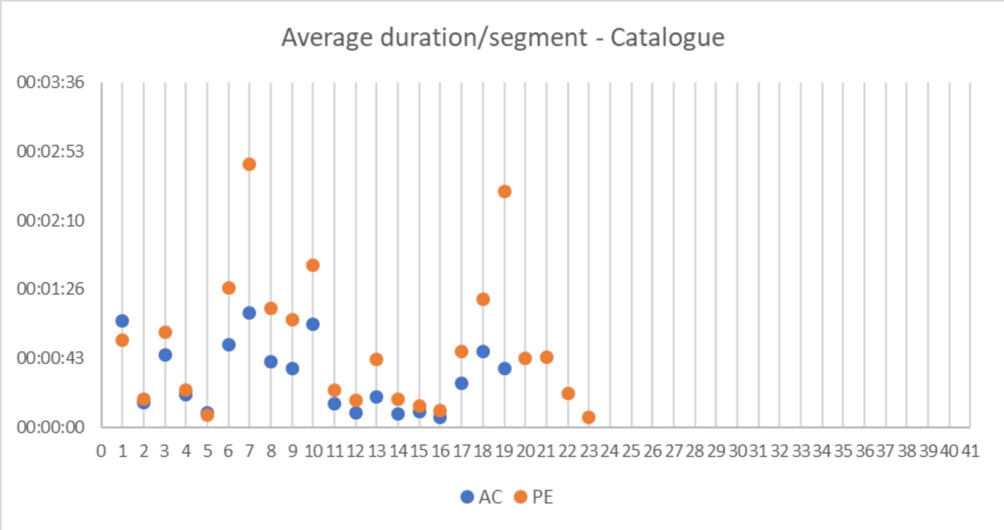


Figure 65 – Average duration/segment in Catalogue.

In Catalogue, two segments pop up as the ones that took, on average, longest to edit. These are, again, the two longest segments in the text. And, as in the previous text, PE seems to be the mode that requires more time. Besides these outlying cases, most other segments are distributed either below the 00:00:43 line or below 00:01:30.

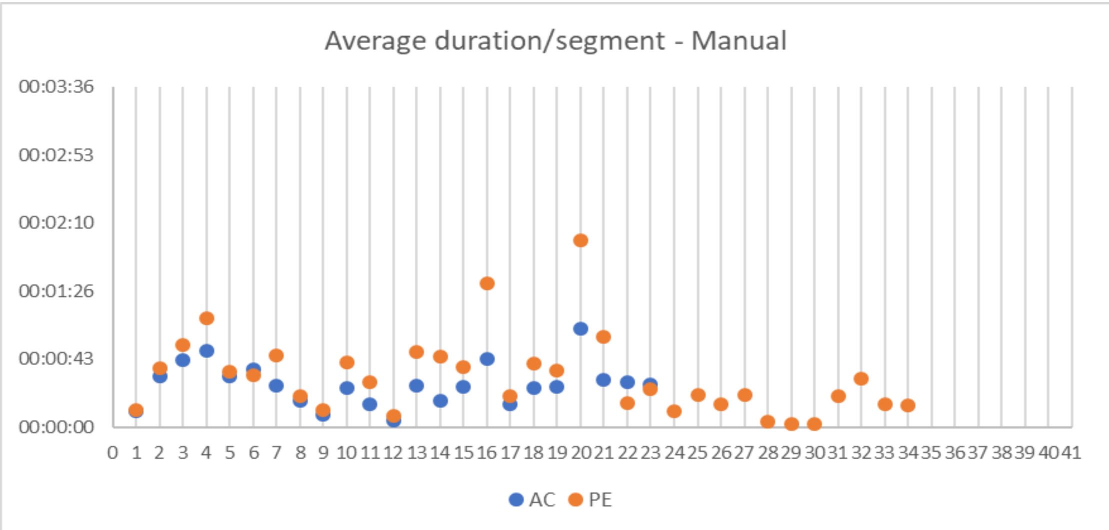


Figure 66 – Average duration/segment in Manual.

“Manual” seems to be the text with the most regular distribution of duration, with most segments showing durations around 00:01:00. In this text, PE mode was again the mode that seems to require more editing time.

Just like “duration”, the number of edits per segment is not very relevant in terms of sums, but rather in terms of averages, which dilute the effect of the differences in the number of times each segment was edited. The three charts below describe the number of edits on each segment, averaged from all users.

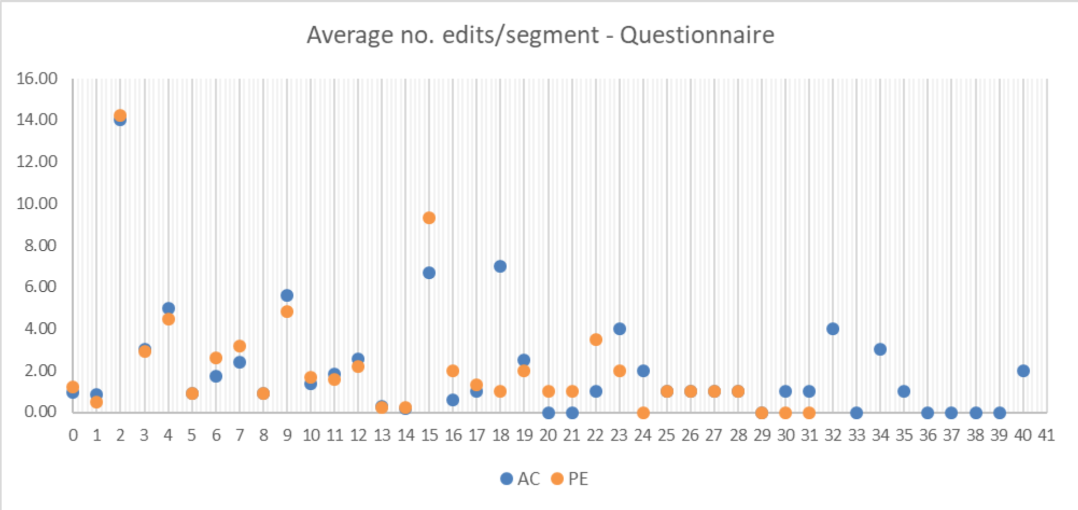


Figure 67 – Average edits/segment in Questionnaire.

Some of the features of the previous chart on duration of this text seem to repeat, with the longest segments at a long distance, here reaching 10 and 14 edits in each. But there is a particular feature that is also visible: in terms of number of edits per segment, AC mode is no longer always below PE mode. In fact, there are 12 segments (nearly half of the 28 segments that were edited in both modes) in which AC involved more edits than PE mode. Another interesting observation is that the number of segments that seem to have required no editing (at the bottom of the chart) is quite low, and it is almost always in AC mode. Most of the segments show numbers of edits between 1 and 5.

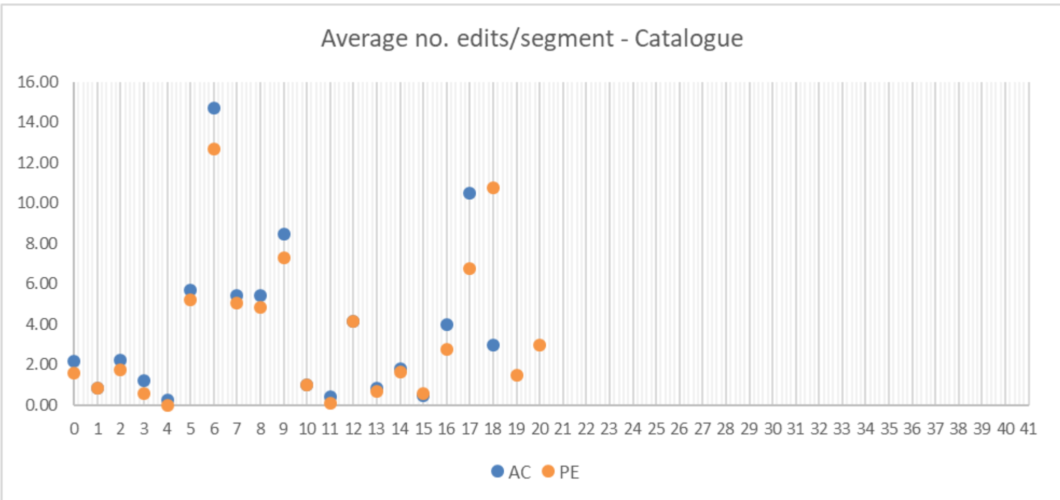


Figure 68 – Average edits/segment in Catalogue.

The same tendencies are visible in the chart for Catalogue. In this case, the number of segments in PE mode that involved more edits than AC is even lower, and there are only two segments that do not require editing. However, this text seems to have more cases above six edits per segment.

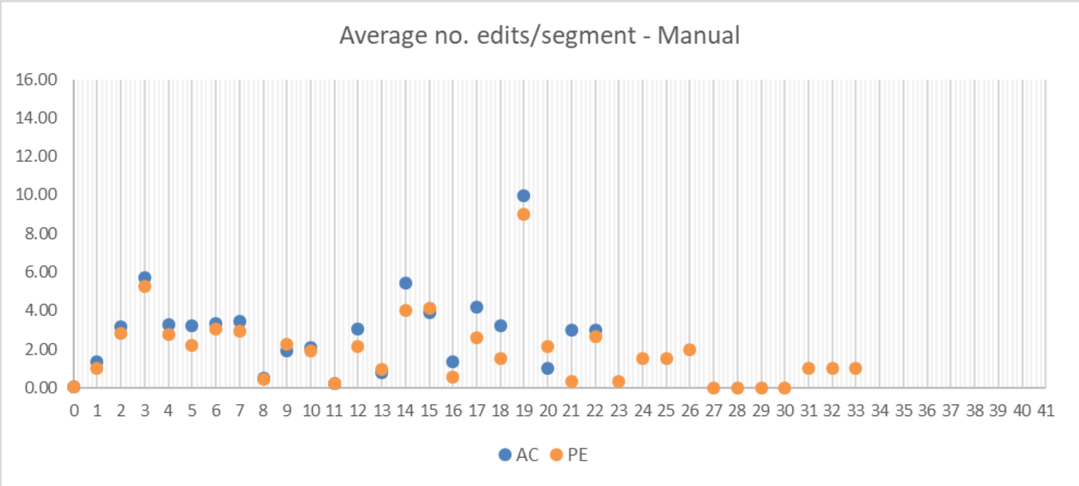


Figure 69 – Average edits/segment in Manual.

Finally, the chart for Manual confirms the tendency, with PE mode virtually only overtaking the number of edits in AC mode when users stopped editing in this mode. Besides, only six segments require no or little editing. The distribution of cases is very consistent, with most segments below four edits per segment.

TER scores describe the density of edits per number of words in the reference sentence. As a ratio of the number of edits per number of words, TER can be expressed by a decimal number or a percentage. For a simpler visualization, TER is always shown as a percentage in this dissertation.

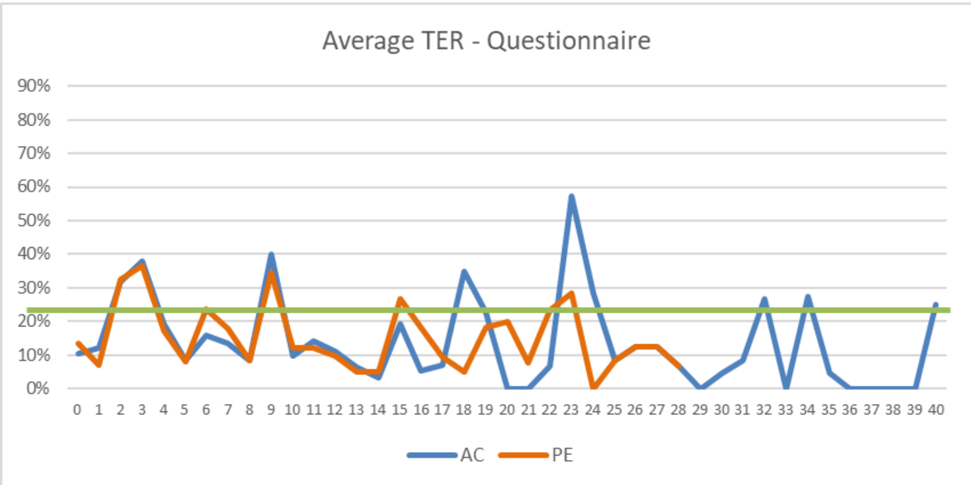


Figure 70 – Average TER score in Questionnaire.

In Questionnaire, users produced TER scores mostly between 10 and 30%. Most cases are, actually, below 13% TER. In the sentence with the highest TER, users edited on average 57% of the words in the reference, in AC mode.

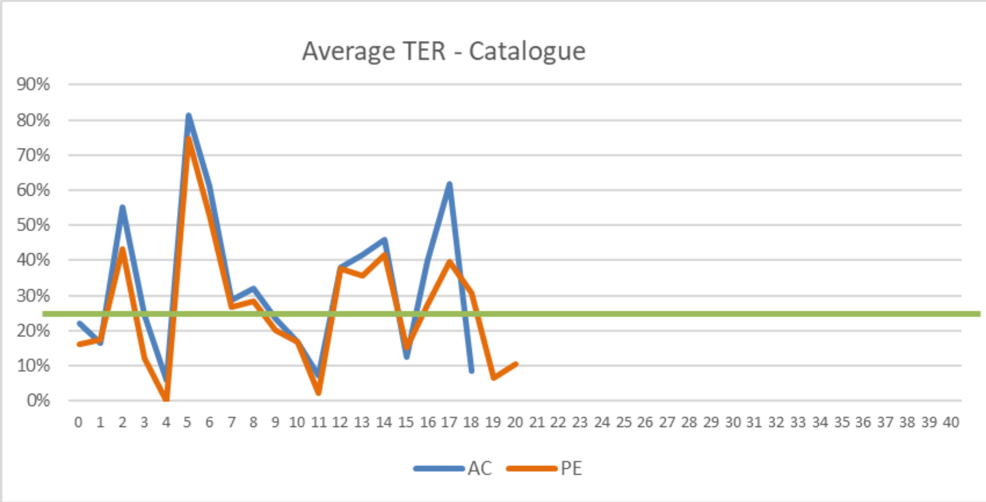


Figure 71 – Average TER score in Catalogue.

In Catalogue, TER scores are, generally, much higher. There are even several cases in which more than 40% of the words were edited, mostly in AC mode, but also in PE mode. The highest TER score is 81% in AC mode, and just below it, in the same segment, 75% in PE mode. It is important to note that this is not the longest segment, so it seems that there is not a correlation between length and TER in this case.

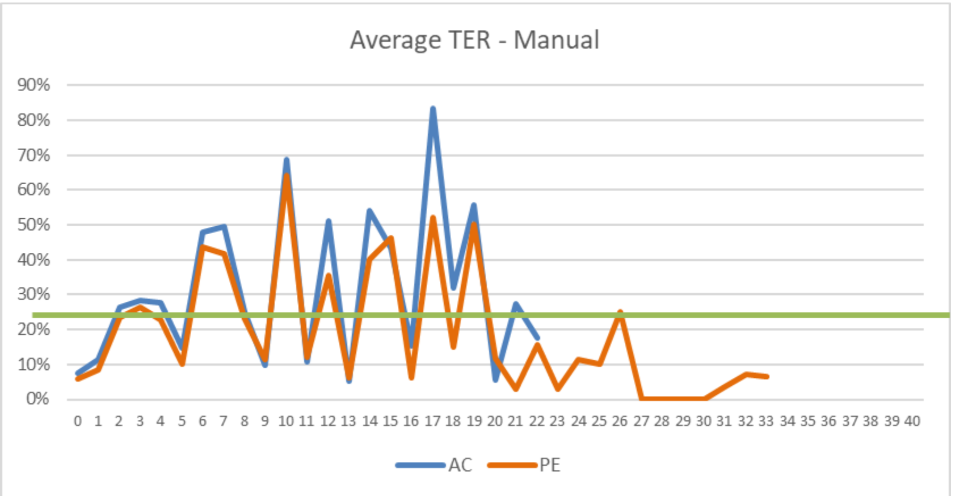


Figure 72 – Average TER score in Manual.

Manual has most cases concentrated around the 25% threshold and below 30%. But there are several cases with TER scores above 40%, in both modes. The highest average TER score is 83% in AC mode and 64% in PE mode, in different sentences.

Analysis of the TER threshold per segment and text

The green line in these three charts shows the 25% editing threshold, which was proposed as the limit for what may be accepted as pure editing. There are several segments in all texts that go over this line, so let us see this in more detail. The table below presents the percentages of segments in each text and mode that had an average TER score equal to 0%, below or equal to 25%, or above the 25% threshold.

	Questionnaire AC	Catalogue AC	Manual AC	Questionnaire PE	Catalogue PE	Manual PE	Global average
TER=0%	20%	0%	0%	3%	5%	12%	8%
TER=<25%	61%	47%	43%	79%	43%	62%	58%
TER>25%	20%	53%	57%	17%	52%	26%	34%

Table 23 – Percentage of segments below and above the editing threshold.

We can see that 58% of all segments show average TER scores below 25%, within the editing threshold. But still, globally, 34% of the segments require an intensity of editing above that threshold. Moreover, two texts in AC mode and one text in PE mode show more than 50% of their segments with TER scores above the editing threshold. The total number of segments that required no editing is only 8%, globally, but in Questionnaire in AC mode this number was 20%. This data can be visualised below in the form of a chart.

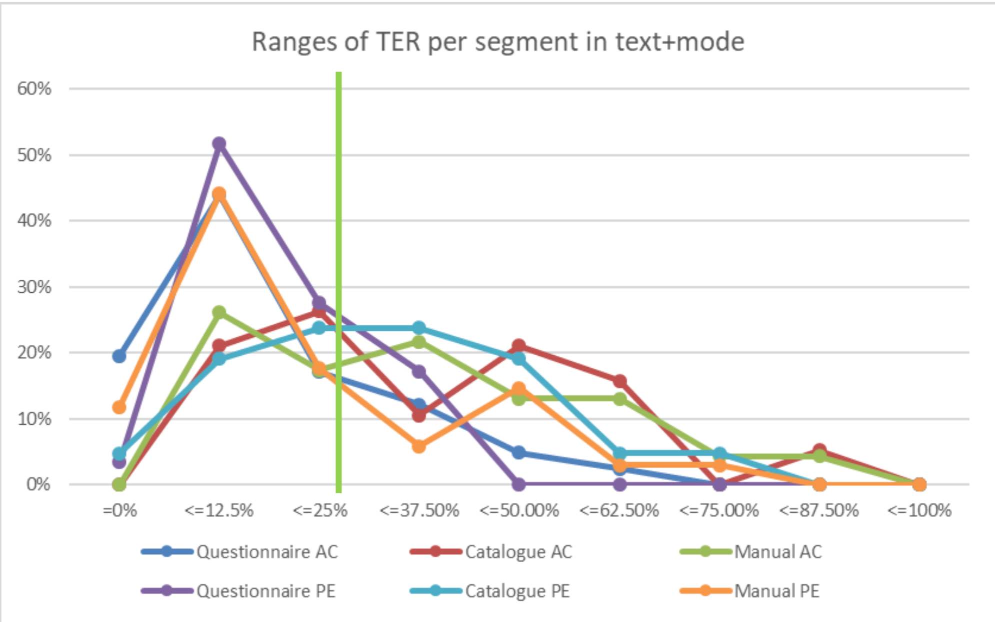


Figure 73 – Distribution of segments in ranges per TER score.

The chart above shows how segment TER averages are distributed in ranges of 12.5%. Each line is a text in one of the two modes. Most segments in virtually all texts are in the range between 1% and 12.5% – only two pairs of text/mode have higher data points in the next level, until 25%, and they are both in Catalogue. However, the number of segments above the editing threshold is still considerable. Percentages of 10% to 20% of the segments in several texts and mode text are in the 25%-37.5%, up to 50% and up to 62.50% ranges. Furthermore, there are several cases of segments with editing scores up to 87.50%. Let us see this distribution in more detail, per text and mode separately.

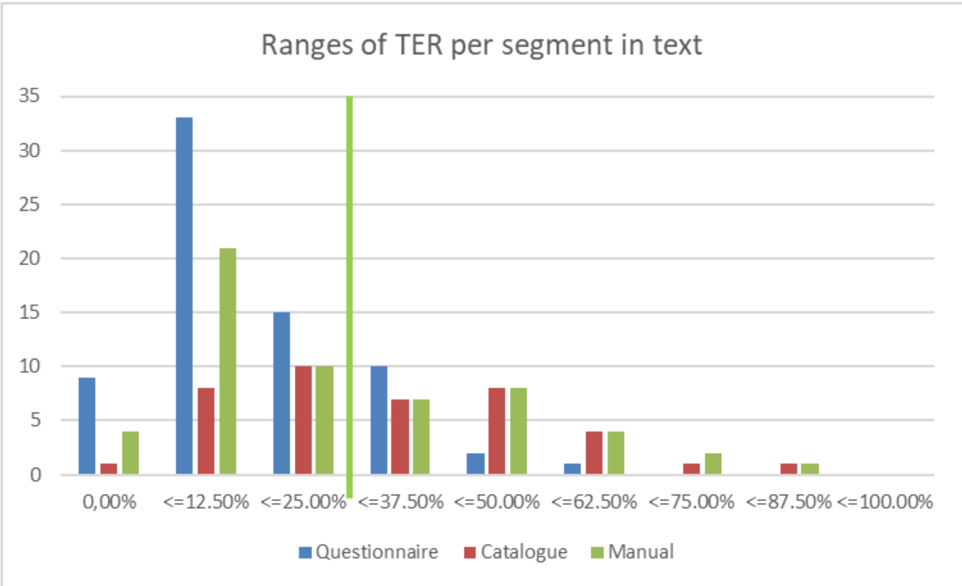


Figure 74 – Distribution of segments in ranges per TER score and per text.

Most of Questionnaire's segments concentrate at below 12.50% TER and virtually disappear after 37.50%. This is what is expected of a regular PE work session. But that is not the behaviour of the other two texts. Catalogue's peak is at 25%, but the second highest number of segments is divided between 12.50% and 50%. And it contains segments that require TER scores of up to 87.50%. Manual peaks at 12.50%, like Questionnaire, but it includes segments that stretch TER scores up to 87.50% too.

	Questionnaire	Catalogue	Manual	Global average
0.00%	13%	3%	7%	8%
<=12.5%	47%	20%	37%	37%
<=25%	21%	25%	18%	21%
<=37.50%	14%	18%	12%	14%
<=50.00%	3%	20%	14%	11%
<=62.50%	1%	10%	7%	5%
<=75.00%	0%	3%	4%	2%
<=87.50%	0%	3%	2%	1%
<=100%	0%	0%	0%	0%
Total average >25%	19%	53%	39%	100%

Table 24 – Percentage of segments per editing score ranges in texts.

This table contains the same data as the chart above. It shows that, for all texts, 58% of their segments are within the ranges below the editing threshold. However, 34% of all segments are above the threshold, and 8% of all segments are at TER score ranges between 50% and 87.50% editing. In Catalogue, 53% of all its segments are above the proposed editing threshold.

Then, assuming the editing threshold shows that above it what is being performed is “translating”, we might say that, in these conditions (taking into account the experimental nature of the tests, the MT hypotheses that were produced and all the other variables), this text should not be considered as appropriate to be translated in a PE process. Moreover, 19% of the segments in Questionnaire and 39% of all segments in Manual should also be accounted for as implying an editing rate superior to that acceptable in a PE project.

Analysis of speed

Speed measures the time taken to make an edit. The analysis of this measure at each segment revealed some interesting results, which question its applicability at this level. It was calculated dividing the total duration of all events in the segment by the number of edits made by all users to each segment, registered only in confirm-segment lines. These are the results for the segments in each text, starting with Questionnaire:

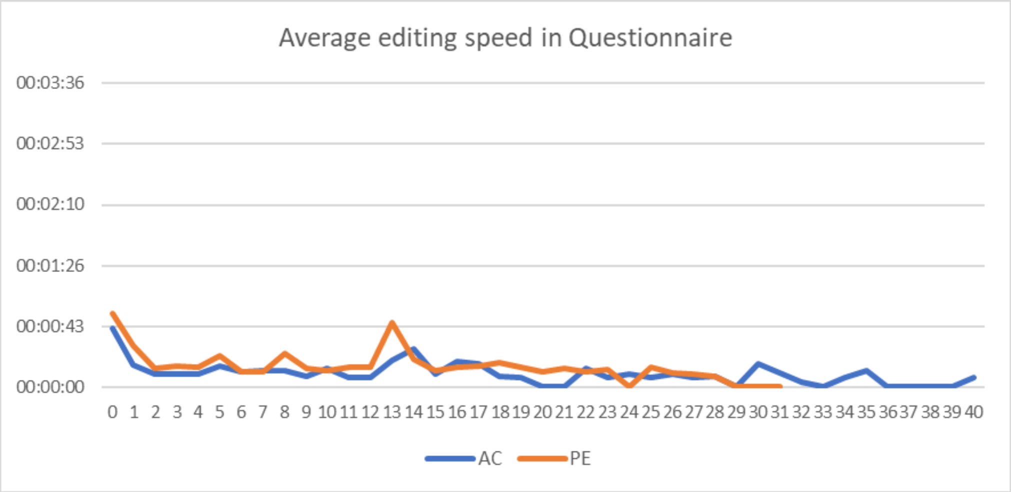


Figure 75 – Average speed in Questionnaire.

The chart for Questionnaire could be more descriptive, zooming in on the details of the distribution below 00:01:00 (one minute), but the reason why the chart has a long vertical case scale is comparability with the charts for the other texts. Here, we can see that the average speed is very low, at around 12 seconds per edit. There are no major differences between the two modes, but PE mode seems to require more time to make each edit. There are a few segments that take longer than the others, but this is not very visible in this chart. The effect of a few outliers is clearer in the other texts.

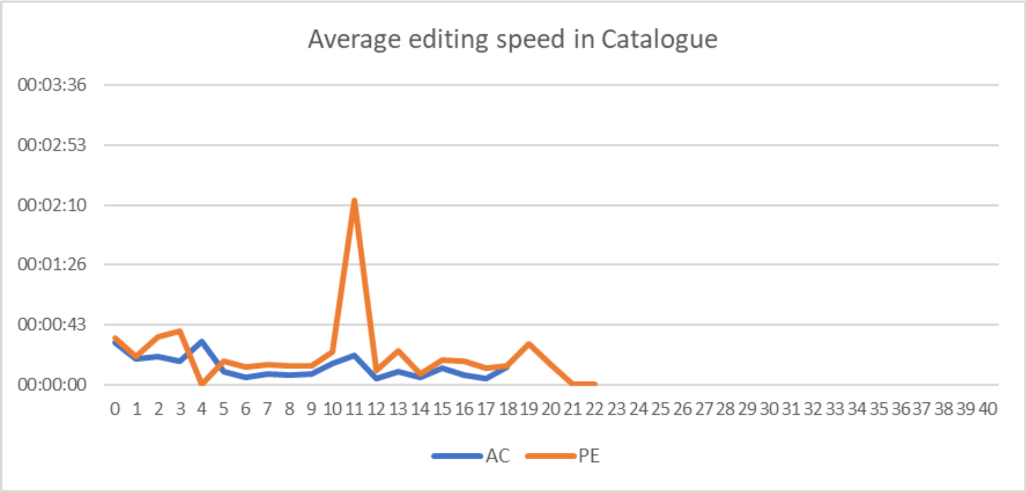


Figure 76 – Average speed in Catalogue.

The highest point in this chart is a segment which was edited by only one user, in which he only did one edit. The average speed of this single edit is, in fact, the total duration of all events registered by all users: a total 00:02:13. This creates this effect, of a segment that goes very much above the average. Without this segment, the global average would be at around 00:00:15. A similar effect is visible in Manual.

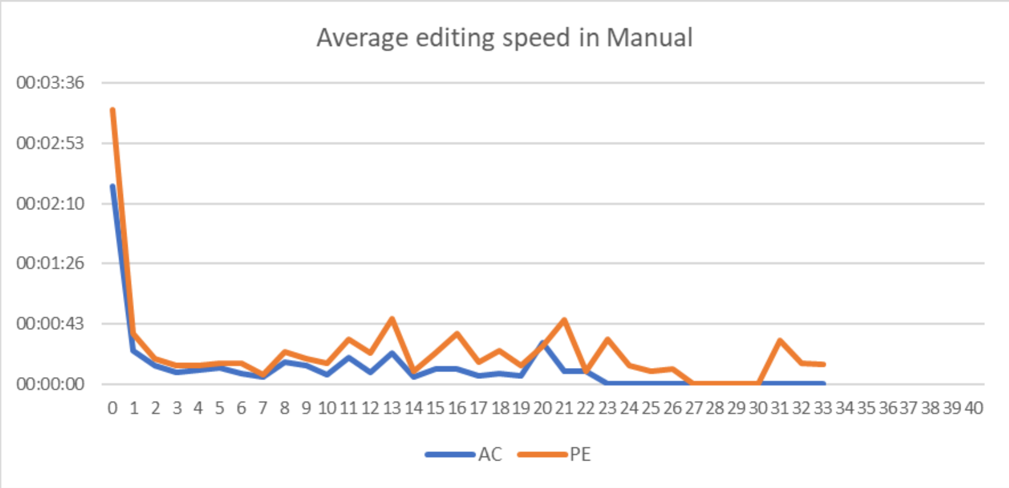


Figure 77 – Average speed in Manual.

The first segment in Manual only had one word. However, two users replaced it, which meant that there were two edits in total to the segment. The whole duration of all users was divided by these two edits to calculate the average speed in that segment. Furthermore, for some reason, this segment was one of the segments with more "open segment" events, some with a fair duration, which added up to the total duration. Apart from this segment, with an average speed that forced the need to raise the scale for all texts up to 00:03:36, the average duration in this text would be at around 00:00:16 (16 seconds).

So, it seems that this estimation of average speed per edit for several users is very sensitive to small numbers of edits which end up being associated with long durations. In the next section, this measure will be tested to compare users, and perhaps it will reveal its usefulness.

6.3.8. Description of main variables – Users

To complement the description of participants from the workshop questionnaires, the activity logs add several parameters to the characterisation of their work. Let us begin by looking at the number of segments each user edited.

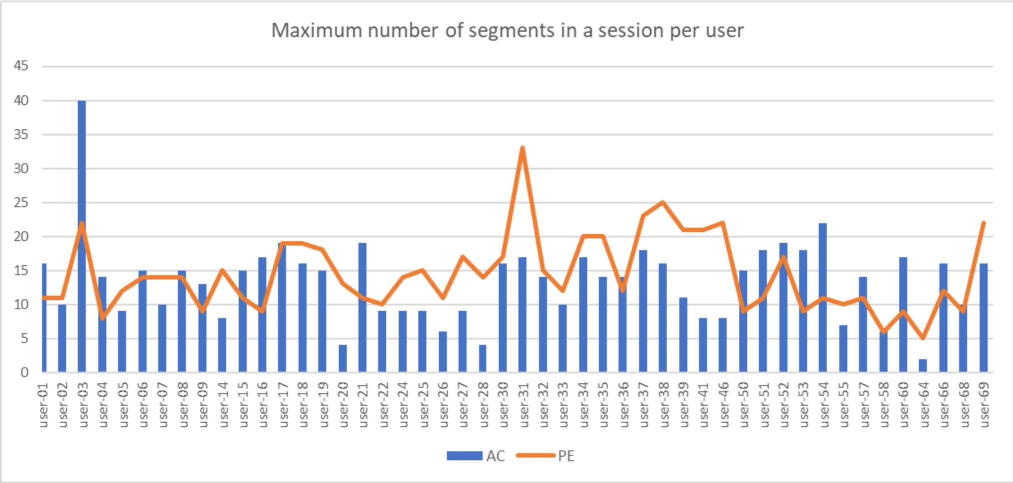


Figure 78 – Number of segments edited per user in a session.

Most users have edited between 10 and 20 segments in a single session. A few users edited less than 10 segments, mostly because they are in the group of users for which there were only records of the two 5-minute sessions. This group also includes the user who did not record the 10-minute session in AC mode, beside one user who did not edit many segments, but kept them open for a long time (see below, on duration). On the opposite end, there are 2 users who have edited more than 30 segments, one in AC mode and the other in PE mode.

Another global count that it is interesting to check is the number of segments that users reopened for re-editing. The distribution of these segments is quite spread out, with 19 users producing a total of 74 such events in AC mode, and 17 users producing 69 of the same type of event in PE mode. The number of these segments per user and mode is as follows.

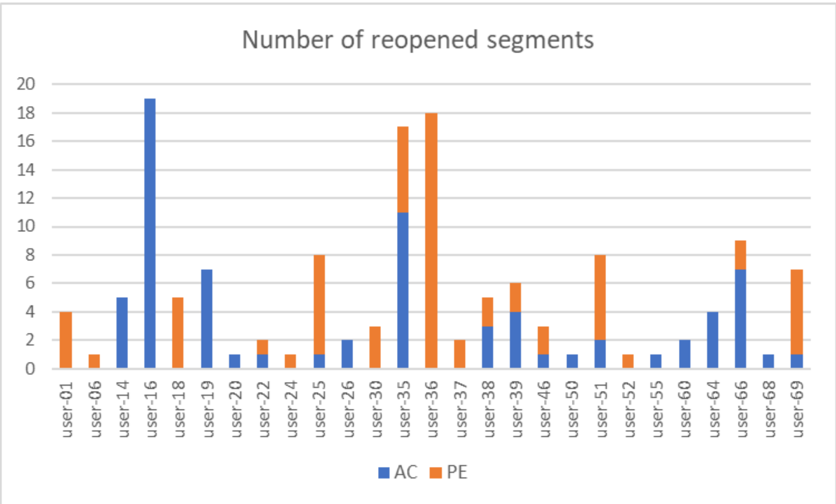


Figure 79 – Number of reopened segments.

Not all users are represented in this chart, but the number of segments that had to be reopened and re-edited is quite large (143). Most of the 27 users in the list reopened segments in only one of the editing modes, 16 users to be more precise. Three users reopened more than 17 segments, or the same segment repeatedly, and two of these did this in only one of the editing modes.

These figures, concerning the numbers of edited segments per user, should be weighted in terms of duration and edit scores, since time and editing effort are better indications of the actual work done by each user. These metrics are described next.

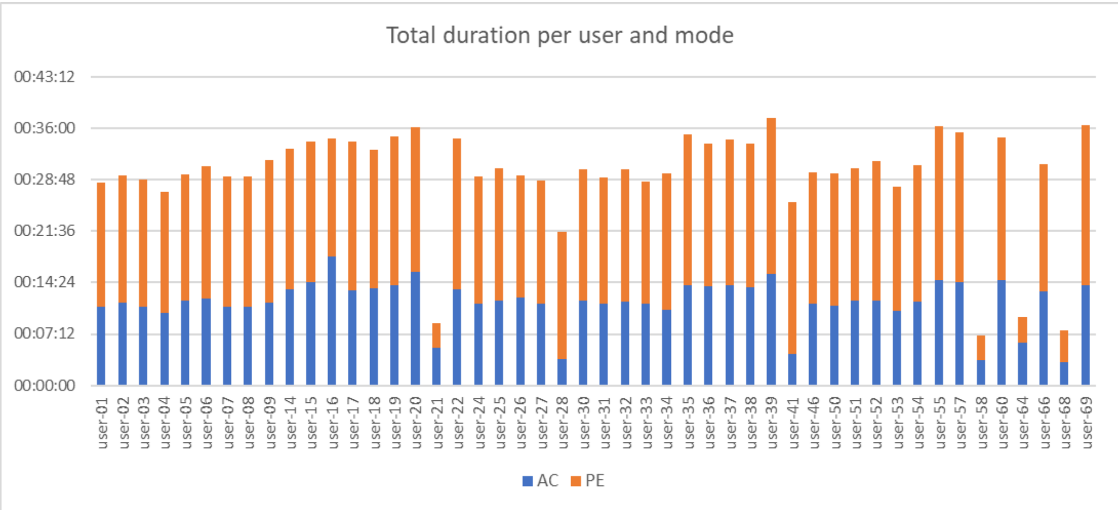


Figure 80 – Total editing time per user and mode.

The total time allowed for the work sessions for each user was 35:00 (15:00 total in AC mode and 20:00 in total in PE mode). Most users kept their editing time (the sum of durations of editing events, excluding initial pauses and pauses between sessions) within that limit, with only a few exceptions. The global average percentage of the available 35 minutes that was used was 78% for AC mode and 89% for PE mode. The four lower bars in the chart above have already been explained, and are associated with those users whose logs only included the two 5:00 sessions. The proportion of total time for each mode follows what was expected, with more time dedicated to PE mode. Average editing times per segment that each user presents are also worthy of analysis.

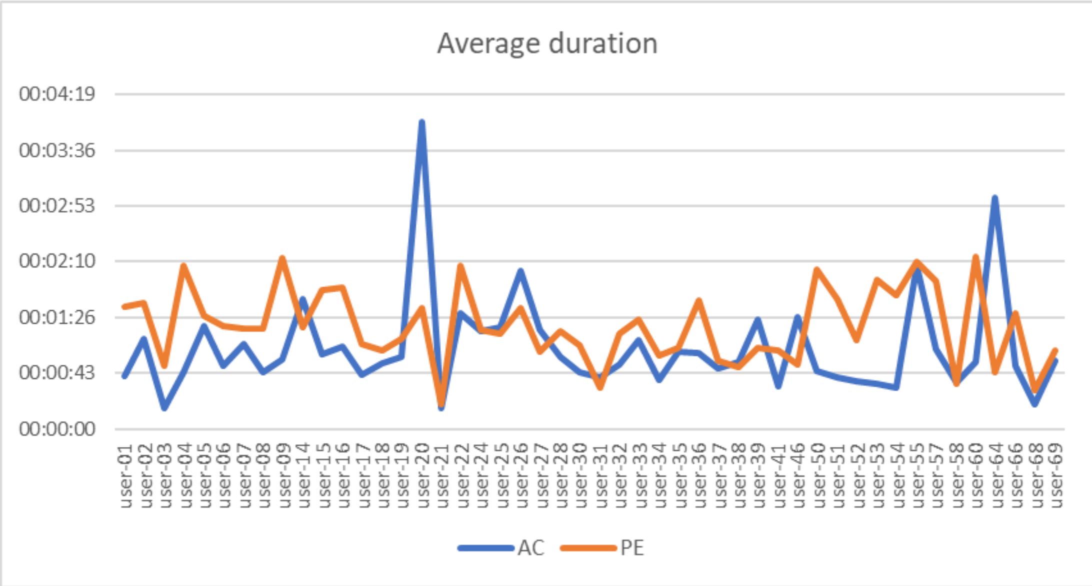


Figure 81 – Average editing time per user and mode.

Knowing that PE mode is more time-consuming than AC mode, it is not surprising to observe higher average editing times for PE. The only exceptions to this are outlying cases of longer editing times in AC mode, by a few users. Zooming in on the user with the longest average duration in a segment (nearly 4 minutes), one may conclude that he kept his segments open for a very long time: he only edited 4 segments in the 15-minute AC session, but he kept the segments open for a long time, up to 6 minutes in a specific case. But even disregarding this unusual case, one can see that AC mode is more heterogeneous than PE in terms of the average time dedicated by each user to each segment edited.

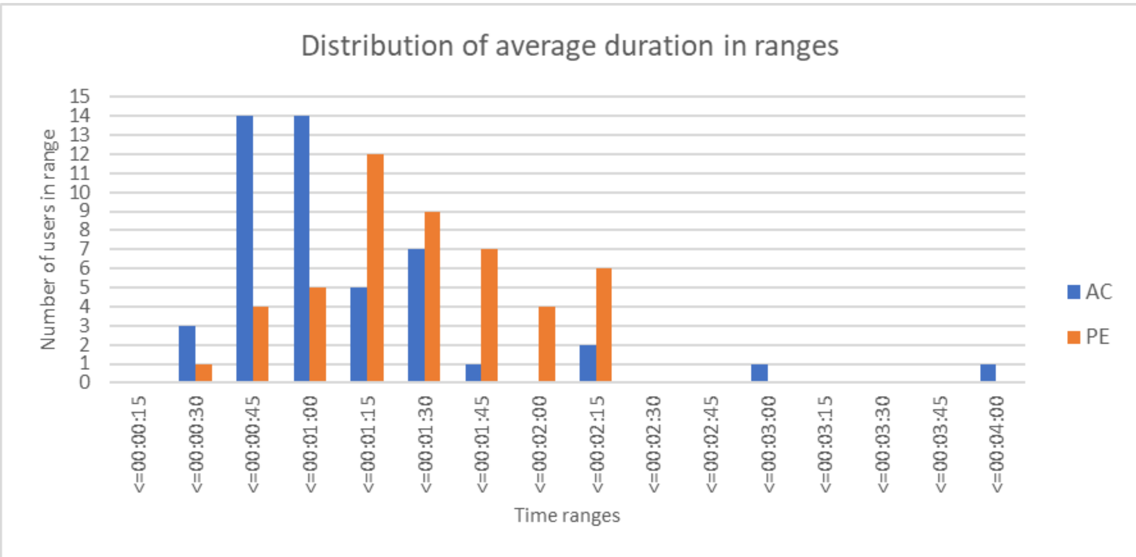


Figure 82 – Distribution of average editing times per user.

The chart above shows the number of users that present average editing times (duration) by time ranges. The times that identify each interval in the chart are the top limits of each range. For example, the columns in the category “<=00:00:45” include all users that have had edited a segment, on average, during a period of between 30 seconds and 45 seconds.

The chart shows that average editing times are more consistent or homogeneous in PE mode than in AC mode. In PE mode, most users edit each segment, on average, between 00:01:15 and around 00:02:15, and there are no users above that average editing time. In AC mode, averages are more varied, with most users showing low averages, below 00:01:00, but with several cases above that, all the way to above 00:03:45.

The first edit score to be analysed in this section is “Edits”, i.e. the number PEC identified as the minimum of editing actions necessary to transform the MT hypothesis into the edited result. The next chart shows the total number of edits in each mode per user.

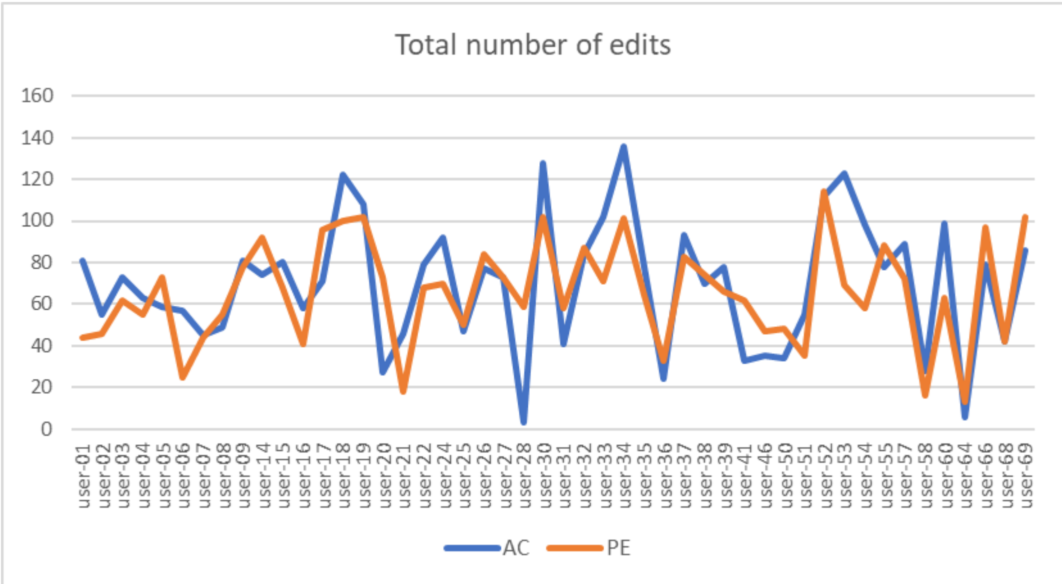


Figure 83 – Total no. edits per user and mode.

This chart shows a high dispersion in terms of total number of edits done by each user. Most users have done between 30 and 80 total edits, but some have gone up to 140 edits, and some stayed below 20 edits – the latter are users that opened several segments but did not edit them.

It is important to note that, in this case, the average is totally comparable between the two modes: the number of edits is measured in the same number of lines, since only confirmed segments are included. The chart below shows the average number of edits in each mode.

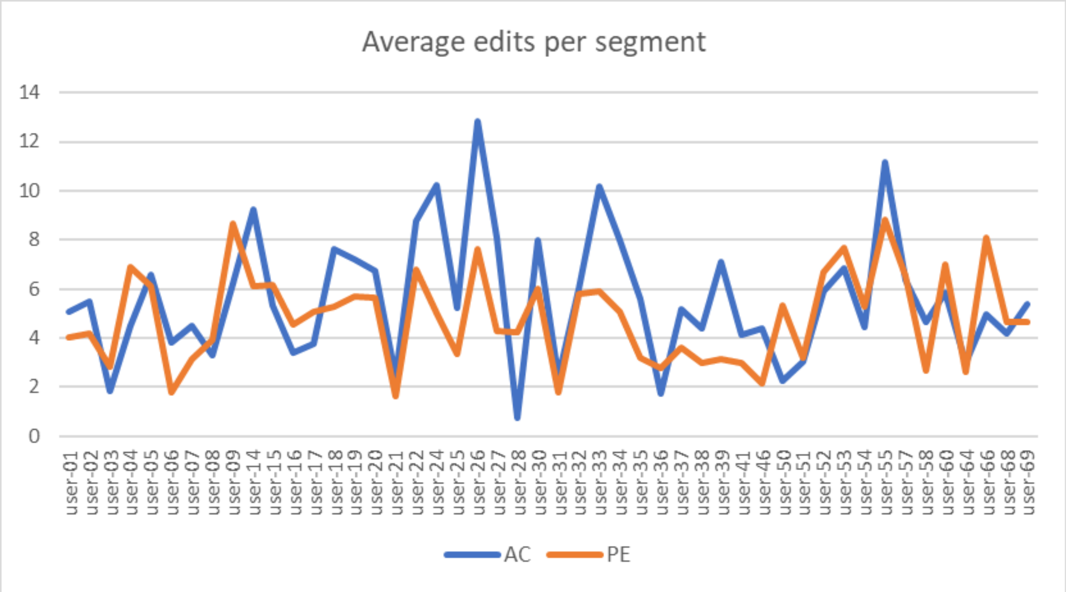


Figure 84 – Average number of edits in AC and PE modes.

The most striking feature highlighted by this chart is the fact that AC mode seems to have been associated with more edits per segment, on average, than PE mode. In AC mode, most users have done 6 edits per segment, whereas this value was 5 in PE mode. The number of edits per segment in AC mode is more varied, extending from 1 to 13 edits per segment, whereas in PE mode it goes from 2 to 9. So, it seems that the two editing modes are different when it comes to intensity of editing.

The next analysis is based on TER, which adds the length of the edited sentence as a factor for the calculation of edit effort. Furthermore, for TER, being a percentage, total values are not as relevant as averages. The global chart below, representing these results, portrays a very dispersed set of data, difficult to interpret. For example, although it seems that there are more users with average TER scores above the 25% editing threshold (the green line), the impact of this is not very clear. So, another way to visualise this data is required.

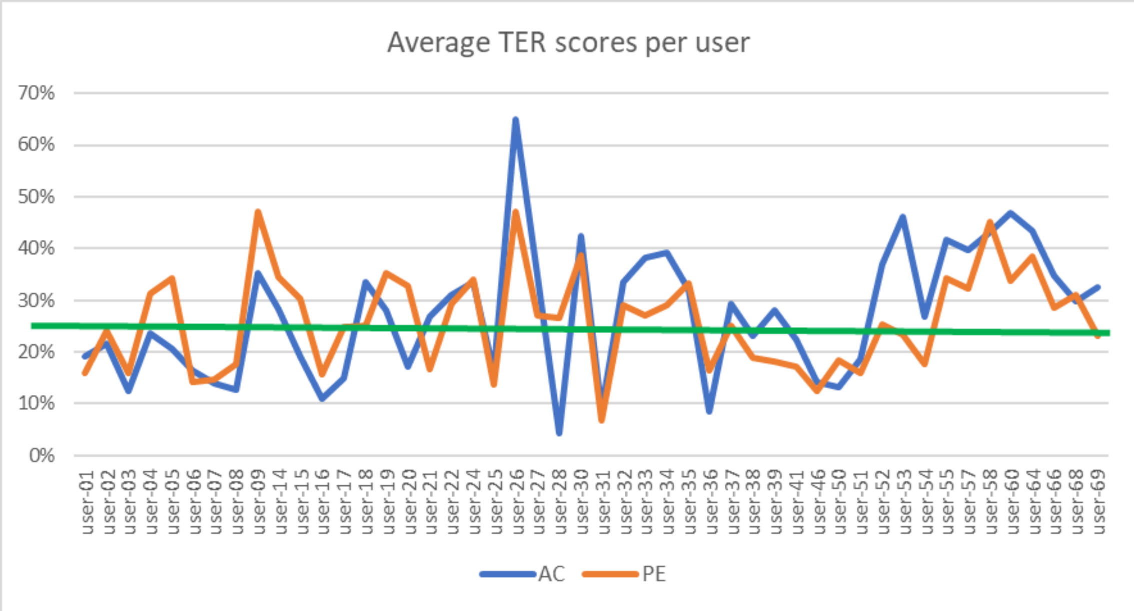


Figure 85 – Average TER scores per user in both modes.

A distribution chart like the one below shows more relevant details of this data.

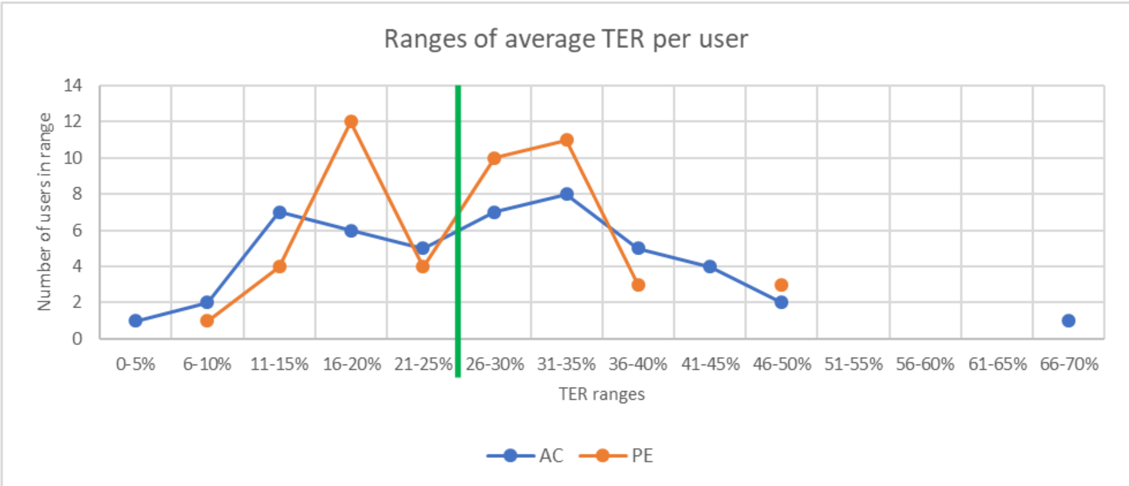


Figure 86 – Distribution of average TER scores in AC and PE modes.

Average TER scores per users are differently distributed in each editing mode. In PE mode, most of the users show average percentages of words edited in each segment that range from 16% to 35%. There are only 4 cases with average TER scores above 40%, in the 46-50% range. In AC, there are more cases in the extremes, below 11%, and above 36%, with seven cases above 40%, one being above 65%. Besides, the number of cases in the highest TER scores are also different for both editing modes. There are 12 cases in PE mode at the top of the distribution, with values between 16% and 20% of the words in the reference sentences edited, while in AC mode, there is a fairly similar number of cases (between 5 and 8) in three ranges, with average TER scores between 11% and 35% of the words edited. The number of users with average TER scores above and below the threshold can be seen in the chart above (to the right and the left of the vertical green line), but it is better explained in a table.

	AC	PE	Total	%
TER=<25%	21	21	42	44%
TER>25%	27	27	54	56%
Total	**48**	**48**	**96**	

Table 25 – Users with average TER scores below/above threshold.

The number of users with average TER scores above 25% is always higher than the number of users who, on average, edited less than 25% of the words in the reference segments. The proportion is close to 60% of the cases. Interestingly, in all levels, there is an even distribution between the two editing modes.

These analyses of distribution of average number of edits and TER seem to indicate that PE mode limits or conditions users in terms of the extent of editing that they do. The comparison between the two modes will be completed in the next section, 6.3.9.

Speed is the average time each edit implies. Since the duration and number of edits per segment is very scattered, the average speed per user and mode is more interesting than its total.

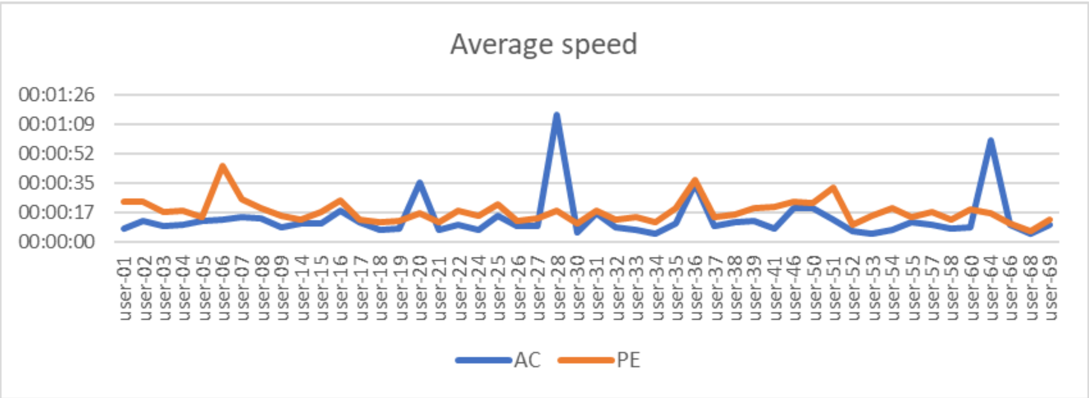


Figure 87 – Average speed per user and mode.

This chart shows a fair concentration of users with speeds between 00:00:05 and 00:00:35. The users with the fastest and slowest sessions registered them in AC mode. The highest point represents the slowest user, with more than one minute per edit, whereas the fastest ones only take on average 5 seconds per edit (the lowest points in the chart). But the distribution of these average speeds is, again, the best way to identify differences in this feature of user performance.

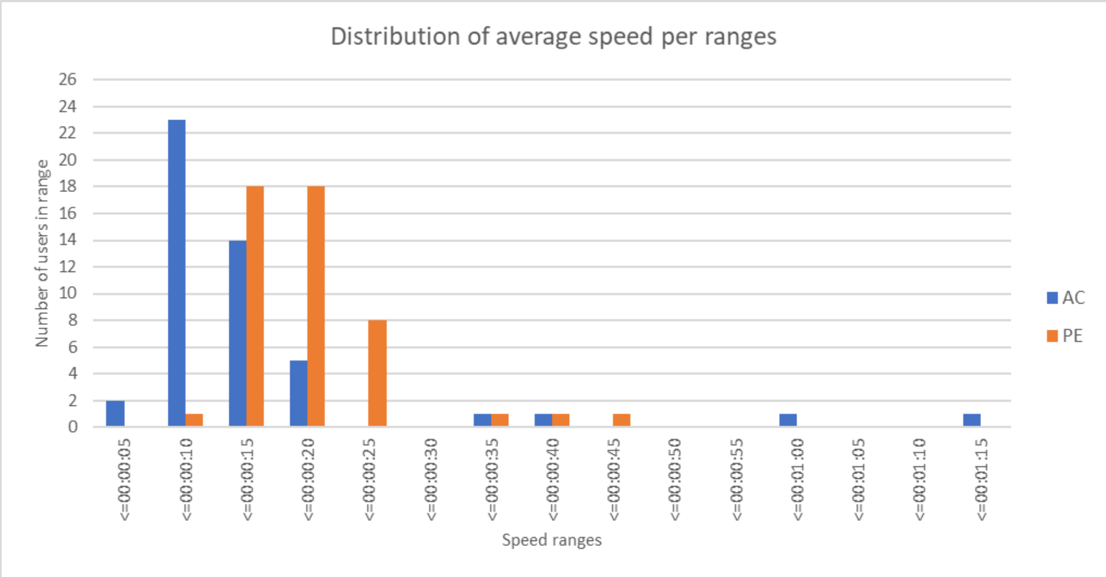


Figure 88 – Distribution of average speed per user in ranges.

Before analysing the data, it is important to note that in this chart the time interval is again represented by its highest value. So, all cases in the 10 seconds column actually represent values between 5 and 10 seconds.

PE mode scores in terms of speed (seconds per edit) seem again to be more homogeneous, as they fall within a shorter range. Virtually all users show average speeds between 00:00:10 and 00:00:25 in this mode, most being below 00:00:20. The longest average speed in PE mode is one case of 00:00:45 seconds. In AC mode, most cases are below 00:00:10, but they extend all the way from 00:00:05 to 00:01:15. So, it seems that PE mode, although not as fast as AC mode, is not associated with such a wide spread range of edits. But the next section focuses specifically on the comparison of AC and PE modes.

6.3.9. Description of main variables – Modes

The two editing modes provided by HandyCAT have been the main terms of comparison for the previous analyses of the main variables. So, it might seem that the description of these modes has been completed. However, there are still some details worth looking at.

Before we proceed, it is important to recall that there was a difference in time allocated to the two modes: 43% of the time was attributed to AC mode against 57% to PE mode. The description of results in this section always puts this difference in perspective. The first paragraphs below include a global analysis of the positioning of the two modes according to the main measures (number of segments, duration, number of edits, TER and speed).

The number of segments edited in PE mode is higher than in AC mode, as the following table shows. The ratio between the two modes in this measure is, however, lower than the one initially setup. This means that the number of segments that users have edited in AC mode is higher than what could be expected in terms of time allocated, whereas in PE mode was lower.

	No. edited segments	Percentage
AC	1045	49%
PE	1103	51%
Total	**2148**	

Table 26 – Number of edited segments per mode.

The number of closed and re-edited segments does not follow the initial proportion, but it actually inverts it:

	No. closed segments	Percentage
AC	74	52%
PE	69	48%
Total	**143**	

Table 27 – Number of re-edited segments per mode.

In this case, AC mode shows a higher number of re-edited segments, which is a relevant result, in view of the lower time allocated to this mode. So, it seems that users tend not to re-edit so much when they use a more intrusive method like PE mode.

In terms of total duration, which is the sum of all editing times by all users in all sessions, the two modes are closer to the proportion that was initially set.

	Total duration	Percentage	Average duration
AC	09:18:51	40%	00:00:32
PE	14:13:13	60%	00:00:46
Total	**23:32:04**		

Table 28 – Total duration per mode.

This means that there were no major imbalances created by some users dedicating more time editing a text in AC mode than in PE mode. The average values confirm that in AC mode users edit segments for a shorter time than in PE mode.

Although in both modes the average number of edits is around 3 per segment, the detailed results for this measure show that AC mode is associated with more editing work. This may be seen in the total number of edits, as well as in their unrounded averages:

	Total no. edits	Percentage		Avg. edits
AC	3351	52%		3.21
PE	3141	48%		2.85
Total	**6492**			

Table 29 – Total and average edits per mode.

This result is confirmed by the average TER scores. The results below show that, on average, users edited a percentage of words per sentence in AC mode that goes above 25%, the proposed threshold that divides translating from editing. In PE mode, average TER scores stay below that threshold.

	Average TER score
AC	26%
PE	24%

Table 30 – Average TER scores per mode.

These total averages are very close to the threshold, but they summarise very different realities. But note should be taken of a specific feature of this threshold. The threshold restricts the TER scores below to the maximum 25%. However, it does not define an upper cap for scores above it. So, in the level above the threshold, there are TER scores from 26% to as high as 171%, but values above that are also accepted. So, the number of segments with TER scores below the 25% threshold is constrained, but the same is not true for those above it. Still, 62% of the segments are below the editing threshold, in both modes, and 38% are above the threshold.

No. of segments	25% or less	above 25%	Total		25% or less	above 25%
AC	638	407	1045		61%	39%
PE	695	408	1103		63%	37%
Total	**1333**	**815**	**2148**		**62%**	**38%**

Table 31 – Number of segments below and above the editing threshold.

The differences between the modes are not very marked in relation to this feature. Still, one needs to be cautious in drawing conclusions from averages taken from data with wide variations. Let us look at these numbers with more detail.

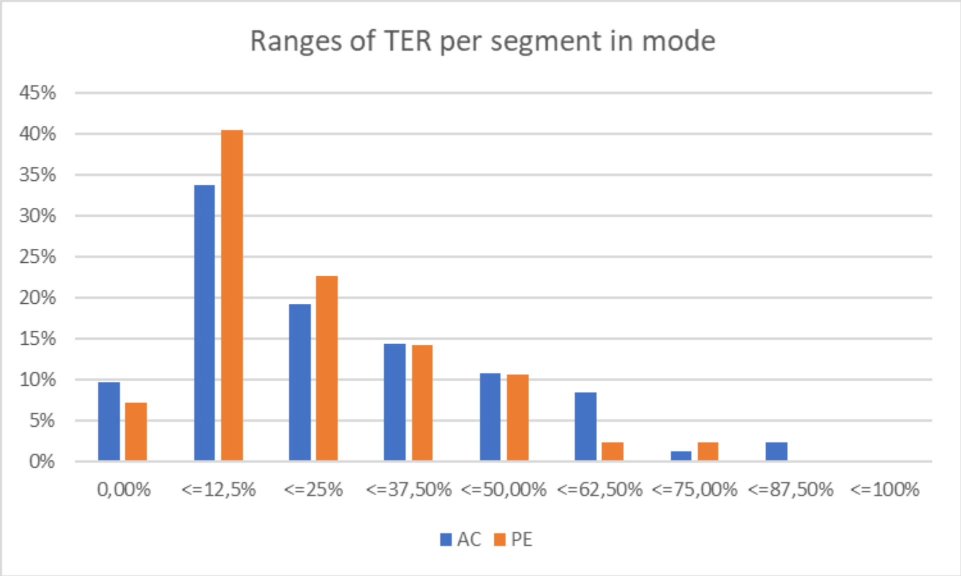


Figure 89 – Distribution of average TER scores per mode in ranges.

Average TER scores of segments organized by ranges shows a different behaviour for both modes. The number of segments that show 0% editing scores is higher in the AC sessions. The same happens at the other extreme: in ranges above 62.50%, AC segments are more common that PE mode segments. PE mode segments are more concentrated in ranges below 25%.

Finally, in terms of average speed, PE mode falls behind AC mode by a large distance, which seems to be a clear indication that this mode requires more time per edit.

	Average speed
AC	00:00:10
PE	00:00:16

Table 32 – Average speed per mode.

Besides this presentation of global results for the two modes, there are two sets of work sessions that should be analysed in comparison. The first set concerns sessions C and D, in which all users edited the same text, first in AC mode and then in PE mode, for 5 minutes in each session. The second set is made up of sessions B and D, both sessions in which PE mode was used, although on different texts.

Comparing two sessions with the same text in the two editing modes

Let us begin with the two last and shorter sessions. A few notes on these two sessions, in which users worked on the same texts with the two editing modes:

- Although only a short time was allocated for these sessions, the fact that all 48 users completed both sessions produced a fair amount of data;
- Since both modes were allocated the same time and the same texts, the two editing modes can be compared directly, without any data normalization;
- The effect created by users always applying AC mode before PE mode cannot be measured. The same sequence (AC before PE) was adopted in the two pairs of sessions (A and B, and then C and D). This sequence may result unfavourably for PE mode, due to it constraining somehow the way translators work, especially when compared with AC mode. Besides, PE mode may have given users a sense of experimentalism, eventually causing some loss of interest over the results. In an opposite view, it might have favoured PE mode, since users were repeating and reconsidering edits that they had applied before in AC mode. However, there is no way to account for these effects.

In sessions C and D, each user edited the same text for a total 10 minutes, 5 minutes in each mode. Globally, users edited a total of 854 segments from the 3 texts. Of this total, 425 segments were edited in AC mode and 429 in PE mode. So, at this level, both modes were similarly balanced.

Total duration of the two sessions was 07:16:24, divided by the two modes in near halves of 03:31:59 (AC mode) and 03:44:25 (PE mode). This represents an average duration per segment of 30 segments for AC mode and 31 seconds for PE mode. Again, there is no major difference between the two modes in this measure.

For edit scores, things are not so balanced between the two modes, as the following tables show. The first one shows how AC mode presents a higher number of edits, globally and on average per segment.

	Total no. edits	Percentage		Average no. edits
AC	1313	54%		3.09
PE	1132	46%		2.64
Total	**2445**			

Table 33 – Total and average number of edits per mode in sessions C and D.

TER score averages confirm these relative positions, since AC mode is the method associated with more edited words per total words in each segment. In these sessions, average TER scores were lower than the global average for both modes.

	Average TER score
AC	24.37%
PE	21.41%

Table 34 – Average TER scores per mode in sessions C and D.

So, this would appear to confirm that AC mode is associated with more edits, even when session times and texts are repeated. It is important to note that there are no clues to identify the reasons for a smaller number of edits in the same texts in PE mode. Even though this confirms a tendency observed in the analyses of the other main variables, there may be several reasons for this. It is possible that users looked at the texts for the second time and decided that some of the edits they had done in AC mode were superfluous, just because they had a chance to rethink them, or they simply could not make the same edits in the same time with the PE mode method. In effect, the next table shows that PE mode is again associated with more time per edit.

	Average speed
AC	00:00:10
PE	00:00:12

Table 35 – Average speed per mode in sessions C and D.

These data might deserve a more detailed analysis, but the purpose in collecting data in these two sessions was to have comparable data between the two modes, all other variables being the same.

Comparing two sessions in PE mode, with different texts and different times

The first time that users applied PE mode, in sessions B, they edited a text for 15 minutes. This was the longest session and it allowed all users to test and apply the PE mode with some detail. Then they edited a different text for the second time, now in PE mode for 5 minutes, in session D. So, they had two sessions in PE mode in a total of 20 minutes, the first session taking up 75% of that time and the second session (editing a text they already knew) taking 25% of the total PE mode time. The results of the comparison of these two modes may be summed up briefly. The main purpose of this analysis is to see whether there has been any gain from the first to the second time they used PE mode.

Users edited 674 segments in total in session B, and 429 in session D. The number of segments edited in session D represents a higher percentage of the total number of segments than the percentage of allocated time. This means that users edited, proportionally, more segments the second time they used PE mode.

	Total segments	Percentage
Session B	674	61%
Session D	429	39%
Total	**1103**	

Table 36 – Total segments edited sessions B and D.

The total and average durations are described by the table below, and these also show an evolution from one session to the next.

	Total duration	Percentage		Average duration
Session B	10:28:48	74%		00:00:56
Session D	03:44:25	26%		00:00:31
Total	**14:13:13**			

Table 37 – Total and average duration in sessions B and D.

Users logged more editing time in session B than in session D, but the result was in the same proportion as allocated time. So, comparatively, they used the same percentage of available time in both sessions. However, average editing time per segment was much shorter in session D (almost half a minute, compared to nearly one minute in session B). So, users were dedicating less time to editing each segment in the second time they applied PE mode.

The total number of edits in session B was naturally higher than in session D, but the proportion is not equal to the different times that were allocated to them. Again, session D is associated with a higher number of edits, if one considers the proportion of time allocated. The average number of edits per segment in session B was higher than in session D, but the difference is not as high as to obscure the gains in the second session.

	Total no. edits	Percentage		Average no. edits
Session B	2009	64%		2.98
Session D	1132	36%		2.64
Total	**3141**			

Table 38 – Total and average number of edits in sessions B and D.

Average TER scores are also different between both sessions: 25.63% of the words on average in each sentence were edited in session B, and 21.41% in session D.

	Average TER
Session B	25.63%
Session D	21.41%

Table 39 – Average TER scores in sessions B and D.

Speed combines all these measures and gives more visibility to an eventual evolution. The result confirms a clear evolution in terms of the time per edit applied by each user, in the two sessions in which PE mode was used, with users taking only 12 segments per edit in the second session, when their average was 19 seconds per edit in the first session.

	Average speed
Session B	00:00:19
Session D	00:00:12

Table 40 – Average speed in sessions B and D.

All these analyses seem to indicate that, although PE mode is admittedly an intrusive method of work (some participants even wrote in the questionnaires that such a method should never be used in production), an evolution in productivity may be observed with familiarity with the method, even in such a short experiment as the one performed during the workshop. Still, this would need to be further tested. One needs to be reminded that other factors may weigh on these decisions, such as the fact that users are working with different texts (the text used in the second session might require less editing) and that they already knew the text used in the second session. However, since all texts were used in all sessions, and all other variables have been analysed separately, this evolution seems reliable.

Apart from the analyses on duration, which involve all events logged, all calculations made so far only focus on confirmed segments. The next section focuses on fundamental events in PE mode that have not been analysed yet: editing actions.

6.3.10. Description of main variables – Actions

PE mode was designed and integrated into HandyCAT to record and allow posterior assessment of the editing actions chosen by users. Each time a user chose a specific action from the context menus in PE mode, HandyCAT recorded a new event. However, these events are very different from the rest of the data collected, and thus they require separate processing. Let us start by discussing the difference between the actions performed by users in HandyCAT and the edits identified by PEC.

Overlap and mismatch between editing actions and edit scores

HandyCAT registers the editing actions chosen by users in the context menus, but this information was not passed on to PEC. So, PEC estimates edit scores (number and type of edits, and TER) by comparing the resulting sentences against their references. Without the information from HandyCAT, when a user chose a Delete action, PEC may have identified a Replace instead. And if the user used the contextual menu in HandyCAT to insert a two-word phrase, even though PEC may have correctly identified the insertion, it still registered 2 edits instead of 1. The distance between what is chosen by the user and what PEC estimates depends not only on what users actually do when they choose each editing option in HandyCAT, but also on PEC's processing (analysed in section 6.3.3). The disparity of the estimation of edits is partly controlled, because, for each event, the reference is the result of the previous action, not the initial MT hypothesis. This reduces the extent of edits recorded in each event and increases the probability of PEC identifying correctly the action chosen by the user.

Ideally, if all editing actions had been accurately chosen and applied by users during the workshop, and if PEC's edit score estimates were adjusted to this process, for each editing action, there would be only one edit identified, with varied lengths, according to the number of words edited in the same action. However, this coincidence was not always achieved. But how good was PEC in identifying the editing actions?

	No. actions	No. edits		Difference	% error
PE.delete	1122	1553		431	28%
PE.insert	570	847		277	33%
PE.move	265	444		179	40%
PE.replace	1470	1658		188	11%
Total	**3427**	**4502**		**1075**	**24%**

Table 41 – Difference between editing actions and edits.

The table above shows the global differences for all editing actions, between the number of editing action selected by the user (second column) and the number of edits identified by PEC (third column). PEC always identified more edits than the ones which had been chosen by users, 1075 more edits in total, an error which corresponds to 24% of the total edits that PEC identified. Move was the editing action more often wrongly identified: 40% of the times PEC identified it, it had not been chosen by users. Next, came Insert, with 33% of errors, and Delete, with 28% of errors in identification. Replace was identified wrongly only 188 times, which corresponds to 11% of the times PEC identified this edit. The next table and chart detail this, per editing action.

		Edits by PEC			
Actions	**No. actions**	**Delete**	**Insert**	**Move**	**Replace**
PE.delete	1122	1280	0	0	273
PE.insert	570	22	650	1	174
PE.move	265	107	46	184	107
PE.replace	1470	118	342	1	1197
Total	**3427**	**1527**	**1038**	**186**	**1751**

Table 42 – Number of edits identified by PEC per editing action.

The majority of edits identified by PEC are correctly associated with the respective action (green cells). Each column in the table above presents the numbers of each type of edits PEC identified; in each line, the table shows the numbers of edits when each action was selected by the user.

For example, users chose 1122 times the Delete action in HandyCAT (line "PE.delete"). In these same events, or lines in the table, PEC identified 1280 deletions, no insertions or movements, and 273 replacements. The 1280 deletions PEC identified are more than the number of times Delete was chosen in HandyCAT. The same goes for the other editing actions. So, PEC was never correct 100% of the times it identified each editing action. The following chart is a visual representation of this data.

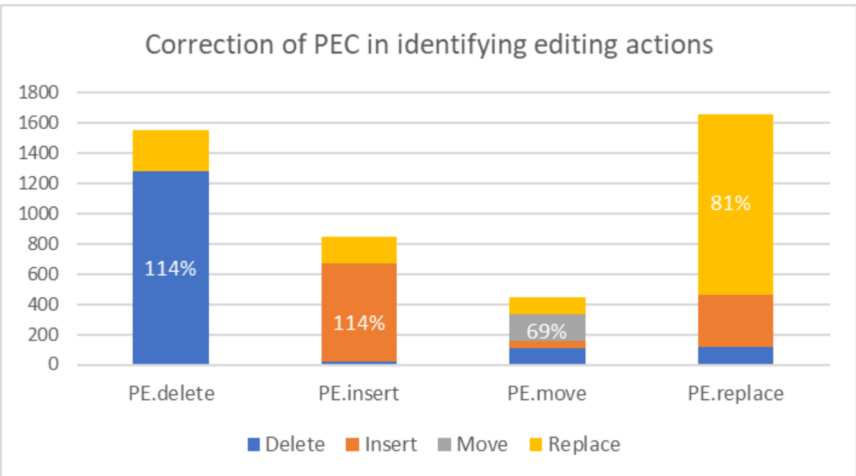


Figure 90 – How PEC classified editing actions.

The bars in the chart represent the editing actions chosen by the users and the colours and values represent the edits identified by PEC. For example, the first bar represents the number of times Delete was chosen by users (coded as "PE.delete"). For these, PEC associated a number of deletions that is 114% the number of actions. Above that, it identified for the same actions a few more (273) replacements. For the editing action Insert, the result is similar: PEC identified more insertions than those that had been chosen by users, plus some replacements and very few deletions. Move does not have a high frequency, not as an editing action chosen by users, nor as an edit identified by PEC. Besides, in 69% of the cases, it was correctly identified. However, all other edits were identified in segments where users had chosen "PE.move". Finally, Replace was correctly identified by PEC 81% of the times, but this does not mean that when Replace was chosen, PEC did not identify other edits, namely insertions and deletions.

It is important to remember that these events are not finished sentences, in which all edits are combined, but events which only register what happens between the moment a user chooses an action from the contextual menu in PE mode and the moment he chooses a different action in the same menu, or closes the segment.

From these results, it seems that movement is the editing action most hard to identify by a metric such as TER. Deletions and insertions are easiest, with replacements close by. For virtually all the editing actions, replacement (in yellow in the chart above) was the second edit most often associated with each action. The exception is obviously replacement, in which case the second edit more often wrongly associated with the action was insertion.

A more detailed analysis of the correlation between editing actions and edits identified by PEC could be made, using evaluation measures such as "precision and recall". However, the purpose of this section is not to evaluate PEC's performance, as it was already established that it did not fully comply with the purposes of identifying the editing actions. Next, HandyCAT's records of editing actions are described, disregarding PEC's results.

Number of events associated with editing actions

Events associated with editing actions selected in HandyCAT were only recorded in sessions B and D (analysed above), and there are 3427 of these, representing 40% of the total 8565 events registered in the activity logs. Their distribution per editing action is as follows:

	Number of events	Percentage		Sess. B	Sess. D	% Sess. D
PE.delete	1122	33%		748	374	50%
PE.insert	570	17%		383	187	49%
PE.move	265	8%		216	49	23%
PE.replace	1470	43%		958	512	53%
Total	**3427**			**2305**	**1122**	49%

Table 43 – Number of total editing actions in activity logs.

As reported by users in the final Questionnaire (Q2.4 – see Figure 28 above), Replace was the action used most often (with more than 40% of the cases), closely followed by Delete, then Insert, and finally Move.

The proportion of actions between the two sessions in PE mode should be 33%, as session D took 5 minutes and session B 15 minutes. However, none of the actions showed that proportion (last column in the table): most actions were used in session D about half of the times they had been used in session B, which may reveal that users found these actions useful in the new contexts of the last session. However, Move was used only one quarter of the times, which may mean that users did not enjoy using this action. Or perhaps it simply means that moving words around was not so relevant in the small portion of text users edited in session D.

In terms of total duration of these events, the order of the editing actions is the same as in the total number of these events, but the percentages are not the same.

	Total duration	Percentage
PE.delete	02:27:12	22%
PE.insert	02:09:49	19%
PE.move	00:57:30	8%
PE.replace	05:46:06	51%
Total	**11:20:37**	

Table 44 – Total duration per editing action.

Replace is again the editing action used most often, but it took more than half of the total editing time registered in total in these events. Delete lost 11% in terms of total time, which means that it is a faster action than Replace. Insert and Move have taken more or less the same percentage of time that they did in percentage of events.

Duration and edit scores associated with editing actions

Average durations should be analysed not only in total times, but also by discriminating the two PE sessions.

	Average duration		
	Session B	Session D	Total
PE.delete	00:00:09	00:00:06	00:00:08
PE.insert	00:00:14	00:00:13	00:00:14
PE.move	00:00:14	00:00:09	00:00:13
PE.replace	00:00:16	00:00:11	00:00:14

Table 45 – Average duration per editing action and session.

One may see that all editing actions took less time, on average, per event, in session D than in Session B. On average, Replace and Insert were the actions with the longest duration, Move was next, and Delete was the action with the shortest average editing time. The fact that Replace and Insert imply inputting text is the natural reason for this longest delay (these are the two actions related to content creation – see section 5.3.1).

Edit scores were estimated by PEC by aligning references and results, and then measuring differences, as described in section 6.3.3 above. The analysis of the relation between HandyCAT events (the editing actions chosen by users) and the edits identified by PEC was done above. So, let us just look at the results presented by PEC as a measure of the editing work done by each user in each of these events.

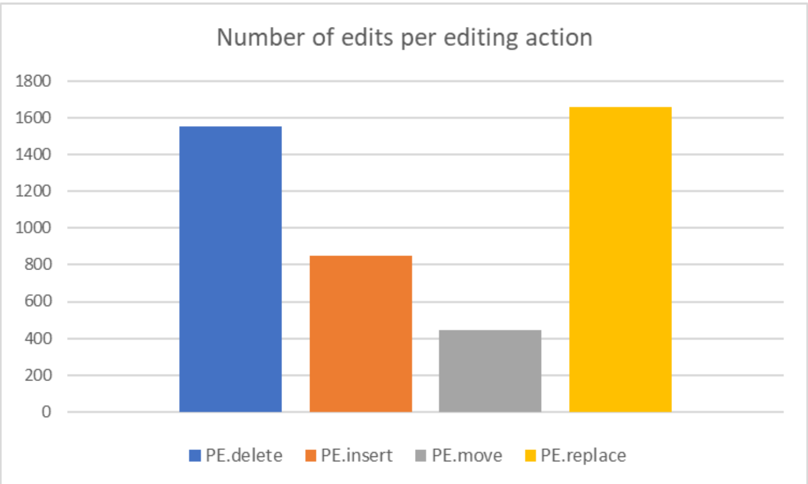


Figure 91 – Number of edits per editing action.

The chart above shows how, in spite of the fact that the same order is maintained, Delete almost has the same number of edits as Replace. Insert has half those numbers, and Move only a quarter.

	Average no. of edits		
	Session B	Session D	Total
PE.delete	1.37	1.41	1.38
PE.insert	1.45	1.55	1.49
PE.move	1.71	1.53	1.68
PE.replace	1.11	1.16	1.13

Table 46 – Average number of edits (identified by PEC) per editing action.

The data for the average number of edits does not confirm the relative positions of the editing actions observed so far. Move seems to be the editing action that produces more edits per event, as identified by PEC, especially in the first and longer session, session B. Then, Insert, followed by Delete, and only then Replace.

It would seem that by choosing a movement, users are creating a more complex result that most often is identified as two edits (or two words edited). This action, moving words around, is very often identified as deletion followed by insertion, but, as may be seen above (section 6.3.3), this is not always the case. In fact, it is even the action that is able to record into one single edit changes to more than one word. Deletion and insertion also produce, on average, results that are above one edit, which means that users apply them to more than one word. Only Replace seems to be closer to the identification of one action chosen and one word edited.

Besides, this data confirms that in session D users produced more edits per event, for all actions except for Move.

Minimum and maximum TER scores

The analysis of maximum and minimum TER scores for each editing action shows that the range of edits in these events is quite wide:

	Minimum TER	TER=0%	Maximum TER	TER>100%
PE.delete	0%	59	100%	5
PE.insert	0%	20	167%	9
PE.move	0%	9	100%	2
PE.replace	0%	264	300%	8
		352		**24**

Table 47 – Minimum and maximum TER scores per editing action.

It should be remembered that all lines concerning editing actions are associated with events that start with the user choosing one of the editing actions in HandyCAT's contextual menu. If users used these options correctly, and TER estimated made by PEC correctly associated edits at the phrase level, there should be only one edit per event/line. As mentioned in 6.3.3, TER counts each edit per word for all actions, even if contiguous words are edited with the same action, except for Move, which may identify shifts of two words as one single edit.

There are, however, several lines in the log table in which PEC identified a TER score of 0%, even though the user chose and applied a specific action to the content of the segment. There are cases like this for all editing actions, but most are associated with users choosing Replace, Delete coming as a distant second. A close analysis of these cases shows that one of the reasons is that users have chosen this editing action, but then they did not make any changes and simply closed the editing window. In other cases, the editing actions were chosen to make edits such as changing capitalisation, deleting and inserting spaces and other changes which are not captured by PEC's method of aligning and comparing references and results.

On the other extreme, there are events that produced TER scores of 100% or more. These cases are all in short segments (with no more than 7 words), and, in these, users applied editing actions to totally change the content of the segment. On different occasions, they deleted the whole content of the segment, but they also moved words around and inserted them together, thus affecting the number of words in the segment. There are two cases of Replace with a TER score of 300%. This was caused by one user who replaced one word with three in two similar segments of the same text.

Although it could be tempting to normalise these numbers, at least by reducing the scores to 100% (as TERp does), it is important to capture and be aware of these behaviours. Besides, these values appear at the level of the editing action events. So, some of these changes are intermediate, and they probably do not survive in the final version of the edited sentence.

The analysis of average TER scores per editing action reveals interesting results.

	Average TER scores		
	Session B	**Session D**	**Average**
PE.delete	11%	11%	11%
PE.insert	14%	15%	14%
PE.move	19%	11%	17%
PE.replace	10%	10%	10%

Table 48 – Average TER scores per editing action.

Average TER scores produced by each editing action are always below 20%. Move is the action that is associated with a higher percentage of edited words in each sentence. For this action, TER scores are higher in Session B than in session D. Insert is the action with the second highest average TER score, and this action shows a higher score in session D. Delete and Replace are the actions with the lowest TER score and they maintain their scores in both sessions, of 11% and 10%, respectively.

Average speed shows that all editing actions were applied faster in session D.

	Average speed		
	Session B	**Session D**	**Average**
PE.delete	00:00:06	00:00:04	00:00:06
PE.insert	00:00:10	00:00:08	00:00:09
PE.move	00:00:08	00:00:06	00:00:08
PE.replace	00:00:14	00:00:10	00:00:13

Table 49 – Average speed per editing action.

Even at such a detailed view (how fast was each edit produced, when users chose only one editing action in PE mode's contextual menu), there was a gain in the last session, compared to the first time that users applied this method. Delete is the fastest action, taking only 6 seconds on average to make an edit, and Replace is the slowest, with an average of 13 seconds. Move is faster than Insert, but only for a small margin. The fact that both Insert and Replace imply typing words is the most likely explanation for these results.

Another conclusion one can draw from this data is that, even without interface improvements, using Move (8 seconds) or Replace (13 seconds) is faster than deleting a word and then inserting it in a new position, or inserting a new one in the same position, a set of actions that takes on average 15 seconds in this editing mode.

6.4. Summary and global analysis of the results

After all the results were collected, several tools were applied to look for regularities and structures that allowed for a global vision of relations and correlations between input variables and results. Due to the wide variety and type of data, this was a hard task and often a frustrating one. Different statistical data methods were tested, adapted to the different types of data, but conclusions often contradicted each other, as methods revealed different perspectives of the same variable, or the relations identified were not strong enough and could not be confirmed in other perspectives.

These results were to be expected, in view of the experimental nature of the creation of the tests. Only after the data collected was analysed was there sufficient knowledge acquired to allow for a more integrated and consistent collection, which could result in solid conclusions. However, as explained at the beginning, the main typical correlations, associated with quality or effort, for example, would not be the measures of success of the project, but the amount of knowledge about the PE process, under the light of new concepts such as the four editing actions and the editing threshold, that could be elicited from a limited experiment with a novel interface.

The next section will present a collection of conclusions retrieved from the data analysis process. The subsections follow the presentation of the main variables from the activity logs: Texts, Segments, Users, Modes and Actions, although for some of these variables it was not possible to collect relevant results. Finally, the two most important result variables (TER and Speed) will also be discussed. The data collected from the Questionnaires during the workshop will be added as relevant to the global analysis.

These analyses were performed on R, a data analysis tool (R Core Team, 2013). Several data treatment operations were done at the beginning, from which different data tables resulted. The main data used in these analyses is organised by users and it combines the main variables from the workshop questionnaires with the main results variables. Appendix 10 includes this table of data before processing. From these variables, new features were extracted, such as averages and medians, minimum and maximum values, rankings, standard deviations and others. A combination of these features was used in the analyses described below.

6.4.1. Summary and analysis of input variable: Texts

The statistical tests that were applied to the texts as input variables did not allow for the identification and statistical validation of their most interesting features. This shows that the texts were not sufficiently different, or behaved in a specific way to reveal statistical relevance. Besides, the different amount of time that was allocated to each text has invalidated many of the analyses, even using methods that were not sensitive to this distortion. Let us examine the analyses of complexity of the texts.

In the second Questionnaire that users answered at the workshop, there were two questions related to their intuitive assessment of the complexity of the three texts, and how that related to the quality of the MT hypotheses that they were presented with for each text (Q2.1 and Q2.2). These answers might be compared with the dimensions of textual complexity measured from the results of the editing sessions. A simple measurement of "type/token ratio" was also estimated, using an online tool (Text Inspector). The next table sums up the results of all these different points of data related with text complexity.

		Questionnaires		Variables of result				
		Q2.1	Q2.2	Re-edited segments				
Texts	Type/token ratio	Very complex	Low quality	No. segments	No. repetitions	Avg.Duration	Avg.TER	Avg.Speed
B-Questionnaire	0,30	6	3	44	5	00:00:10	18%	00:00:12
C-Catalogue	0,35	14	9	37	10	00:00:11	31%	00:00:13
D-Manual	0,47	8	3	57	10	00:00:09	26%	00:00:14

Table 50 – Comparison of different measures for text complexity.

In both questions in the questionnaire, users mostly classified "Catalogue" as the text that was most complex and which presented the worst quality, "Manual" would occupy the second place and "Questionnaire" seemed to be the text that would present fewest problems. However, this does not coincide with most other measures – the only measure that keeps this order is TER. However, this is not statistically relevant enough to enable any conclusions in terms of correlation or other aspects. Type/token ratio classifies "Manual" as the most complex text, and this is the same order indicated by variables like number of re-edited segments, average duration and average speed.

One important conclusion from the analysis of the editing threshold at the text level concerned the global average TER score of 25%. As mentioned, a global analysis of a project composed by these three texts might disguise the fact that, what seemed to be a project that fell below the editing threshold, included one text that had an average editing score of 31%, above that threshold.

So, text-level effort analyses need to be studied with other parameters and other methods, if one wants to find stronger relations and reach stronger conclusions.

6.4.2. Summary and analysis of input variable: Segments

The segment-level was also not very fruitful in a statistical analysis either. Some of the reasons for this have to deal with the variety of factors that characterised this variable, each with a different behaviour. For example, segment length was not a constant factor, as the text with the longest segments was also the one with the shortest (if there were more texts, this factor would not have a major impact, but in one out of three texts, the impact is strong). Another factor was the number of segments edited in each text and mode: while in Questionnaire, one may find 40 segments edited in AC mode, in Catalogue in the same mode only 17 segments were edited. Again, this variable alone could be normalised and analysed separately. However, forcing the extraction of global conclusions would require an excessive manipulation of the data. So, the effort was concentrated on collecting the most important conclusions from the specific results described above and from looking at specific sets of segments.

The first conclusion taken from the results at the segment level is that average duration is always higher in PE mode, whereas the number of edits is higher in AC mode. However, this finding would need to be confirmed with more data.

As for speed and TER, the results were very different. Speed data per segment is very concentrated in a line of up to 45 seconds, with a major and difficult to analyse variation in that range, but it is a measure that is very sensitive to segments with a long reading time and a small number of edits. So, it seems to be a measure that is not very appropriate for the segment level.

Analyses of TER scores seem to be adjusted to the segment level, perhaps because this takes into account the segment length. A detailed analysis of the different ranges of TER scores in segments showed a wide range with interesting cases, that are worth further study. Let us look at the ranges that are above the suggested editing threshold of 25% edited words per segment.

Globally, 34% of the segments are above the editing threshold. In Catalogue, this number is 53% of the segments. In this sample, there were 5 segments (2 in Manual and 3 in Catalogue) that required on average more than 60% of editing, in both modes. The following table shows a summary of these segments. In Appendix 10, the full list of translations submitted by the workshop participants in these segments may be consulted.

Text	Segment	Mode	Avg TER	Sessions	Source	Reference	Result (sample)	TER of sample
Manual	17	AC	83%	A & C	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque-Perigo de choques.	80%
Manual	17	PE	52%	B	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque - perigo de choque.	60%
Manual	10	AC	69%	A & C	Product/Feature	Característica do produto	Produto/Característica	100%
Manual	10	PE	64%	B & D	Product/Feature	Característica do produto	Característica/ Produto	67%
Catalogue	5	AC	81%	A & C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies húmidas e secas	86%
Catalogue	5	PE	75%	B & D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza de superfícies Wet and Dry	57%
Catalogue	6	AC	61%	A & C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da mesa, laptop e teclado limpas e livres de pó e sujidade com os toalhetes de limpeza <tag> para superfícies húmidas e secas.	67%
Catalogue	6	PE	53%	B & D	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies da sua mesa, laptop e teclado limpas de sujidade e poeiras com os toalhetes de limpeza para superfícies secas e húmidas.	54%
Catalogue	17	AC	62%	A	Thoroughly clean telephone sets, keyboards and other hard surfaces with the <tag> Anti-Static Cleaning Cloths.	conjuntos completamente limpas de telefone, teclados e outras superfícies duras com a <tag> Anti-estático panos de limpeza.	Limpe telefones, teclados e outras superfícies rígidas com os toalhetes de limpeza anti-estática <tag>.	71%
Catalogue	17	PE	53%	B	Thoroughly clean telephone sets, keyboards and other hard surfaces with the <tag> Anti-Static Cleaning Cloths.	conjuntos completamente limpas de telefone, teclados e outras superfícies duras com a <tag> Anti-estático panos de limpeza.	Limpeza profunda de telefones, teclados e outras superfícies rígidas com os panos de limpeza Anti-estático .	53%

Table 51 – Segments with highest TER scores.

This table shows that, in this list of the five most edited segments, three are very short ones. Two of these are from Manual and one from Catalogue. So, the assumption that there is a correlation between sentence length and editing effort needs further testing, in view of the content that appears in each segment. For each pair of segments (each segment is presented in AC mode and in PE mode), the AC mode segment always has a higher TER score, sometimes with a major gap, as in the first case, with 83% in AC and 52% in PE. A sample of an edited result was chosen to illustrate each case in the table above. Sometimes, the results seem very similar between segments with different TER scores, but in short segments every edit makes a strong impact on TER. Moreover, these samples are as close as possible to the average, but when the average is 52% and there are only edited segments with either 60% TER score or 33%, there is no way to show a sample of the average score. All work sessions are represented in this list, from the longer sessions A and B, to the shortest sessions C and D.

6.4.3. Summary and analysis of input variable: Users

Data collected at the workshop questionnaires included interesting data on the characterisation of the users. Although this was not the focus of this study, there are a few results that are worth recalling.

There were several questions on the use of technologies, and most participants appeared comfortable with translation technologies and there were several experienced users in PE. One of the questions that did not have the expected response was the one on the reasons for using MT as an aid to translation work (Q1.8b). Still, it was interesting to realise that the answers presented could be grouped according to the four dimensions to estimate MT quality described in section 3.5.2:

- **Complexity of the ST:** the type of text (technical texts as the main type that calls for the support of MT), the number of words and repetitiveness of ST (long and repetitive texts are seen as a good reason to use MT);
- **Confidence of the MT system:** one user said that whenever the quality of the MT hypotheses justified it, he used MT;
- **Adequacy of the MT output:** a few users said that they used MT to look for specialised words, while others said that they never used MT to look for specialised vocabulary; a few said that they used MT to look for the translation of Named Entities, such as names of countries, or as an aid to the interpretation of the ST;
- **Fluency of the MT output:** the use of MT was also justified as a support to writing the TT, mainly when this was a foreign language.

Many more results could be commented on here, and different professional profiles could be composed of the users that participated in the workshop, but these analyses were not conclusive.

From the activity data collected at the workshop, the most interesting results may also be emphasised:

- Twenty one users reopened and re-edited 143 segments. Of these, three users re-edited more than 16 segments each, but the rest re-edited less than 10;
- Average TER scores for users range from 4% to 65% in AC mode; and from 7% to 47% in PE mode – there are three users with this top TER score in PE mode, and one user registered the highest TER score in both modes;
- 56% of all users show average TER scores above the 25% editing threshold;
- Most users stay within the 13 seconds/26 seconds per edit in PE mode, but the variation is much higher in AC mode.

TER and Speed are the result variables that show most interesting results. Techniques like Principal Component Analysis and clustering were tested to identify regularities among users' behaviours. However, it was considered that it was difficult to find a balance between the need to select and tune the adequate variables and the reasonability of the interpretation of the results that were obtained.

The best results from this analysis were obtained by simply crossing TER and Speed average results by users. This simple approach showed that users could be organised in four distinct groups. The tables with the data from ranking users according to average values for TER and Speed are included in Appendix 10. When these groups are visualised in a matrix, this is the result:

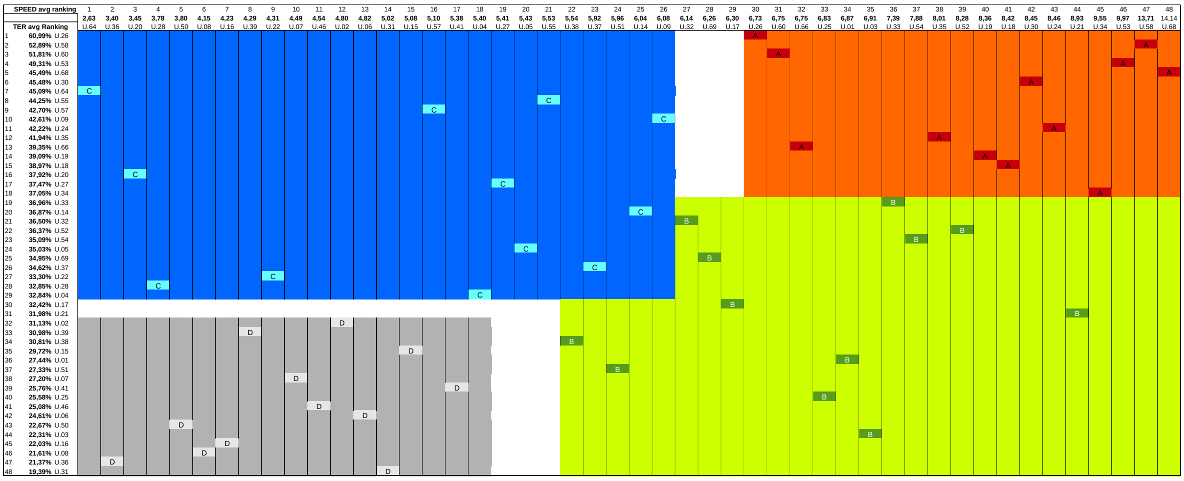


Figure 92 – Four groups of users by combination of TER and Speed results.

The values for TER and Speed averages create four very clear groups of users, separated by higher and lower averages for each of these variables. The composition of this table follows a very intuitive process:

- Rank all users vertically by their average TER scores, from highest at the top to lowest at the bottom;
- Rank all users horizontally by their average Speed scores, from slowest on the left to fastest on the right;
- Mark the cell in the matrix where the line and column with the values for each user meet;
- Mark four groups of the 12 users closest to each corner; check whether it is possible to draw vertical and horizontal lines separating each group.

If the values were too dispersed, it would not be possible to obtain a distribution so clear as in this case. So, this results in four groups, which may be identified as:

Group A: users with the highest scores both in TER and in Speed;

Group B: users with the highest scores in Speed, but lowest scores in TER;

Group C: users with highest TER scores, but lowest Speed;

Group D: users with low speed and low TER scores.

These groups have been formed naturally from the data, with very clear separation lines, and no overlaps, except for one small area in the middle, between groups C and B, but even in that area, members of each group are separated, with no members crossing over to the other area.

One can see that even inside each group, users concentrate in the top and rightmost areas, where the highest results are. For example, from the quarter of users in the D area, only four users are in the area that represents the lowest results for both variables, TER and Speed. If we divided each ranking in five sections, this area represented the lowest fifth for both variables.

To allow for this crossing, it was required that progression was in the same direction. So, the measure for Speed in this table was the inverse of the one which was used in the previous analyses: in this table, Speed is presented as edited words/minute.

These groups were confirmed using other methods. The data visualisation represented by the two box and whiskers charts below show the groups obtained by analysing the two result variables separately. One may see that the groups are cohesive and there are not many outliers in any of the representations. The data and explanation of these charts is included in Appendix 10.

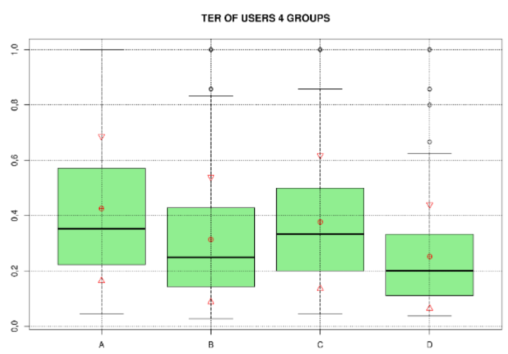


Table 52 – Four groups of users by TER.

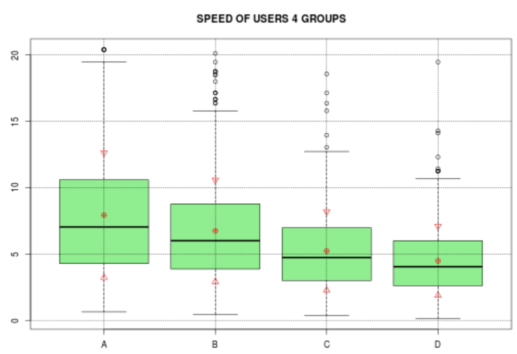


Table 53 – Four groups of users by Speed.

The four groups of users could have been further explored, namely by detailing their composition, in terms of the professional profiles and the answers that users in each group had given at the workshop. However, all tests revealed that it was difficult to establish relations or correlations. This was not considered as a shortcoming, since the characterisation of the translators was not one of the purposes of the study.

6.4.4. Summary and analysis of input variable: Mode

The two editing modes have been analysed quite thoroughly throughout the section on description of the results, and the statistical tests did not add more useful information. Here are some of the most relevant findings of the description of results.

- AC mode has a typically higher TER average score (26%) than PE mode (24%);
- In AC mode, there are more segments above the editing threshold (39%) than in PE mode (37%);
- AC mode allows for a shorter time per edit (10 seconds on average, against 16 seconds in PE mode);
- The two comparable sessions (C and D) confirmed higher TER scores, faster editing time and a higher number of edits in AC mode;
- There are more segments with zero edits in PE mode, which may indicate a more detailed editing, like correction of spaces, caused by issues with the interface.

All these findings were to be expected, in view of the assumed more complex user interface element and interaction that was proposed for PE mode. So, the next step was to check whether even in those conditions, it was possible for users to improve on their performance with PE mode. The two sessions in PE mode showed a progression. This evolution was measured in comparison with the different proportion of time between the two sessions, the first one with 75% of the total time. These were the results of this comparison:

- On average, more segments were edited in session D;
- In the same session, the editing times per segment were shorter;
- There was a higher number of edits in session D;
- TER scores were lower in session D;
- Speed was faster in session D.

Although some of these factors, like shorter editing times per segment, may be attributed to the fact that users had just edited the same text in session C in AC mode, the evolution shows that performance scores in PE mode may improve with familiarisation. Considering all these findings, one may think of PE mode as more appropriate for environments where editing behaviour must be constrained, such as in pedagogical uses.

These findings from the data about the differences between the two modes confirm the impressions that the users had given in their answers to the questionnaires: AC mode is faster, easier to use and they make more changes in it. The fact that PE mode could be more adapted to PE work and for shorter sequences could be elicited from part of the editing behaviours, but it would require specific testing for a conclusion to be drawn.

6.4.5. Summary and analysis of input variable: Actions

The four editing actions have also been extensively studied, and there were no major findings from statistical tests. The lessons taken from the data are summed up below.

Previous experiments in which TER was used have found that Replace is the action most used during PE sessions. Matthew Snover and colleagues in his paper on TER (Snover et al, 2006) already mention this result. Usually, Delete is the second most used action. As was mentioned above, this may be due to the way TER is implemented, which favours Replace to all other editing actions. This has been explained in section 6.3.3.

In these experiments, this result was confirmed, both as an editing action selected in HandyCAT, and as an edit identified by a TER method.

- Replace is, by far, the action most often used;
- Delete is the second action most used;
- Insert is next;
- And Move is the least used of all editing actions.

Move usually lags behind in all measures. For example, it was the only action which was used fewer times in the second PE session and it is the most difficult to identify by TER. However, it was also applied faster in session D, and it has the highest average TER score – this may be caused by TER interpreting some of the movements as an insertion and a deletion. Besides, the average number of edits that results from users choosing Move is close to two. This may happen not only because of the interpretation as delete and insert, but also because move is the only edit which is identified in two-word phrases by TER methods. In fact, average TER scores invert the order of use:

- Move is the edit with the highest TER average score,
- Then Insert,
- Delete,
- And finally Replace.

In terms of speed, delete and insert actions are, in this order, the fastest to apply. Move is faster than replace. So, speed is determined first by the difference between primary and secondary editing actions, and then by whether or not they imply content creation. The fastest editing action is the primary one that does not require creation of content, and the slowest is the secondary one that implies content creation. Both secondary actions are faster than the sum of the two primary editing actions:

Movement	00:00:08
Replace	00:00:13
Delete + Insert	00:00:15

Table 54 – Speed of secondary actions vs. sum of primary actions.

The analyses of the mismatch between HandyCAT editing actions and TER edit scores also reveals the more visible role of Replace, as it is the action that most frequently appears in the place of the other editing actions.

The assessments from the users confirm most of these findings. According to the users, Replace and Insert, in this order, are the actions that require most assistance from an editing tool. These are the two actions that are used most frequently and they require entering text. As such, they benefit from writing aids like predictive writing. The actions that do not involve editing content are virtually finished once they are selected, which makes it hard to anticipate the interest of any suggestion related to these.

However, in the question about usefulness of support to these actions, users placed Delete in the first place, followed by Insert, Replace, and Move at the end. This ordering of actions seems to follow the sequence of editing actions that was proposed, with one change in the last step. Users start by cleaning the segment, analysing it at the paradigmatic axis, and only at the end do they sort it at the syntagmatic level.

As mentioned before, Move and Delete are the actions that are harder to implement in a useful way. Delete's usefulness totally depends on its predicting power. The same happens with Move, but this also requires a good interaction.

6.4.6. Analysis of result variables: TER and Speed

This section tries to identify which factors weighed more on the values presented in terms of two main results: TER scores and Speed. These analyses were performed with decision trees. These are algorithms that run through sequences of data, organised in variables, and pick up those that most determine a result. This process is performed consecutively, until there are no more factors defining a result. The representation as a tree describes the shortest path between a big number of variables to a final result.

This analysis used the data that is described at the beginning of Appendix 10, plus several features extracted from these variables. This appendix also shows the data that results from each of these decision trees.

Main factors for TER scores

As will be demonstrated next, for all tested variables, the one that determines more strongly the results in terms of TER score is Text. A variable like Mode has virtually no effect on TER. What about users, or specific users, as revealed by specific answers to the workshop questionnaires? All the data was codified for processing by a decision tree, which first analysed the TER scores. At the top of the tree are the factors on which TER scores most strongly depend. If there are other influencing factors, they will appear in lower branches. The calculations stop when the algorithm finds a terminal, i.e. a value that cannot be further broken into dependent factors.

This is the first decision tree from the analysis of TER:

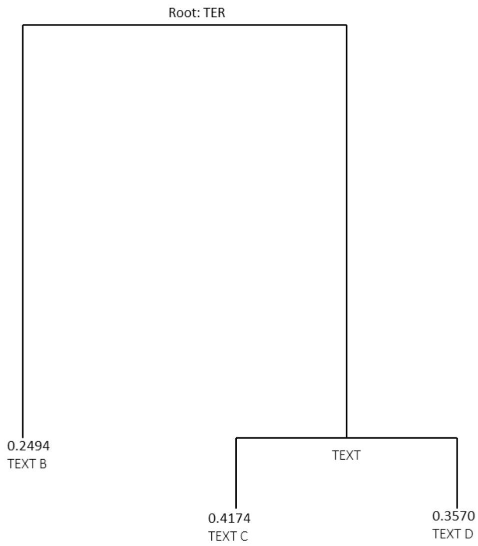


Figure 93 – Decision tree for TER scores.

The main result from this decision tree is that Text determines almost by itself the TER scores. Text B is at a specific level, and texts C and D share a separate branch. It is important to note that these tests are based on numerical data only. The fact that the three texts show such a clear division in TER scores (text B-18%, C-31%, D-26%), each in a very specific quadrant, is the probable reason for this result. All other variables present more varied and inconstant results.

The second factor that determines TER is the "experience in PE" profile, determined by the answer users gave to question Q1.9. The following decision tree presents this detail.

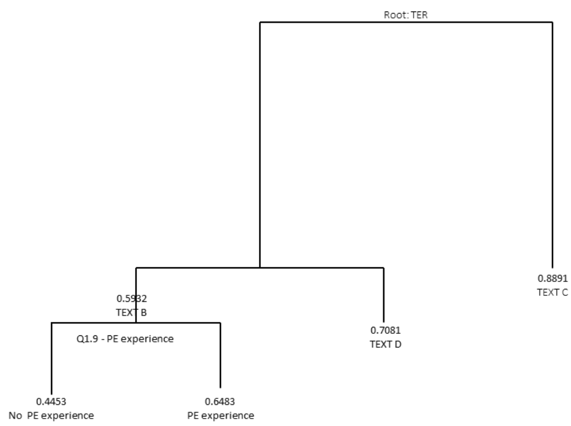


Figure 94 – Decision tree for TER scores – second level.

Although PE experience might look like a factor that could determine TER strongly, the data says otherwise. Its influence is so feeble, that if we remove the main branch, or said otherwise, if we remove the data from texts from the inputs for the calculations, the decision tree cannot calculate this relation and presents an error.

Main factors for Speed

The analysis with a decision tree for Speed presents different results. In this case, there is no single influencing factor, but Mode is the most important one. The Text variable is not even considered in the list of the factors that determine Speed. Some of the features in the branches are: the use or not of predictive writing, the use or not of MT, and whether or not PE is one of the tasks frequently performed by the user.

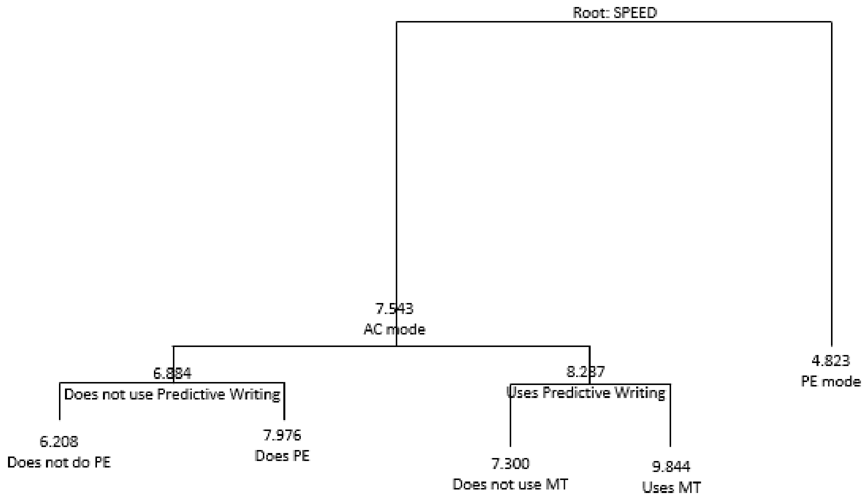


Figure 95 – Decision tree for Speed.

These analyses raise many questions, such as why is the influence of texts so strong on TER and why does Mode determine Speed so much. Perhaps Speed is a more technical variable, and, as such, it is determined by the technical characteristics of the two editing modes, followed by factors that separate users with technical expertise from those that do not have it. TER is a variable that depends more on content: users decide what to edit based on the content of the segment.

Although the data did not allow for more fruitful conclusions, decision trees seem to be reliable methods to look for and describe aspects that are linked to variations in main results. However, further tests would be required to confirm this, in view of the findings of other methods of analysis. For example, the results of the decision tree on TER (in which Mode does not appear as a contributing factor to the variation of results in the sample) might be interpreted as a contradiction to the assumption taken from previous analyses that the editing mode conditions the number of edits, with PE mode forcing users to make less edits. This is an open question for future studies on these variables.

6.5. Extended and improved testing

This section presents the lessons learnt with the tests and points to ways of building on these lessons so as to improve on the preparation of the testing stage and the generation of results.

6.5.1. Lessons learnt

The workshop was a very compact experiment, with contributions from several professional translators. These heard a proposal of a new view of PE, thought about proposals of new concepts, and tested a new approach to a translation editor. The tests were conceived to encompass a wide array of questions, so as to elicit informed feedback from participants, besides collecting a fair amount of data for analysis in a short time.

The data collection through the questionnaires was very fruitful, as some very interesting results were retrieved. However, a more focused approach might have helped get a better connection between the results of the questionnaires and the activity data.

Still, as mentioned in the introduction to Chapter 6, there was no previous methodology for such a study, so the approach was a mix of a pilot-test of new software and a case study of different speculative approaches to a work process. As will be seen in the conclusions, there were many fruitful results from this approach.

Most of the lessons learnt from the experiments concentrate on the practical tests with HandyCAT – and that is where the novelty of the approach lay. Admittedly, the tool was not developed to allow for a better experience in terms of usability and of focused data analysis. Again, the root cause of this was lack of previous comparable knowledge. Chris Hokamp did a great job with the specifications that he received, and HandyCAT has the potential to become a very useful tool for advanced research, not only into the translation process, but also into ways to interact with tools and aids during the process.

The ecological validity of the tests

As explained in the introduction to the tests, there were no claims about the ecological validity of these tests. The typical procedures of TPR were not followed, as the focus of the tests was not on a specific question about the process. So, information on pauses, users' verbal description of processes, or even keystroke data, was not collected. The focus was on collecting feedback from a specific use of an interface that is still at an experimental stage.

As there was to be no assessment of the quality of the results, participants received no instructions on the expected level of quality, or, rather, the rate of errors that were permitted. So, users were allowed to decide what was "adequate editing" in each segment. Actually, this is a guideline with a strong ecological validity, as this happens very frequently in the real world.

The texts that were edited during the workshop had been completely pre-translated by a MT process. However, section 4.2.2 had described real PE projects as including both MT and TM segments. It was also mentioned that this adds to the complexity of the decision process. So, it was decided to leave that global complexity of a real PE process behind, in order to allow for users to focus on the specific dimension of the PE process that was under scrutiny.

Use of the editing tools

The complexity and variability of editing actions and procedures was visible in this sample situation. From a wide variety of editing, with many segments that show values above the proposed editing threshold, to many different uses of the two editing modes, a lot of knowledge was acquired from these tests.

Users showed how an unrestricted editing mode like AC mode, even with predictive writing help, leads to results that are farther from the unedited versions. A more restrictive mode, like PE mode, seems to be associated with a conditioning of the process that forces users to reflect about what it is really important to edit, and also in which sequence they should do the edits. PE mode takes longer to apply each edit, so users need to make more efficient choices, and eventually reduce the number of edits.

The downside of the restrictions imposed by PE mode and the difficulties in handling them was that there were many wrong uses of the interface elements. One could observe users dedicating time to correcting spaces and capitalisation, and this gave rise to many segments with zero edits. Besides, a few users used the Replace action to delete a whole segment and rewrite it. (This may be an indication that this method is not adapted to segments that require long sequences of changes.) Others chose the Delete action repeatedly to erase a word, character by character. All of these misuses were improved or even eliminated with the familiarisation, as may be testified with the results from last session D.

Another effect of the use of these two modes was when users decided to leave a segment open and go back and re-edit another segment – this is called backtracking (Koehn, 2015). This effect could have been restrained, but it yielded valuable information. So, it should have been anticipated.

Information from the logs

The effects of the practice of backtracking could only be found in the logs after the data was analysed with some detail, and problems with divergent lengths for the same segment were detected, or outliers in duration, segments with zero edits and similar issues. This is a usual and expected result, but, in a future experiment, the information collected in logs may anticipate some of these situations and report them in a clearer way.

6.5.2. Looking ahead

Data collection in a workshop

The format of the workshop seems to be a good choice for the collection of data from specialised users in a concentrated way. Participants benefit from the initial presentation and discussion, and also from the possibility to test a new way of working. The choice of the audience for such an event should still be purposive, directed towards specialised users, namely because of the demonstration that PE is a specialised task.

In terms of the sampling of texts, improvements in this feature may be required. Still, it is not clear which characteristics the texts should have to improve the collection of data. There are no standard procedures to measure features like complexity or repetitiveness. Besides, one needs to find a balance between representativeness and comparability (as Zanettin points out), which means that one needs to be cautious in avoiding the creation of a situation that is so controlled that it does not correspond to any real scenario.

Quality and productivity evaluations may be excluded from such a study, contrary to what is usual in these studies, as long as the focus on the technical details yields useful and applicable conclusions.

The use of questionnaires is important to retrieve not only context data on the users and their opinions, but also fundamental feedback for the progression of the project. However, the purposes of the workshop should be more focused and the questionnaires should include specific questions that could be measured in the activity data.

Guiding the use of the software

This experiment should be used as a pilot test for a possible future similar project. One of the lessons to be learnt is that there needs to be more time for familiarisation with the software, ideally including proper training time. One may also consider specific sessions for evaluation of the pedagogical use of the approach and the support features. Taking into account the learning curve in this type of project is a fundamental requirement for its success. A few users admitted that the usefulness of this approach depended on the capacity to use the software.

Learning to be comfortable with this type of interactive tool may sound easy, as the tool should be intuitive and provide suggestions that guide the user, but to take advantage of all the potential of the approach requires an intentional attitude.

A proficient user with an interactive tool based on the four editing actions has to know how to make the most of each action. The system may present too many suggestions and hinder the translator's work, or too few, and depend on requests from the user. Each action has a different degree of usefulness in the prediction moment and in the typing moment. The expectation of a focus only on predictive writing may need to be forcefully changed, into a scenario in which the user anticipates the Delete and Move actions more often. So, not only must the interface achieve a good level of interaction, but the user needs to be ready for that too.

The proposal of an efficient sequence of editing actions, from cleaning the MT hypothesis to making contextual replacement actions, needs further testing, to confirm if it is one of the approaches that should be included in such a training stage.

Capturing and measuring user activity data

Implementation difficulties (see section 5.4.7), together with the mismatch between how TER reports edits and how the actual editing actions are performed (section 6.3.3.), are clear evidence that it is not easy to set up the logging feature in a way that captures all the details of this process. Nevertheless, the inclusion in the log of data related to each action's duration, the textual unit that it relates to, and how it affects the editing score, would be welcomed. An upgraded form of capturing the data could also allow for the analysis of the sequence of actions each user applies to each segment.

However, one knows that if the data collection is too detailed, it may become useless, or require an extra data treatment stage that could be avoided. An example of this is capturing the details at the character-level. Another is the number of interim changes, which do not survive in the final version of the segment; some of these may contain useful information, but one must consider that only approved segments contain approved edits.

I believe that it has been demonstrated that edit distance estimations based on products are not reliable, and that an improved logging feature is more capable of capturing process data. But lessons should be learnt from the methods to estimate edit distances. The roles played by the paradigmatic and the syntagmatic axes in a process need to be understood, for the models to be adaptable to actual use. Finally, further analysis of the weight costs employed in TER may also give important guidance for the implementation of such a tool.

User and process-centred software development deals with conscious and unconscious processes. This means that a project to improve on these tools must rely on extensive use tests, so that the best information may be collected from such simple actions as the usual keyboard and mouse operations.

Beyond the lessons taken from this experiment, there are still many open questions that might be explored on a second take on a project on interactive editing tools. In the final chapter, there is room for a reflection on these open questions.

7. CONCLUSIONS

This section will present the findings of this dissertation, starting with a verification of its grounding in the discipline of Translation Studies, followed by a description of the evolution of the propositions and research questions that were presented at the beginning. The impact of the redefinition of editing and post-editing, and of the information that was extracted from the testing stage, will also be discussed.

7.1. Context and evolution of the dissertation

7.1.1. Type of Translation Studies research

This dissertation started by describing Holmes' map of TS and the two areas from which this study sprung: Descriptive Translation Studies (DTS) with the analysis of the Translation Process (TP), and Applied Translation Studies (ATS) with the analysis of translation tools. According to another more recent attempt at describing the structure of the discipline – "The Map", (Williams and Chesterman, 2010), this research falls into the domain of "Translation and Technology", with related topics areas such as evaluation of software, effects of technology and the place of technology in translator training.

The approach of the work described in the dissertation was observational, and provided interpretive and descriptive claims. In the practical experiments described in Chapter 6, none of the initial propositions became a hypothesis subject for testing. This was related to the fact that the research questions put forward could only be tested in an extensive scientific experiment, after the appropriate tools had been developed.

7.1.2. Evolution of initial propositions

The initial propositions and research questions presented in section 1.4.2 were open and speculative, as guides to a sequence of theoretical and practical questioning, in the search for the terms, concepts and practical data that enabled the answer to these questions. On this journey, some of these questions needed to be reformulated, and new ones have arisen. Let us recall the two initial propositions:

- **First proposition:** PE is a mixed process, by which translators not only edit but also revise and translate sub-units of a sentence pre-translated by an MT process.

- **Second proposition:** PE can be described, defined and analysed based on four simple editing actions: deleting, inserting, moving and replacing words and multi-word units from a MT sentence.

Since these two descriptions of Post-editing (PE) are almost incompatible, neither of them could be transformed into a claim that competently and accurately defined what PE was. The debate between these two propositions moved to a different level as soon as the concept of “editing” was proposed. This was considered as the best description of the task of applying four specific actions (deleting, inserting, moving, and replacing) to TL material, as part of the main processes of Translation and Revision. These clear-cut concepts (editing and the four editing actions), which could be identified in several books and references from TS literature, accompanied the progression of the dissertation and helped achieve a clear view of complex debates. So, the second proposition was reshaped into a claim, which was demonstrated early in the dissertation and explored throughout it:

- **First claim:** Editing can be described, defined and analysed based on four simple editing actions: deleting, inserting, moving and replacing words and multi-word units in a TL sentence.

The first proposition required more labour. After having analysed the Translation Process, the Revision Process (both in Chapter 2), and Machine Translation theory and practice (Chapter 3), a search began on studies, practice and support to the technical dimension of PE (Chapter 4), as a way to not only approach the definition of PE, but also to guide this theoretical effort towards the two initial questions that had been presented. In Chapter 5, a new claim was presented, as the proposal for a redefinition of PE:

- **Second claim:** Post-editing is a generic name that describes a set of tasks by which a translator modifies language content that has previously been converted from a Source Language into a Target Language by a Machine Translation system, in order to make it conform to the objectives defined for the Target Text. The set of tasks required for the modification of the machine-translated content may include translating, editing and revising. **Post-editing** may be identified as being composed only of editing, but this is only possible if the purpose of the Target Text can be achieved by performing only the four technical actions (deleting, inserting, replacing and moving) over the machine-translated content, within a defined editing effort threshold.

This claim was explained in Chapter 4 and Chapter 5. One of the advantages introduced by these two claims is that they solve the incompatibility between the two initial propositions. Furthermore, it helps clarify the debate over whether PE is a form of revision or translation. The role of editing in this definition is highlighted by the proposal to set an editing threshold that may contribute to the separation of editing and translating tasks in a PE project. And the definition of PE as a form of translation helps underline its value as a specialised service.

7.1.3. Evolution of the research questions

Besides the two initial propositions, two research questions were presented:

- **First question:** Can Machine Learning (ML) techniques used by MT be used to present suggestions to post-editors, for each of these four editing actions?
- **Second question:** If translators had these PE aids incorporated into the computer tools that they use every day, would they be more efficient during PE?

To present a full and unequivocal answer to these questions would require access to software development skills, so as to build and test a tool that presented interactive suggestions to translators. The confirmation of the first question would enable the transformation of the second question into a hypothesis, and in a practical test, this hypothesis might then be tested and confirmed.

Chapters 3, 4 and 5 made the case for the confirmation that ML techniques can be used to model PE and to present suggestions in an interactive tool. Chapter 5 presented a proposal for the development of such a tool, as a speculative and explorative exercise. After Chris Hokamp confirmed his willingness to adjust HandyCAT to test part of the ideas and models described in this dissertation, the conditions were met to move this dissertation from the mere speculative to partially answer some of its questions.

It was not possible to implement a full learning module in HandyCAT, so Question 1 and 2 could not be answered. However, the concepts that arise from the two propositions (the definition of PE, the definition of editing, the four editing actions and the editing threshold) were tested and validated by professional translators, in a workshop with a strong practical component. The experiments are described and their results are presented and discussed in Chapter 6.

As described in section 6.5 above, important lessons have been learnt from this experiment and testing stage. Besides, as will be discussed in the next sections, several new research questions have been raised. So, the results that are required from a speculative and explorative approach have been met.

7.1.4. Approach to testing and developing software

Although software development is not one of the branches of TS, TP research and ATS have long focused on the need to participate in the development of tools for translation. Most of these attempts are approached as guidelines or evaluation tasks. In this dissertation, the study of the technical specificities of the editing process has enabled the presentation of a proposal for a tool that explores the potential of current ML methods in an interactive work method.

ATS research indicates usefulness and usability as the two main evaluation lines for software projects. The usefulness of this project may be evaluated in terms of solving two existing problems:

- Lack of tools that adequately support the growing complexity and increase in cognitive load brought by MT content into translation projects;
- Lack of adjustment of predictive writing and other sub-segmental approaches to the actual actions translators must perform during editing.

The pilot test approach to the testing stage was a necessary step towards the goal of contributing to the development of systems that solve the two problems stated here. The usability dimension, related to improving on existing tools, is linked to the difficulties of implementation of these concepts in dynamic and interactive tools that may be used every day by translators.

As the next sections demonstrate, this project may not have solved these problems, but it has probably contributed to focusing the discussions on practical preoccupations and to guiding the development of useful and usable translation software.

7.1.5. User-centred translation software

The second research question focused on the translators' capacity to deal with a tool that dynamically presents suggestions to support their work. The workshop in which about 50 translators participated was an opportunity to assess their receptivity to the proposals and approaches that were being prepared for this dissertation. In the introductory presentation, the main proposition of a simplified view was presented, and then translators performed the practical tests with that view in mind.

Despite the evident limitations of the interface, most translators recognised that the approach based on the four editing actions was appropriate for PE work, especially when the editing effort was not excessive. In this context, the purpose of the editing threshold, although not implemented in the tool, was received positively. Besides, most of the answers that translators presented based on their reaction to the tests, and to an intuitive reading of the activities, closely followed the results observed in the activity data that was collected from the software's logs.

The interface elements that presented the four editing actions forced users to make a conscious effort in every editing decision. This had the intended effect of bringing an unconscious process to the surface, and making the participants reflect on a process that is often seen as purely mechanic. At the end of the workshop, most users admitted that their view on PE had changed.

This feedback, and the more detailed feedback on each editing mode and the four editing actions, was globally positive, and may function as an incentive to improve on the implementation of the concepts that were presented. Furthermore, it was possible to conclude that the learning module that might enable an interactive support system is fundamental to improve the usability of such software.

The objective of the workshop to function as a pilot-test was fulfilled: users confirmed the interest in the approaches that were being followed and offered guidance to the development of tools that are useful for their work.

7.2. Contributions to the theories of the translation and post-editing processes

The starting point of this dissertation was the notion that natural languages should not be seen as barriers, because translation has the role of a bridge that enables communication across languages. As if the complexity of these languages and these bridges were not sufficient, technology advances often create their own barriers.

Machine Translation (MT) is a form of translation. Not a form of Human Translation (HT), not a way to produce finished translations, since its output is only a hypothesis or a suggestion of a translation, but it still should be recognised as a form of translation, in the strict sense that it transforms SL content into TL content. As such, together with PE, it should be studied by TS.

The main contribution of this dissertation to the theories of the translation and PE processes is the notion that editing is a task that should be defined from the bottom-up, from the micro actions that are performed when content is already presented to translators. This notion was developed from several sources, not just from TS but also from MT research. And it culminates with the proposal of KAT – Knowledge Assisted Translation, an all-encompassing technology that brings together CAT tools and MT, and which must be based on proper modelling of all the processes.

I hope I have demonstrated that not only TS theories benefit from the knowledge and approaches to translation that it may incorporate from MT, but also that MT may benefit from knowledge that is currently held by TS.

It was mentioned earlier that MT often considers that only Linguistics can contribute with useful knowledge to its work, and that the more this appears in the form of annotated corpora, the better. TS has also been described as only adding a description of style to Linguistics.

As a contribution to removing this ideological barrier, let us mention a few pieces of theory that are practice-based, and which come mostly from TS, but which have the potential to change views on both TS and MT.

- The ST is just one of the references that translators use to make decisions, so lack of parallelism between ST and TT is to be expected;

- Fragmented texts, like lists and tables of items, challenge linguistic analyses that are usually developed from properly constructed sentences; automated processes that are sensitive to length also tend to fail in this type of content;
- There are multiple translation shifts, at different levels and dimensions of the TP, which should be studied and modelled, instead of pruned out of the data;
- The view of the technical dimension of translation, or of the shallow dimension of texts, is often a simplification strategy. It is important to recognise that the complexity of this level implies several things:
    - Not everything learnt at this level scales up to other levels;
    - Data from technical effort does not necessarily align with temporal data, let alone with cognitive data;
    - Being able to compose a sentence by chunks does not mean that a meaning construction process is being approximated;
    - Approaching translation as a simple puzzle-fitting process, in the paradigmatic axis, loses a lot of information on other dimensions;
- Any claims about the decision process must be recognised as mainly speculative;
- It should be acknowledged that modelling a translation process is not the same as modelling a revision or an editing process; the fact that there is a new layer of text between the ST and the TT has a major impact on such models.

One of the advantages of bringing together knowledge from both disciplines, and using it to guide the collection of real process data, is that inefficient or unreasonable processes may be discarded. Approaches like monolingual PE, the simulation of processes artificially constructed from product data, or the estimation of details of processes such as editing distances from product data, are some of the activities that may be challenged by theory, and compensated by methods that are developed with input from real processes.

One final note on the discussions about efficiency, productivity and quality. The second research question posed the problem of whether translation aids improve efficiency. However, in the literature of ST and MT, this term was replaced by productivity. As work in the dissertation progressed, it became clear that these measures did not describe the full extent of improvements that one should expect from a deep analysis of the PE process, so it was decided not to analyse them. The focus of the work moved instead to the evaluation of usefulness and usability in the context of the development of interactive tools. But a short comment on measures of quality may be required at this concluding stage.

It should be clear that quality is not discussed in this dissertation not because of some notion of it having a secondary role to productivity, or of the need to avoid another complex discussion. The main reason for this is that quality is an elusive issue of intersubjectivity, a term that both James Holmes (1988) and Gideon Toury (1995) associate with research on translation. This means that even when research is focused on objective and measurable items, it is rare to achieve a clear and unambiguous notion of quality.

Section 6.4.3 presented a matrix of individual performances at the workshop. Some users are fast and insert many edits, others are fast but insert few edits, and others are slow but insert many edits. Which ones of these are the best post-editors? One would probably choose those that are fast and are capable of making more edits. However, that is the group that most probably includes the type of user who does not look carefully at the contents he edits. Or, on the contrary, since PE is so often associated with “less is better”, are the best post-editors in those groups with fewer edits?

Discussions on quality cannot be tainted by an excessive attention to measurable metrics, such as time and number of actions. That is the field of productivity. Quality brings together knowledge from many sides, not just the four levels of Quality Estimation (complexity, adequacy, confidence, and fluency), no matter how far-reaching this perspective is, and not only from purpose-based translation theories. Quality is an open issue that goes far beyond what is possible in a dissertation on improving how translators work with technology.

7.3. Contributions to research on interactivity in translation

After looking at editing and post-editing from the perspective of the four editing actions, analysis was focused on the way translation tools support these actions. A different method was proposed, and it was then tested with professional translators. Chapter 5 presented a model of interactivity that builds on technologies that are currently available, and which are being employed in MT, but which are more useful if they are applied in an interactive support environment for translators. This model proposes several approaches that may help develop the translation tools of the future. However, many more questions arise, with impact on how the interactivity in translation tools must be approached. Since Chapter 5 describes the proposal for an interactive tool adequately, this section will mention a few of the questions this approach opened for future research.

The main challenge in this area is the implementation of an interactive learning module. The amount of data this module must manage is quite high, and it is updated dynamically. A combination of offline and online data management is proposed, but this is recognised as a complex feature. So, the combination of different learning methods may have to be studied.

In terms of the interaction with the user, the PE mode that was used at the workshop must clearly be improved. It makes sense to combine it with the predictive writing functionality, especially if the user determines the availability of each of these modes when he chooses each editing action. The agility of this combination, together with a notion of the sequence of actions, is instrumental to the capacity to help a translator facing a segment with complex editing. This means that it would be advisable to study the methods and sequences by which the actions are applied by different users and in different editing contexts.

Another dimension of software implementation is the process to capture textual units. For the ones that depend on selection, like Move and Delete, the length of the selection is the indication of the unit. For those that imply typing, it may be difficult to define the start and finish of a unit. In most recent PE interfaces analysed in section 4.3.4, units are inside tiles, but maybe these can be made more flexible. Still, this data and the one that relates to unit alignment, should not be presented in a way that may be considered too intrusive, since data collection should not be the central preoccupation of the post-editor. The notions of segmentation of keystroke sequences and of macro and micro units used in TPR may also play a part here.

Once a tool that features such an interactivity functionality is made available, it will be possible to test and tune the role and usefulness of the different editing actions. Movement is recognised as a difficult feature to implement. Should it be prioritised and promoted, since it provides the system with validated aligned units? The adaptation of the tool to different uses and roles can also only be learnt in the presence of a proper tool. A different version for use in pedagogical contexts, and a different one for revision may also be called for. So, user adaptation studies and tests must be considered as a necessity early in the development process, so as to avoid overfitting the interface to particular uses.

An editing tool can also be the ideal environment to test new metrics, guided, for example, by the editing threshold, or targeted at high correlation with notions of effort, adapted from the different fuzzy match levels and editing levels, and many other metrics that may be used to collect different levels of editing and process data.

The definition of the editing threshold and the study of its impact in the technological environment and also in the service workflow is another new and interesting area of research.

The data relative to each editing action and the textual units they manipulate may enable further analyses. The classification of the editing actions in terms of content and position, or as primary and secondary, may lead to models that may be innovative in interactive systems. The frequency of selection of different sized units is also a possible theme for research, linked for example to the claim that these are linguistically-motivated. Another unit selection behaviour that may be studied and improved on is the long-distance effect: is it possible for users to select simultaneously all units that are affected by the same morphological process, even if this implies skipping words, and can these be replaced in a single action? As for predictive writing features, for Replace and Insert actions, the ratio of accepted vs. rejected suggestions may also be an important feature for the tool learning system.

Moreover, data analysis methods should be tested so to identify which are the best to treat such volatile data. A standard methodology for normalisation, simplification and preparation of this data would be required to enable fast and reliable conclusions. The effect of mode over TER and Speed, for example, as discussed in section 6.4.6, is worth further analysis.

This shows how such an interactive tool may be, more than just a good support for translation work, an invaluable source of data for research.

7.4. Contributions for an industry that manages knowledge

The industrial context of translation is a constant background for any research work on translation technology. In a project that aims at learning from translators' practices, and then to reapply the knowledge acquired, the industrial context plays a very important role.

Section 5.5 discusses a few of the implications of a system that will collect process data that is valuable for the adaptation of the tool to its user. The way this data is managed is central to a good flow of communication and service in the production chain.

In other sections of this dissertation, concepts that involve the organisation of services have also been discussed, since these are also dependent on the technologies used. Methods of accounting for work, quality, and effort are open themes for debate, as they are not determined solely by technological solutions. The editing threshold may help draw a line that identifies segments and texts that fall below the expectable quality produced by MT systems. While this may determine technology evaluation scores, it may also help build a more realistic notion of what actual editing work in a PE project implies.

It is the role of research to highlight the implications that technological advances have on the lives of people, namely of those who produce knowledge. The central role that the human factor plays in translation must be clearly recognised, even more so in the face of technological advances. This recognition should have an effect on the adjustment of the value that the industry puts on projects that involve humans with MT. PE is not adequately described by the current strategies that ask for more while requesting less.

By giving PE its deserved role as a specialised service, recognising that this service requires specialised users as much as translation without MT, the industry may also be guaranteeing that its role as a specialised service provider survives the different generations of translation technologies that are yet to come.

BIBLIOGRAPHY

According to APA (6th edition): http://www.apastyle.org/about-apa-style.aspx

Abekawa, T., & Kageura, K. (2008). Constructing a Corpus that Indicates Patterns of Modification between Draft and Final Translations by Human Translators. *Proceedings of the 6th International Language Resources and Evaluation*, 2000–2005.

Alabau, V., Bonk, R., Buck, C., Carl, M., Casacuberta, F., García-Martínez, M., … Tsoukala, C. (2013). CASMACAT: An Open Source Workbench for Advanced Computer Aided Translation. *The Prague Bulletin of Mathematical Linguistics*, *100*(100), 101–112. https://doi.org/10.2478/pralin-2013-0016.

Alabau, V., Carl, M., Casacuberta, F., Martínez, M. G., González-Rubio, J., Mesa-Lao, B., … Sanchis-Trilles, G. (2016). Learning Advanced Post-editing. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research* (pp. 95–110). Heidelberg: Springer International Publishing Switzerland. https://doi.org/10.1007/978-3-319-20358-4_5.

Allen, J. (2003). Post-editing. In H. Somers (Ed.), *Computers and Translation*. Amsterdam/Philadelphia: John Benjamins.

Alves, F. (2003). Foreword: Triangulation in Process-Oriented Research in Translation. In F. Alves (Ed.), *Triangulating translation: perspectives in process oriented research* (pp. vii–x). Amsterdam/Philadelphia: John Benjamins.

Alves, F., & Hurtado Albir, A. (2010). Cognitive approaches. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies - volume I*. Amsterdam/Philadelphia: John Benjamins.

Alves, F., Koglin, A., Mesa-Lao, B., Martínez, M. G., Fonseca, N. B. de L., Sá, A. de M., … Aquino, M. (2016). Analysing the Impact of Interactive Machine Translation on Post-editing Effort. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research*. Heidelberg: Springer International Publishing Switzerland. https://doi.org/10.1007/978-3-319-20358-4_4.

Alves, F., Pagano, A., Neumann, S., Steiner, E., & Hansen-Schirra, S. (2010). Translation units and grammatical shifts: towards an integration of product and process-based translation research. In G. Shreve & E. Angelone (Eds.), *Translation and Cognition*. Amsterdam/Philadelphia: John Benjamins.

Alves, F., Pagano, A., & Silva, I. L. da. (2011). Modeling (Un)Packing of Meaning in Translation: Insights from Effortful Text Product. In B. Sharp, M. Zock, M. Carl, & A. L. Jakobsen (Eds.), *Proceedings of the 8th International NLPCS Workshop* (pp. 153–163). Copenhagen: Samfundslitteratur.

Alves, F., & Vale, D. C. (2011). On drafting and revision in translation: a corpus linguistics oriented analysis of translation process data. *Translation: Computation, Corpora, Cognition*, *1*(1), 105–122.

Angelone, E. (2010). Uncertainty, uncertainty management, and metacognitive problem solving in the translation task. In G. Shreve & E. Angelone (Eds.), *Translation and Cognition*. Amsterdam/Philadelphia: John Benjamins.

Anzulewicz, A., Asanowicz, D., Windey, B., Paulewicz, B., Wierzchoń, M., & Cleeremans, A. (2015). Does level of processing affect the transition from unconscious to conscious perception? *Consciousness and Cognition*, *36*(June), 1–11.

Aranberri, N., & Choudhury, R. (2012). *Advancing Best Practices in Machine Translation Evaluation*. Retrieved 5 November 2016, from https://www.taus.net/file-downloads/download?path=Reports%252FMember%2BReports%252Ftaus-bestpracticesmtevaluation-july2012.pdf.

Arnold, D., Balkan, L., Meijer, S., Humphreys, R. L., & Sadler, L. (1994). *Machine Translation: An Introductory Guide*. NCC Blackwell.

Austermühl, F. (2001). *Electronic tools for translators*. Manchester: St. Jerome.

Austermühl, F. (2013). Future (and not-so-future) Trends in the Teaching of Translation Technology. *Tradumàtica: Tecnologies de La Traduciò*, (11), 326–337.

Aymerich, J. (2005). Using Machine Translation for fast, inexpensive, and accurate health information assimilation and dissemination: experiences at the Pan American Health Organization. In *9th World Congress on Health Information and Libraries*. Salvador - Bahia, Brazil.

Aymerich, J., & Camelo, H. (2009). The Machine Translation Maturity Model at PAHO. In *Proceedings of the Machine Translation Summit XII*. Ottawa, Canada.

Aziz, W., Castilho, S., & Specia, L. (2012). PET: a Tool for Post-editing and Assessing Machine Translation. *The Eighth International Conference on Language Resources and Evaluation*, 3983–3987.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate, arXiv:1409.0473 1–15.

Baker, M. (1993). Corpus Linguistics and Translation Studies: Implications and applications. In *Text and Technology: in Honour of John Sinclair* (pp. 233–250). Amsterdam/Philadelphia: John Benjamins.

Baker, M. (1999). *Routledge Encyclopedia of Translation Studies*. London: Routledge. https://doi.org/10.1080/07374836.1999.10524085.

Bar-Hillel, Y. (1951). The present state of research on mechanical translation. *American Documentation*, *2*, 229–237.

Barrachina, S., Bender, O., Casacuberta, F., Civera, J., Cubel, E., Khadivi, S., … Vilar, J.-M. (2009). Statistical Approaches to Computer-Assisted Translation. *Computational Linguistics*, *35*(1), 3–28.

Bartolomé, D., & Ramírez, G. (2016). Beyond the Hype of Neural Machine Translation. In *TAUS Translation Technology Showcase*. Dublin: TAUS.

Bassnett-McGuire, S. (1980). *Translation Studies*. London & NY: Methuen.

Béchara, H., Ma, Y., & Genabith, J. van. (2011). Statistical post-editing for a statistical MT system. *MT Summit XIII*, 308–315.

Beeby, A., Fernández, M., Fox, O., Albir, A. H., Kozlova, I., Kuznik, A., … Wimmer, S. (2009). Results of the validation of the PACTE translation competence model: Acceptability & decision making. *Across Languages & Cultures*, *10*(2), 207230.

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus Phrase-Based Machine Translation Quality: a Case Study. *CoRR*, *2*.

Bertoldi, N., & Federico, M. (2012). Practical Domain Adaptation in SMT. In *AMTA Tutorial*. San Diego, USA.

Bertoldi, N., Simianer, P., Cettolo, M., Wäschle, K., Federico, M., & Riezler, S. (2014). Online adaptation to post-edits for phrase-based statistical machine translation. *Machine Translation*, *28*(3–4), 309–339. https://doi.org/10.1007/s10590-014-9159-7.

Blain, F., Logacheva, V., & Specia, L. (2016). Phrase-Level Segmentation and Labelling of Machine Translation Errors. *Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, (May), 2240–2245.

Blain, F., Senellart, J., Schwenk, H., Plitt, M., & Roturier, J. (2011). Qualitative Analysis of Post-Editing for High Quality Machine Translation. *MT Summit XIII*, 164–171.

Bojar, O., Chatterjee, R., Federmann, C., Graham, Y., Haddow, B., Huck, M., … Zampieri, M. (2016). Findings of the 2016 Conference on Machine Translation (WMT16). *Wmt-2016*, *2*, 131–198.

Bojar, O., Chatterjee, R., Federmann, C., Haddow, B., Huck, M., Hokamp, C., … Turchi, M. (2015). Findings of the 2015 Workshop on Statistical Machine Translation. *Proceedings of the Tenth Workshop on Statistical Machine Translation*, (September), 1–46.

Bojar, O., Federmann, C., Haddow, B., Koehn, P., Post, M., & Specia, L. (2016). Ten Years of WMT Evaluation Campaigns: Lessons Learnt. In *Proceedings of the LREC 2016 Workshop 'Translation Evaluation – From Fragmented Tools and Data Sets to an Integrated Ecosystem'*. Portoroz, Slovenia.

Bowker, L. (2002). *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.

Bowker, L. (2005). Productivity vs quality? A pilot study on the impact of translation memory systems. *Localisation Focus*, *4*(March), 13–20.

Bowker, L., & Fisher, D. (2010). Computer-aided translation. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies - volume I* (pp. 60–65). Amsterdam/Philadelphia: John Benjamins.

Bravo, C., & Enríquez, V. (2006). Summary of discussion on Localization and Translator Training. In A. Pym, A. Perekrestenko, & B. Starink (Eds.), *Translation Technology and its Teaching (with much mention of localization)*. Intercultural Studies Group Universitat Rovira i Virgili.

Brown, P., Cocke, J., Della Pietra, S., Della Pietra, V., Jelinek, J., Mercer, R., & Roossin, P. (1988). A Statistical Approach to Language Translation. In *Proceedings of the 12th International Conference on Computational Linguistics* (Vol. 1, pp. 22–27). Budapest.

Brown, P., Pietra, S. A. Della, Pietra, V. J. Della, & Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, *19*, 263–311.

Byrne, J. (2004). *Textual Cognetics and the Role of Iconic Linkage in Software User Guides*. Dublin City University.

Byrne, J. (2006). *Technical Translation: Usability strategies for translating technical documentation*. Dordrecht: Springer.

Callison-Burch, C., Bannard, C., & Schroeder, J. (2004). Improved Statistical Translation Through Editing. In *European Association for Machine Translation (EAMT-2004) Workshop* (pp. 26–32).

Callison-Burch, C., Koehn, P., Monz, C., Post, M., Soricut, R., & Specia, L. (2012). Findings of the 2012 Workshop on Statistical Machine Translation. *Proceedings of the Seventh Workshop on Statistical Machine Translation*, 10–51.

Callison-Burch, C., Osborne, M., & Koehn, P. (2006). Re-evaluation the Role of BLEU in Machine Translation Research. *EACL 2006*, *6*, 249–256.

Carl, M. (2009). Triangulating product and process data: quantifying alignment units with keystroke data. In I. Mees, S. Göpferich, & F. Alves (Eds.), *Methodology, technology and innovation in translation process research: a tribute to Arnt Lykke Jakobsen* (pp. 225–247). Copenhagen: Samfundslitteratur.

Carl, M. (2012). Translog-II: a Program for Recording User Activity Data for Empirical Translation Process Research. In *The 8th International Conference on Language Resources and Evaluation. LREC 2012* (Vol. 3, pp. 153–162). Istanbul.

Carl, M., Bangalore, S., & Schaeffer, M. (Eds.). (2016). *New Directions in Empirical Translation Process Research*. Heidelberg: Springer.

Carl, M., Dragsted, B., & Jakobsen, A. L. (2011). On the systematicity of human translation processes. In *Tralogy 2011. Translation Careers and Technologies: Convergence Points for the Future*. Paris, France.

Carl, M., & Jakobsen, A. L. (2009). Towards statistical modelling of translators' activity data. *International Journal of Speech Technology*, *12*(4), 125–138.

Carl, M., Kay, M., & Jensen, K. T. H. (2010). Long distance revisions in drafting and post-editing. *Research in Computing Science--Special Issue: Natural Language Processing and Its Applications*, *46*, 193–204.

Carl, M., Lacruz, I., Yamada, M., & Aizawa, A. (2016a). Comparing spoken and written translation with post-editing in the ENJA15 English-Japanese Translation Corpus. In *22nd Annual Meeting of the Association for Natural Language Processing* (pp. 1–4). Sendai, Japan.

Carl, M., Lacruz, I., Yamada, M., & Aizawa, A. (2016b). Measuring the Translation Process. In *The 22nd Annual Meeting of the Association for Natural Language Processing, NLP 2016* (pp. 0–3). Japan.

Carl, M., & Way, A. (Eds.). (2003). *Recent Advances in Example-Based Machine Translation*. Dordrecht: Kluwer Academic Press.

Carmo, F. do. (1998). *A variação do número de palavras em tradução: análise de sintagmas nominais*. Faculdade de Letras da Universidade do Porto.

Carmo, F. do. (2010). *Why do translations have more words than original texts?* (MA thesis reviewed - unpublished).

Carmo, F. do. (2017). Tradução Automática Interactiva e Pós-edição. In *Workshop on Interactive Machine Translation and Post-editing*. Porto.

Carmo, F. do, & Maia, B. (2013). Sleeping with the enemy? Or should translators work with Google Translate? In P. Sánchez-Gijón, O. Torres-Hostench, & B. Mesa-Lao (Eds.), *Conducting Research in Translation Technologies*. Bern: Peter Lang.

Carmo, F. do, & Maia, B. (2016). A Description of Post-editing, from Translation Studies to Machine Learning. In *II Congreso Internacional T3L: Tradumática, Tecnologías de la Traducción y Localización*. Barcelona, Spain.

Carmo, F. do, Trigo, L., & Maia, B. (2016). From CATs to KATs. In J. Esteves-Ferreira, J. Macan, R. Mitkov, & O.-M. Stefanov (Eds.), *Proceedings of the 38th Conference Translating and the Computer (TC38)*. Geneva: Editions Tradulex.

Carpuat, M., & Simard, M. (2012). The Trouble with SMT Consistency. *Proceedings of the 7th Workshop on Statistical Machine Translation*, 442–449.

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). https://doi.org/10.1515/pralin-2017-0013.

Catford, J. C. (1965). *A Linguistic Theory of Translation*. London: OUP.

Cettolo, M., Bertoldi, N., Federico, M., Schwenk, H., Barrault, L., & Servan, C. (2014). Translation project adaptation for MT-enhanced computer assisted translation. *Machine Translation*, *28*(2), 127–150. https://doi.org/10.1007/s10590-014-9152-1.

Chatterjee, R., de Souza, J. G. C., Negri, M., & Turchi, M. (2016). The FBK Participation in the WMT 2016 Automatic Post-editing Shared Task. In *Proceedings of the 1st Conference on Machine Translation: shared task papers* (Vol. 2, pp. 745–750).

Chatterjee, R., Negri, M., Turchi, M., & Kessler, F. B. (2016). Instance Selection for Online Automatic Post-Editing in a Multi-domain Scenario. In *AMTA 2016 - Researcher's Track*. Austin, Texas.

Chatterjee, R., Weller, M., Negri, M., & Turchi, M. (2015). Exploring the Planet of the APEs: a Comparative Study of State-of-the-art Methods for MT Automatic Post-Editing. *Association for Computational Linguistics (ACL)*, (3), 156–161.

Chesterman, A. (2016). *Memes of Translation: The spread of ideas in translation theory*. Amsterdam/Philadelphia: John Benjamins.

Chiang, D. (2007). Hierarchical Phrase-based Translation. *Computational Linguistics*, *33*(2), 201–228. https://doi.org/10.1162/coli.2007.33.2.201.

Cho, K. (2015a). Introduction to Neural Machine Translation with GPU (part 2). Retrieved 5 November 2016 https://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-with-gpus-part-2.

Cho, K. (2015b). Introduction to Neural Machine Translation with GPUs (part 1). Retrieved 5 November 2016 https://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-with-gpus/.

Cho, K. (2015c). Introduction to Neural Machine Translation with GPUs (part 3). Retrieved 5 November 2016 https://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-with-gpus-part-3.

Cho, K., van Merrienboer, B., Bahdanau, D., & Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. In *Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8)*. https://doi.org/10.3115/v1/W14-4012.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Christensen, T. P. (2011). Studies on the Mental Processes in Translation Memory-assisted Translation - the State of the Art. *Trans-Kom*, *4*, 137–160.

Christensen, T. P., & Schjoldager, A. (2010). Translation-Memory (TM) Research: What Do We Know and How Do We Know It? *Hermes*, *44*(44), 89–101.

Christensen, T. P., & Schjoldager, A. (2011). The Impact of Translation-Memory (TM) Technology on Cognitive Processes: Student-translators' retrospective comments in an online questionnaire. In B. Sharp, M. Zock, M. Carl, & A. L. Jakobsen (Eds.), *Proceedings of the 8th International NLPCS Workshop (Special theme: Human-Machine Interaction in Translation)*. Copenhagen: Samfundslitteratur.

Commission, E. (1994). *Final report of the results of Eurotra*. Retrieved 5 November 2016 from http://aei.pitt.edu/36888.

Costa-jussà, M., Rapp, R., Lambert, P., Eberle, K., Banchs, R., & Babych, B. (Eds.). (2015). *Hybrid Approaches to Machine Translation*. Springer Verlag.

Crego, J., Kim, J., Klein, G., Rebollo, A., Yang, K., Senellart, J., … Zoldan, P. (2016). SYSTRAN's Pure Neural Machine Translation Systems. *CoRR*.

Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, Calif: Sage Publications.

Cronin, M. (2003). *Translation and Globalization*. Abingdon and New York: Routledge.

Cronin, M. (2013). *Translation in the digital age*. Abingdon and New York: Routledge. https://doi.org/10.4324/9780203073599.

Culo, O., Gutermuth, S., Hansen-Schirra, S., & Nitzke, J. (2014). The Influence of Post-editing on Translation Strategies. In S. O'Brien, L. W. Balling, M. Carl, M. Simard, & L. Specia (Eds.), *Post-editing of Machine Translation: Processes and Applications*. Cambridge Scholars Publishing.

Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, *7*(3), 171–176.

Declercq, C. (2014). Editing in Translation Technology. In C. Millán & F. Bartrina (Eds.), *Routledge Encyclopedia of Translation Technology* (pp. 480–493). Abingdon: Routledge.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B Methodological*, *39*(1), 1–38.

DeNero, J., Kumar, S., Chelba, C., & Och, F. (2010). Model combination for machine translation. In *Proceedings of the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Los Angeles, USA.

Denkowski, M., & Lavie, A. (2012). Challenges in Predicting Machine Translation Utility for Human Post-Editors. *Proceedings of AMTA 2012*.

Denkowski, M., & Lavie, A. (2014). Meteor Universal: Language Specific Translation Evaluation for Any Target Language. *Wmt*, (June 2014), 376–380.

Denkowski, M., Lavie, A., Lacruz, I., & Dyer, C. (2014). Real Time Adaptive Machine Translation for Post-Editing with cdec and TransCenter. *Workshop on Humans and Computer-Assisted Translation*, (April), 72–77.

DePalma, D., & Lommel, A. (2017). Augmented Translation Powers up Language Services. Retrieved from http://www.commonsenseadvisory.com/Blogs.aspx?1=1&moduleID=390&Contenttype=ArticleDetAD&AId=37907.

Depraetere, I. (2010). What counts as useful advice in a university post-editing training context? Report on a case study. *EAMT 2010 - 14th Annual Conference of the European Association for Machine Translation*.

Devlin, J., Zbib, R., Huang, Z., Lamar, T., Schwartz, R., & Makhoul, J. (2014). Fast and robust neural network joint models for statistical machine translation. In *ACL 2014* (pp. 1370–1380). https://doi.org/10.3115/v1/P14-1129.

Diamond, B. J., & Shreve, G. M. (2010). Neural and physiological correlates of translation and interpreting in the bilingual brain: recent perspectives. In G. M. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 289–321). Amsterdam/Philadelphia: John Benjamins.

Dillon, S., & Fraser, J. (2006). Translators and TM: An investigation of translators' perceptions of translation memory adoption. *Machine Translation*, *20*(2), 67–79. https://doi.org/10.1007/s10590-006-9004-8.

Dimitrova, B. E. (2005). Combining Product and Process Analysis: Explicitation as a Case in Point. *Bulletin Suisse de Linguistique Appliquée*, (81), 25–39.

Dimitrova, B. E. (2010). Translation process. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies - volume I*. Amsterdam/Philadelphia: John Benjamins.

Diño, G. (2016). Translation Research Gets almost a Third of Ireland's €50M ADAPT Investment. Retrieved 5 November 2016, from http://slator.com/academia/translation-research-gets-almost-a-third-of-irelands-e50m-adapt-investment/.

Doddington, G. (2002). Automatic evaluation of machine translation quality using N-gram co-occurrence statistics. *HLT 2002*, 138–145.

Doherty, S. (2012). *Investigating the effects of controlled language on the reading and comprehension of machine translated texts: A mixed-methods approach*. Dublin City University.

Doherty, S., & Kenny, D. (2014). The design and evaluation of a Statistical Machine Translation syllabus for translation students. *The Interpreter and Translator Trainer*, *8*(2), 295–315. https://doi.org/10.1080/1750399X.2014.937571.

Doherty, S., Kenny, D., & Way, A. (2012). Taking Statistical Machine Translation to the Student Translator. *Conference of the Association for Machine Translation in the Americas (AMTA) 2012*.

Domingo, M., Peris, Á., & Casacuberta, F. (2016). Interactive-Predictive Translation based on Multiple Word-Segments. *Baltic Journal of Modern Computing*, *4*(2), 282–291.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78. https://doi.org/10.1145/2347736.2347755.

Dorr, B. (editor). (2010). Part 5: Machine Translation Evaluation. In J. Olive, C. Christianson, & J. McCary (Eds.), *Handbook of Natural Language Processing and Machine Translation* (pp. 801–894). Springer Verlag.

Dragsted, B. (2010). Coordination of reading and writing processes in translation: an eye on uncharted territory. In C. Shreve & E. Angelone (Eds.), *Translation and Cognition*. Amsterdam/Philadelphia: John Benjamins.

Dragsted, B., & Carl, M. (2013). Towards a classification of translation styles based on eye-tracking and keylogging data. *Journal of Writing Research*, *5*(1), 133–158.

Dugast, L., Senellart, J., & Koehn, P. (2007). Statistical Post-Edition on SYSTRAN Rule-Based Translation System. In *Proceedings of the Second Workshop on Statistical Machine Translation* (pp. 220–223). Prague, Czech Republic.

Dugast, L., Senellart, J., & Koehn, P. (2009). Statistical Post Editing and Dictionary Extraction: Systran/Edinburgh submissions for ACL-WMT2009. *Proceedings of the 4th Workshop on Statistical Machine Translation*, 110–114.

Eetemadi, S., Lewis, W., Toutanova, K., & Radha, H. (2015). Survey of data-selection methods in statistical machine translation. *Machine Translation*, *29*(3–4).

Ehrensberger-Dow, M., & Heeb, A. H. (2016). Investigating the ergonomics of a technologized translation workplace. *Reembedding Translation Process Research*, (October), 69–88. https://doi.org/10.1075/btl.128.04ehr.

Ehrensberger-Dow, M., & Massey, G. (2016). Why ergonomics matters in translation. In *Next Generation Translation Tools Workshop*. Swansea.

Eisele, A., & Lavecchia, C. (2011). Using Statistical Machine Translation for Computer-Aided Translation at the European Commission. In V. Zhechev (Ed.), *Proceedings of the Third Joint EM+/CNGL Workshop Bringing MT to the Users: Research Meets Translators (JEC '11)*. Luxembourg.

Enríquez Raído, V. (2014). *Translation and Web Searching*. Abingdon and New York: Routledge. https://doi.org/10.1556/Acr.15.2014.1.8.

España-Bonet, C., & Màrquez, L. (2010). Robust estimation of feature weights in Statistical Machine Translation. *EAMT 2010 - 14th Annual Conference of the European Association for Machine Translation*.

Esplà, M., Sánchez-Martínez, F., & M. Forcada. (2011). Using word alignments to assist computer-aided translation users by marking which target-side words to change or keep unedited. *Proceedings of the 15th Annual Conference of the European Association for Machine Translation*, (May), 81–88.

Esselink, B. (1998). *A practical guide to software localization*. Amsterdam/Philadelphia: John Benjamins.

Esselink, B. (2000). *A practical guide to localization*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/liwd.4

Faes, F. (2015). That's Big: Spain Pours $100 Million into Language Technology. Retrieved from http://slator.com/demand-drivers/thats-big-spain-pours-100-million-into-language-technology/.

Faes, F. (2016). Hyperbolic? Experts Weigh in on Google Neural Translate. Retrieved from https://slator.com/technology/hyperbolic-experts-weigh-in-on-google-neural-translate/.

Fantinuoli, C. (2016). The influence of translation technologies on language production. *Input a Word, Analyse the World: Selected Approaches to Corpus Linguistics*, (June).

Fawcett, P. (1997). *Translation and Language: Linguistic Theories Explained*. Manchester: St. Jerome Publishing.

Federico, M., Bertoldi, N., Cettolo, M., Negri, M., Turchi, M., Trombetti, M., … Germann, U. (2014). The Matecat Tool. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations*, *7*(287688), 129–132.

Fidel, R. (2012). *Human Information Interaction: An Ecological Approach to Information Behavior*. MIT Press.

Filkin, P. (2013a). Life without Trados. Retrieved 14 January 2016, from https://multifarious.filkin.com/2013/09/30/life-without-trados/.

Filkin, P. (2013b). Solving the post-edit puzzle. Retrieved 25 May 2017, from https://multifarious.filkin.com/2013/12/28/solving-the-post-edit-puzzle/.

Firth, J. R. (1957). A synopsis of linguistic theory 1930-1955. In *Studies in Linguistic Analysis. Special volume of the Philological Society* (pp. 1–32). Oxford: Blackwell.

Flanagan, K. (2013). Lift – towards true sub-segment TM recall (poster). In *Translating Europe*. Brussels: Directorate-General for Translation.

Flanagan, K. (2014). Filling in the gaps: what we need from TM subsegment recall. In *Proceedings of the 36th Conference Translating and the Computer (TC36)* (pp. 12–23). London: AsLing.

Flanagan, K. (2015). Subsegment recall in Translation Memory — perceptions, expectations and reality. *The Journal of Specialised Translation*, (23), 64–88.

Flanagan, K. (2016). Fine-grained alignment, subsegment recall and match repair. In *Next Generation Translation Tools Workshop*. Swansea.

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

Fomicheva, M., Bel, N., & da Cunha, I. (2015). Neutralizing the Effect of Translation Shifts on Automatic Machine Translation Evaluation. In A. Gelbukh (Ed.), *CICLing 2015. Lecture Notes in Computer Science*. Springer Verlag.

Fonseca, N. B. de L., & Alves, F. (2016). Assessing complexity and difficulty levels of machine-translated texts. *Letras & Letras*, *32*(1), 306–337.

Forcada, M. (2014). On the annotation of TMX translation memories for advanced leveraging in computer-aided translation. *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC-2014)*, 4374–4378.

Forcada, M., & Sánchez-Martínez, F. (2015). A general framework for minimizing translation effort: towards a principled combination of translation technologies in computer-aided translation. *Proceedings of the 18th Annual Conference of the European Association for Machine Translation*, 27–34.

Forcada, M., & Ñeco, R. (1997). Recursive hetero-associative memories for translation. In J. Mira, R. Moreno-Díaz, & J. Cabestany (Eds.), *Biological and Artificial Computation: From Neuroscience to Technology*. Berlin: Springer Verlag.

Frankenberg-Garcia, A., & Gough, J. (2016). Next Generation Translation Tools. In *Next Generation Translation Tools Workshop*. Swansea.

Gambier, Y., & Doorslaer, L. van (Eds.). (2010). *Handbook of Translation Studies - volume I*. Amsterdam/Philadelphia: John Benjamins.

García-Martínez, M., Singla, K., Tammewar, A., Mesa-Lao, B., Thakur, A., Anusuya, M. A., … Carl, M. (2014). SEECAT: ASR & Eye-tracking Enabled Computer Assisted Translation. In *Proceedings of the 17th Annual Conference of the European Association for Machine Translation: EAMT2014*. Basel, Switzerland.

García, I. (2006). Translators on translation memories: a blessing or a curse? In A. Pym, A. Perekrestenko, & B. Starink (Eds.), *Translation Technology and its Teaching (with much mention of localization)* (pp. 97–105). Tarragona.

García, I. (2007). Power shifts in web-based translation memory. *Machine Translation*, *21*(1), 55–68. https://doi.org/10.1007/s10590-008-9033-6.

García, I. (2009). Research on translation tools. In A. Pym & A. Perekrestenko (Eds.), *Translation Research Projects 2*. Tarragona: Intercultural Studies Group.

García, I. (2012). A brief history of postediting and of research on postediting. *Revista Anglo Saxonica*, *3*, 293–310.

García, I. (2014). Computer-aided translation. In S. Chan (Ed.), *The Routledge Encyclopedia of Translation Technology* (pp. 68–87). Abingdon & NY: Routledge.

Gentzler, E. (1993). *Contemporary Translation Theories*. London: Routledge.

Giménez, J., & Màrquez, L. (2009). Discriminative Phrase Selection for SMT. In C. Goutte, N. Cancedda, M. Dymetman, & G. Foster (Eds.), *Learning Machine Translation* (pp. 205–236). Cambridge, Massachusetts/London, England: MIT Press.

González-Rubio, J., Ortiz-Martínez, D., Casacuberta, F., & Benedí, J. M. (2016). Beyond Prefix-Based Interactive Translation Prediction. *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, 198–207.

Google. (2016). Google History. Retrieved 5 November 2016, from https://www.google.com/about/company/history/.

Gouadec, D. (2007). *Translation as a Profession*. Amsterdam/Philadelphia: John Benjamins.

Gough, N., & Way, A. (2003). Example-Based Controlled Translation. In *EAMT-04 - 9th Workshop of the European Association for Machine Translation* (pp. 73–81). La Valleta, Malta.

Goutte, C., Cancedda, N., Dymetman, M., & Foster, G. (2009). *Learning Machine Translation*. Cambridge, Massachusetts/London, England: MIT Press.

Gow, F. (2003). *Metrics for evaluating translation memory software*. University of Ottawa.

Green, S., Chuang, J., Heer, J., & Manning, C. D. (2014). Predictive Translation Memory: A Mixed-Initiative System for Human Language Translation. In U. Honolulu (Ed.), *UIST 2014*. https://doi.org/http://dx.doi.org/10.1145/2642918.2647408.

Green, S., Heer, J., & Manning, C. D. (2013). The efficacy of human post-editing for language translation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 439.

Green, S., Heer, J., & Manning, C. D. (2015). Natural Language Translation at the Intersection of AI and HCI: Old questions being answered with both AI and HCI. *Queue 13*, *13*(6), 1–13. https://doi.org/http://dx.doi.org/10.1145/2791301.2798086.

Grönroos, M., & Becks, A. (2005). Bringing Intelligence to Translation Memory Technology. In *Proceedings of the 27th International Conference on Translating and the Computer (TC27)*.

Guerberof Arenas, A. (2008). *Productivity and Quality in the Post-editing of Outputs from Translation Memories and Machine Translation*. Universitat Rovira i Virgili.

Guerberof Arenas, A. (2014). The Role of Professional Experience in Post-editing from a Quality and Productivity Perspective. In S. O'Brien, L. W. Balling, M. Carl, M. Simard, & L. Specia (Eds.), *Post-editing of Machine Translation: Processes and Applications*. Cambridge Scholars Publishing.

Guerberof Arenas, A., Depraetere, H., & O'Brien, S. (2012). What we know and what we would like to know about post-editing. *Tradumàtica: Tecnologies de La Traducció*, (10), 211–218. https://doi.org/10.1080/08853908.2012.682022.

Guzmán, R. (2007). Manual MT Post-editing: 'If It's Not Broken, Don't Fix It'. *Translation Journal*, *11*(4), 21–26.

Halevy, A., Norvig, P., & Pereira, F. (2009). The Unreasonable Effectiveness of Data. *IEEE Intelligent Systems*, *24*(2), 8–12. https://doi.org/10.1109/MIS.2009.36.

Halliday, M. A. K., & Martin, J. R. (1993). *Writing Science: Literacy and Discursive Power*. London: The Falmer Press.

Han, A. L.-F., & Wong, D. F. (2016). Machine Translation Evaluation: A Survey. *CoRR*.

Hansen, G. (2013). The Translation Process as Object of Research. In C. Millán & F. Bartrina (Eds.), *The Routledge Handbook of Translation Studies* (pp. 88–101). Abingdon: Routledge.

Hartnett, P., & SDL Community. (2014). Post-Edit Compare. Retrieved 25 May 2017, from http://posteditcompare.wiki-site.com/index.php/Main_Page.

He, Y. (2011). *The Integration of Machine Translation and Translation Memory*. Dublin City University.

Hearne, M., & Way, A. (2011). Statistical Machine Translation: A Guide for Linguists and Translators. *Linguistics and Language Compass*, *5*(5), 205–226.

Heyn, M. (1996). Integrating Machine Translation into Translation Memory Systems. In *Proceedings of the EAMT Machine Translation Workshop, TKE'96* (pp. 29–30). Vienna, Austria.

Hokamp, C. (2015a). A Component-Centric Design Framework for Translation Interfaces. In H. Costa, A. Zaretskaya, G. Corpas Pastor, L. Specia, & M. Seghiri (Eds.), *Proceedings of the EXPERT Scientific and Technological Workshop* (pp. 91–102). Malaga, Spain.

Hokamp, C. (2015b). Component-Centric Design HandyCAT Prototypes and Experiments Software. In *EXPERT Workshop*. Malaga.

Hokamp, C., & Liu, Q. (2015). HandyCAT: The Flexible CAT Tool for Translation Research. In *EAMT 2015*. Istanbul.

Holmes, J. S. (1988). *Translated! Papers on Literary Translation and Translation Studies*. Amsterdam: Rodopi.

Horvitz, E. (1999). Principles of mixed-initiative user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI Is the Limit - CHI '99*, 159–166. https://doi.org/10.1145/302979.303030.

House, J. (2000). Consciousness and the strategic use of aids in translation. *Tapping and Mapping the Processes of Translation and Interpreting*, 149–162.

Hu, K., & Cadwell, P. (2016). A Comparative Study of Post-editing Guidelines. *Baltic Journal of Modern Computing*, *4*(2), 346–353.

Huang, J., & Minocha, A. (2014). Cognitive Process of Revision: the Behaviours and the Motivations Behind. In *Translation in transition - Between Cognition, Computing and Technology*. Copenhagen.

Hutchins, J. (1986). *Machine translation: past, present, future*. Chichester: Ellis Horwood.

Hutchins, J. (1996). ALPAC: the (in)famous report. *MT News International*, (14), 9–12.

Hutchins, J. (1998). The origins of the translator’s workstation. *Machine Translation*, *13*(4), 287–307.

Hutchins, J. (2004). Two precursors of machine translation: Artsrouni and Trojanskij. *International Journal of Translation*, *16*(1), 11–31.

Hutchins, J. (2010). Machine translation: a concise history. *Journal of Translation Studies - Special Issue: The Teaching of Computer-Aided Translation*, *13*(1 & 2), 29–70.

Hutchins, J. (2012). Milestones in the history of machine translation. Science Museum, London.

Hutchins, J. (2005). Towards a definition of example-based machine translation. In *MT Summit X second workshop on example-based machine translation* (pp. 63–70). Phuket, Thailand.

Hutchins, J., & Somers, H. (1992). *An Introduction to Machine Translation*. London: Academic Press. https://doi.org/10.1016/B978-0-85709-710-1.50015-X.

ISO. (2015). 17100:2015 Translation services - Requirements for translation services. Geneva: International Organization for Standardization.

Jakobsen, A. L. (2003). Effects of think aloud on translation speed, revision, and segmentation. In F. Alves (Ed.), *Triangulating translation: perspectives in process oriented research* (Vol. 45, pp. 69–95). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/btl.45.08jak.

Jakobsen, A. L., & Schou, L. (1999). Translog Documentation, version 1.0. In G. Hansen (Ed.), *Probing the Process in Translation: Methods and Results* (pp. 1–36). Copenhagen: Samfundslitteratur.

Jawaid, B., & Bojar, O. (2014). Two-Step Machine Translation with Lattices. *LREC 2014*. Reykjavik, Iceland. (pp. 682–686=.

Jiménez-Crespo, M. A. (2009). The effect of Translation Memory tools in translated Web texts: Evidence from a comparative product-based study. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, 213–232.

Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., … Dean, J. (2016). Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *CoRR*, 1–16. https://doi.org/10.3115/v1/P15-1002

Junczys-Dowmunt, M., Dwojak, T., & Hoang, H. (2016). Is Neural Machine Translation Ready for Deployment? A Case Study on 30 Translation Directions. *CoRR*.

Junczys-Dowmunt, M., & Grundkiewicz, R. (2016). Log-linear Combinations of Monolingual and Bilingual Neural Machine Translation Models for Automatic Post-Editing. In *Proceedings of the First Conference on Machine Translation, WMT 2016* (Vol. 2, pp. 741–748). Berlin, Germany: Association for Computational Linguistics.

Jurafsky, D., & Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (2nd ed., Vol. 21). New Jersey: Prentice Hall.

Kalchbrenner, N., & Blunsom, P. (2013). Recurrent Continuous Translation Models. *EMNLP 2013*, (October), 1700–1709.

Kay, M. (1997). The Proper Place of Men and Machines in Language Translation. *Machine Translation*, *12*(1–2), 3–23.

Kenny, D., & Doherty, S. (2014). Statistical machine translation in the translation curriculum: Overcoming obstacles and empowering translators. *Interpreter and Translator Trainer*, *8*(2), 276–294. https://doi.org/10.1080/1750399X.2014.936112.

King, G. W. (1956). Stochastic Methods of Mechanical Translation. *Mechanical Translation*, *3*(2), 38–39.

Kitchenham, B., Pickard, L., & Pfleeger, S. (1995). Case Studies for Method and Tool Evaluation. *Software, IEEE*, *12*(4), 52–62. https://doi.org/10.1109/52.391832.

Knight, K. (1999). Decoding Complexity in Word-Replacement Translation Models. *Computational Linguistics*, (25).

Koby, G. S., & Melby, A. K. (2013). Certification and job task analysis (JTA): Establishing validity of translator certification examinations. *Translation and Interpreting*, *5*(1), 174–210. https://doi.org/10.12807/ti.105201.2013.a10.

Koehn, P. (2008). *Statistical Machine Translation*. Cambridge: Cambridge University Press.

Koehn, P. (2009a). A process study of computer-aided translation. *Machine Translation*, *23*(4), 241–263. https://doi.org/10.1007/s10590-010-9076-3.

Koehn, P. (2009b). A Web-Based Interactive Computer Aided Translation Tool. *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations*, (August), 17–20.

Koehn, P. (2010). Enabling Monolingual Translators: Post-Editing vs . Options. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL* (pp. 537–545).

Koehn, P. (2015). Advances in Computer Aided Translation Beyond Post-Editing. In *Invited Talk at MT Summit 2015*.

Koehn, P. (2016a). Neural network models. Retrieved 16 January 2017 http://mt-class.org/jhu/assets/papers/neural-network-models.pdf.

Koehn, P. (2016b). The State of Neural Machine Translation. Retrieved from https://omniscien.com/wp-content/uploads/2016/02/20161124-Omniscien_Technologies-NMT-Webinar.pdf.

Koehn, P. (2017). Introduction to Neural Machine Translation. Retrieved from http://omniscien.com/more/resources/webinar/#-introduction-to-neural-machine-translation--nmt-----technology-webinar-series.

Koehn, P., Haddow, B., Williams, P., & Hoang, H. (2010). More Linguistic Annotation for Statistical Machine Translation. *Proceedings of the Joint 5th Workshop on Statistical Machine Translation and MetricsMATR*, (July), 115–120.

Koehn, P., Och, F., & Marcu, D. (2003). Statistical phrase-based translation. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, *1*, 48–54.

Koehn, P., Saint-Amand, H., & Alabau, V. (2015). *Final Report on Interactive Editing*. Retrieved 17 January 2017 http://cordis.europa.eu/docs/projects/cnect/6/287576/080/deliverables/001-deliverable33Ares20152495853.pdf.

Koehn, P., & Senellart, J. (2010). Convergence of translation memory and statistical machine translation. *Proceedings of AMTA Workshop on MT Research and the Translation Industry*, 21–31.

Koehn, P., & University of Edinburgh. (2015). Machine Learning. Retrieved from http://www.statmt.org/survey/Topic/MachineLearning.

Koehn, P., Zens, R., Dyer, C., Bojar, O., Constantin, A., Herbst, E., … Moran, C. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions - ACL '07*, (June), 177.

Koponen, M. (2016). *Machine Translation Post-editing and Effort Empirical Studies on the Post-editing Process*. University of Helsinki.

Krings, H. P. (2001). *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. (G. S. Koby, Ed.). Kent, Ohio & London: The Kent State University Press.

Kuhns, B. (2007). It's not MT, and it's not TM. Retrieved from https://www.taus.net/thinktank/%0Aarticles/translate-articles/its-not-mt-and-its-not-tm

Labaka, G., & Melero, M. (Eds.). (2011). Proceedings of the International workshop LIHMT and of the Shared Task MLK4HMT. In *LIHMT and ML4HMT*. Barcelona: Universitat Politècnica de Catalunya.

Lacruz, I., Shreve, G. M., & Angelone, E. (2012). Average Pause Ratio as an Indicator of Cognitive Effort in Post-Editing: A Case Study. *AMTA 2012 Workshop on Post-Editing Technology and Practice (WPTP 2012)*, 21–30.

Lagarda, A.-L., Alabau, V., Casacuberta, F., Silva, R., & Díaz-de-Liaño, E. (2009). Statistical Post-Editing of a Rule-Based Machine Translation System. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers* (pp. 217–220).

Lagoudaki, E. (2006). Translation Memories Survey 2006: Users' perceptions around TM use. In *Proceedings of the 28th International Conference on Translating and the Computer (TC28)* (Vol. 2006, pp. 1–29). London: Aslib.

Langlais, P., Foster, G., & Lapalme, G. (2000). TransType: a Computer-Aided Translation Typing System. In *Proceedings of the 2000 NAACL-ANLP Workshop on Embedded machine translation systems*. Seattle, Washington.

Langlais, P., Sauvé, S., Foster, G., Macklovitch, E., & Lapalme, G. (2000). Evaluation of TransType, a computer-aided translation typing system: a comparison of a theoretical and a user-oriented evaluation procedure. *Conference on Language Resources and Evaluation (LREC)*, 8.

Lapshinova-Koltunski, E. (2013). VARTRA: A Comparable Corpus for Analysis of Translation Variation. In *ACL 2013* (pp. 77–86). Sofia: Association for Computational Linguistic.

Lavie, A. (2011). Evaluating the output of machine translation systems. *AMTA Tutorial*, *19*, 189–201.

Lavie, A., & Agarwal, A. (2007). METEOR: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments. In *Proceedings of the Second Workshop on Statistical Machine Translation* (Vol. 0, pp. 65–72).

Lavie, A., Sagae, K., & Jayaraman, S. (2004). The Significance of Recall in Automatic Metrics for MT Evaluation. In R. E. Frederking & K. B. Taylor (Eds.), *6th Conference of the Association for Machine Translation in the Americas - Machine Translation: From Real Users to Research* (pp. 134–143). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-30194-3_16.

Le, H.-S., Allauzen, A., & Yvon, F. (2012). Continuous Space Translation Models with Neural Networks. *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 39–48.

LeBlanc, M. (2013). Translators on translation memory (TM). Results of an ethnographic study in three translation services and agencies. *Translation and Interpreting*, *5*(2), 1–13. https://doi.org/105202.2013.a01.

Lederer, M. (1986). Transcoder ou Réexprimer? In D. Seleskovitch & M. Lederer (Eds.), *Interpréter pour Traduire*. Paris: Didier Érudition.

Leuven-Zwart, K. van. (1989). Translation and Original: Similarities and Dissimilarities I. *Target*, *1:2*, 151–181.

Leuven-Zwart, K. van. (1990). Translation and Original: Similarities and Dissimilarities II. *Target*, *2:1*, 69–95.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, *10*(8), 707–710.

Li, L., Way, A., & Liu, Q. (2016). Phrase-Level Combination of SMT and TM Using Constrained Word Lattice. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 275–280.

Liddy, E. D. (2001). Natural Language Processing. In *Encyclopedia of Library and Information Science*. New York: Marcel Decker.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22 140*, 55. https://doi.org/2731047.

Ling, W., Graça, J. V., Trancoso, I., & Black, A. W. (2012). Entropy-based Pruning for Phrase-based Machine Translation. *EMNLP 2012*, (July), 962–971.

Liparini Campos, T. (2010). *O Efeito do Uso de um Sistema de Memória de Tradução e da Pressão de Tempo no Processo Cognitivo de Tradutores Profissionais*. Universidade Federal de Minas Gerais.

Löffler-Laurian, A.-M. (1996). *La Traduction Automatique*. Villeneuve d'Asc (Nord): Presses Universitaires du Setentrion.

Logacheva, V., & Specia, L. (2015). Phrase-level Quality Estimation for Machine Translation. *International Workshop on Spoken Language Translation (IWSLT)*, 143–150.

Loinaz, I., Arantzabal, I., Forcada, M., Gómez Guinovart, X., Padró, L., Pichel Campos, J. R., & Waliño, J. (2006). OpenTrad: Traducción automática de código abierto para las lenguas del Estado español. *Procesamiento del Lenguaje Natural, 37,* 357–360.

Lommel, A. (2017a). Is Neural MT Really as Good as Human Translation? Retrieved 15 November 2016 from http://www.commonsenseadvisory.com/default.aspx?Contenttype=ArticleDetAD&tabID=63&Aid=37896&moduleId=390.

Lommel, A. (2017b). Zero-Shot Translation Is Both More and Less Important Than You Think. Retrieved 15 November 2016 from http://www.commonsenseadvisory.com/Default.aspx?Contenttype=ArticleDetAD&tabID=63&Aid=37905&moduleId=390.

LT-Innovate.eu. (2013). *LT2013: Status and Potential of the European Language Technology Markets*.

Ma, Y., He, Y., Way, A., & van Genabith, J. (2011). Consistent Translation using Discriminative Learning: A Translation Memory-inspired Approach. In *ACL 2011* (pp. 1239–1248).

Machado, M. J., & Fontes, H. L. (2011). Machine Translation: Case Study – English Into Portuguese — Evaluation of Moses in DGT Portuguese Language Department Using Moses for Mere Mortals. *A Folha*, (37).

Macklovitch, E. (2006). TransType2: The Last Word. In *Proceedings of the 5th International Conference on Languages Resources and Evaluation (LREC 06)* (pp. 167–172). Genoa, Italy.

Maia, B. (2003). What are comparable corpora? In *Proceedings of pre-conference workshop Multilingual Corpora: Linguistic Requirements and Technical perspectives at Corpus Linguistics 2003* (pp. 27–34). Lancaster, UK.

Maia, B. (2008). Machine Translation and Human Translation: Using machine translation engines and corpora for teaching and research. In M. de los Á. G. González, J. L. Mackenzie, & E. M. G. Álvarez (Eds.), *Current Trends in Contrastive Linguistics: Functional and cognitive perspectives.* (Vol. 21, pp. 123–145). Amsterdam: John Benjamins.

Malone, J. L. (1988). *The Science of Linguistics in the Art of Translation: Some Tools from Linguistics for the Analysis and Practice of Translation*. New York: State University of New York Press.

Manning, C. D. (2012). Segmentation from 97% to 100% Is It Time for Some Linguistics? In Gelbukh Alexander (Ed.), *Computational Linguistics and Intelligent*

Text Processing, 12th International Conference, CICLing 2011, Proceedings, Part I (pp. 171–189). Springer.

Manning, C. D. (2015). Computational Linguistics and Deep Learning. *Computational Linguistics*, *41*(4). https://doi.org/10.1162/COLI.

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press.

Martín-Mor, A. (2017). MTradumàtica: Statistical Machine Translation Customisation for Translators. *SKASE Journal of Translation and Interpretation*, *11*(1), 1–18.

Martín-Mor, A., & Sánchez-Gijón, P. (2013). Differences between translations made with and without CAT tools: an empirical approach. In P. Sánchez-Gijón, O. Torres-Hostench, & B. Mesa-Lao (Eds.), *Conducting Research in Translation Technologies* (pp. 209–226). Bern: Peter Lang AG.

Martín, R. M. (2014). A Blurred Snapshot of Advances in Translation Process Research. *MonTI Special Issue – Minding Translation*, 49–84.

Martínez-Gómez, P., Minocha, A., Huang, J., Carl, M., Bangalore, S., & Aizawa, A. (2014). Recognition of translator expertise using sequences of fixations and keystrokes. *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA'14*, 299–302. https://doi.org/10.1145/2578153.2578201.

Massardo, I., van der Meer, J., O'Brien, S., Hollowood, F., Aranberri, N., & Drescher, K. (2016). MT Post-Editing Guidelines. Retrieved 17 January 2017 http://www.commonspaces.eu/document/39/download/.

Melby, A. K. (1982). Multi-level Translation Aids in a Distributed System. *Proceedings of the 9th International Conf. on Computational Linguistics*, 215–220.

Melby, A. K. (1983). Computer-assisted translation systems: the standard design and a multi-level design. *Proceedings of the First Conference on Applied Natural Language Processing*, 174–177.

Melby, A. K. (1992). The translator workstation. In J. Newton (Ed.), *Computers in Translation: A practical appraisal* (pp. 147–165). London & New York: Routledge.

Melby, A. K. (2006). MT+TM+QA: the future is ours. *Tradumàtica: Tecnologies de La Traducció*, *4*.

Melby, A. K. (2015). QT21: A New Era for Translators and the Computer. In *Proceedings of the 37th Conference Translating and the Computer (TC37)* (pp. 1–11). London: AsLing.

Melby, A. K., Fields, P., Hague, D., Koby, G. S., & Lommel, A. (2014). Defining the Landscape of Translation. *Tradumàtica: Tecnologies de La Traducció*, (12), 392–403.

Melby, A. K., & Warner, C. T. (1995). *The Possibility of Language: A discussion of the nature of language, with implications for human and machine translation*. Amsterdam/Philadelphia: John Benjamins.

Mesa-Lao, B. (2011). Explicitation in translation memory-mediated environments. Methodological conclusions from a pilot study. *Translation and Interpreting*, *3*(1), 44–57.

Mesa-Lao, B. (2013). Introduction to post-editing. In *SEECAT - Speech & Eye-Tracking Enabled CAT*. Copenhagen Business School.

Millán, C., & Bartrina, F. (Eds.). (2013). *The Routledge Handbook of Translation Studies*. Abingdon: Routledge.

Mitkov, R., & Corpas, G. (2008). Improving Third Generation Translation Memory systems through identification of rhetorical predicates. *Proceedings of LangTech'2008*, (December 2014).

Molina, L., & Hurtado Albir, A. (2002). Translation Techniques Revisited: A Dynamic and Functionalist Approach. *Meta: Journal des Traducteurs*, *47*(4), 498.

Moorkens, J. (2012). *Measuring Consistency in Translation Memories: A Mixed-Methods Case Study*. Dublin City University.

Moorkens, J., Lewis, D., Reijers, W., Vanmassenhove, E., & Way, A. (2016). Translation Resources and Translator Disempowerment. In *Proceedings of ETHI-CA2 2016: ETHics In Corpus Collection, Annotation & Application* (pp. 49–53). Portoroz.

Moorkens, J., & O'Brien, S. (2013). User attitudes to the post-editing interface. In *Proceedings of Machine Translation Summit XIV: Second Workshop on Post-editing Technology and Practice* (pp. 19–25). Nice, France.

Moorkens, J., & O'Brien, S. (2017). Assessing User Interface Needs of Post-Editors of Machine Translation Interfaces for Editing Human Translation and Post-Editing Machine. In D. Kenny & J. Williams (Eds.), *Human Issues in Translation Technology*. Taylor & Francis.

Moorkens, J., O'Brien, S., da Silva, I. A. L., Fonseca, N. B. de L., & Alves, F. (2015). Correlations of perceived post-editing effort with measurements of actual effort. *Machine Translation*, *29*(3–4), 267–284. https://doi.org/10.1007/s10590-015-9175-2.

Moorkens, J., O'Brien, S., & Vreeke, J. (2016). Developing and testing Kanjingo: A mobile app for post-editing. *Revista Tradumàtica: Tecnologies de La Traducció, 14*, 58–66.

Moorkens, J., & Way, A. (2016). Comparing Translator Acceptability of TM and SMT Outputs. *Baltic Journal of Modern Computing*, *4*(2), 141–151.

Moran, J., Saam, C., & Lewis, D. (2014). Towards desktop-based CAT tool instrumentation. In *Third Workshop on Post-Editing Technology and Practice, The 11th Conference of the Association for Machine Translation in the Americas* (pp. 99–112).

Mossop, B. (2006). Has Computerization Changed Translation? *Meta: Journal des Traducteurs*, *51*(4), 787–805. https://doi.org/10.7202/014342ar.

Mossop, B. (2014). *Revising and Editing for Translators* (3rd ed.). Abingdon and New York: Routledge.

Muñoz Miquel, A. (2015). El desarrollo de la competencia traductora a través de la socialización con el experto en la materia: una experiencia didáctica. *The Journal of Specialised Translation*, (23), 333–351.

Nagao, M. (1984). A framework for a mechanical translation between Japanese and English by analogy principle. *Artificial and Human Intelligence*, 351–354.

Neubig, G., Morishita, M., & Nakamura, S. (2015). Neural Reranking Improves Subjective Quality of Machine Translation: NAIST at WAT2015. *CoRR*, *2014*.

Newmark, P. (1984). *Approaches to Translation*. Oxford: Pergamon Institute of English.

Newmark, P. (1988). *A Textbook of Translation*. London & NY: Prentice Hall.

Newton, J. (1992). *Computers in translation: a practical appraisal*. Abingdon and New York: Routledge. https://doi.org/10.4324/9780203320969.

Ni, Y., Saunders, C., Szedmak, S., & Niranjan, M. (2010). Exploitation of machine learning techniques in modelling phrase movements for machine translation. *Journal of Machine Learning Research*, *12*, 1–30.

Nida, E. A. (1964). *Toward a Science of Translating*. *Meta: Journal des traducteurs*. Leiden: E. J. Brill. https://doi.org/10.7202/003030ar.

Nirenburg, S. (1989). Knowledge-Based Machine Translation. *Machine Translation*, *4*, 5–24.

Nirenburg, S. (1992). Tools for Machine-Aided Translation: The CMU TWS. *Meta: Journal Des Traducteurs, 37*(4), 709–720. https://doi.org/10.7202/003739ar.

Norvig, P. (2011). On Chomsky and the Two Cultures of Statistical Learning. Retrieved from http://norvig.com/chomsky.html.

NYU - Department of Computer Science. (2011). Machine Translation. Retrieved 8 June 2017, from http://nlp.cs.nyu.edu/mt/index.html.

O'Brien, S. (2002). Teaching Post-editing: A Proposal for Course Content. *Proceedings of the 6th EAMT Workshop, Teaching Machine Translation*, 99–106.

O'Brien, S. (2009). Eye tracking in translation process research: methodological challenges and solutions. In I. M. Mees, F. Alves, & S. Göpferich (Eds.), *Methodology, technology and innovation in translation process research: a tribute to Arnt Lykke Jakobsen* (pp. 251–266). Copenhagen: Samfundslitteratur.

O'Brien, S. (2012). Translation as Human Computer Interaction. *Translation Spaces*, *1*(1), 101–122. https://doi.org/http://dx.doi.org/10.1075/ts.1.05obr.

O'Brien, S. (2013). The Borrowers: Researching the cognitive aspects of translation. *Target*, *25*(1), 5–17. https://doi.org/10.1075/target.25.1.02obr.

O'Brien, S. (2016). A Broad Overview of Ergonomics: Related research at the Centre for Translation and Textual Studies - 2006-2016. In *Next Generation Translation Tools Workshop*. Swansea.

O'Brien, S., Moorkens, J., & Vreeke, J. (2014). Kanjingo: A Mobile App for Post-Editing. *Third Workshop on Post-Editing Technology and Practice*, 125.

O'Brien, S., O'Hagan, M., & Flanagan, M. (2010). Keeping an eye on the UI design of Translation Memory How do translators use the 'Concordance' feature? In *European Conference on Cognitive Ergonomics* (pp. 187–190). Delft.

O'Hagan, M. (2012). The impact of new technologies on translation studies. In C. Millán & F. Bartrina (Eds.), *The Routledge Handbook of Translation Studies* (pp. 503–518). Abingdon and New York: Routledge.

Och, F. (2003). Minimum Error Rate Training in Statistical Machine Translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics* (pp. 160–167). Sapporo - Japan.

Och, F., & Ney, H. (2002). Discriminative training and maximum entropy models for statistical machine translation. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, (July), 295.

Och, F., & Ney, H. (2004). The alignment template approach to statistical machine translation. *Computational Linguistics*, *30*(4), 417–449.

Olsen, M. (1988). INK TextTools: Bilingual glossary management for commercial and academic translators. *Language Technology, 6*.

Ortiz-Martínez, D. (2016). Online Learning for Statistical Machine Translation. *Computational Linguistics*, *42*(1 (March 2016)), 121–161.

Ortiz-Martínez, D., & Casacuberta, F. (2014). The New Thot Toolkit for Fully-Automatic and Interactive Statistical Machine Translation. In *EACL-2014: 14th Conference of the European Chapter of the Association for Computational Linguistics, Demonstrations* (pp. 45–48). Gothenburg, Sweden.

Ortiz-Martínez, D., González-Rubio, J., Alabau, V., Sanchis-Trilles, G., & Casacuberta, F. (2016). Integrating Online and Active Learning in a Computer-Assisted Translation Workbench. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research*. Heidelberg: Springer International Publishing Switzerland. https://doi.org/10.1007/978-3-319-20358-4-3.

Ortiz-Martínez, D., Sanchís, G., Casacuberta, F., Alabau, V., Vidal, E., Benedí, J.-M., … González, J. (2012). The CASMACAT Project: The Next Generation Translator's Workbench. *Proceedings of the 7th Jornadas en Tecnología del Habla and the 3rd Iberian SLTech Workshop (IberSPEECH),* 326–334.

PACTE Research Group. (2014). First Results of PACTE Group's Experimental Research on Translation Competence Acquisition: The Acquisition of Declarative Knowledge of Translation. In *MonTI Special Issue – Minding Translation Special issue 1. Minding Translation* (pp. 85–115). Universitat d'Alacant | Universitat Jaume I | Universitat de València. https://doi.org/10.6035/MonTI.2014.ne1.2.

Pal, S. (2015). Statistical Automatic Post Editing. *Proceedings of the EXPERT Scientific and Technological Workshop*, 13–22.

Pal, S., Naskar, S. K., Vela, M., & van Genabith, J. (2016). A Neural Network based Approach to Automated Post-Editing. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.* 281–286.

Pal, S., Zampieri, M., Naskar, S. K., Nayak, T., Vela, M., & Van Genabith, J. (2016). CATaLog Online: Porting a Post-Editing Tool to the Web. In *LREC 2016* (pp. 599–604). Portoroz, Slovenia.

Palumbo, G. (2009). *Key Terms in Translation Studies*. New York: Continuum.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting* *of the Association for Computational Linguistics* (pp. 311–318). Philadelphia: Association for Computational Linguistics. https://doi.org/10.3115/1073083.1073135.

Parra Escartín, C. (2015). Creation of new TM segments: Fulfilling translators' wishes. *Proceedings of the International Workshop on Natural Language Processing for Translation Memories (NLP4TM)*, 1–8.

Parton, K., Habash, N., Mckeown, K., & Iglesias, G. (2012). Can Automatic Post-Editing Make MT More Meaningful? *EAMT 2012: Proceedings of the 16th Annual Conference of the European Association for Machine Translation, Trento, Italy, May 28-30 2012*, (May), 111–118.

Pérez, C. R. (2012). A Flexible Decision Tool for Implementing Post-Editing Guidelines. *Localisation Focus*, *11*(1).

Pierce, J., J. Carroll, Hamp, E., Hays, D., Hockett, C., Oettinger, A., & Perlis, A. (1966). *Language and Machines: Computers in Translation and Linguistics - The ALPAC Report*. Washington D.C. https://doi.org/10.17226/9547.

Planas, E. (2005). SIMILIS. Second-generation translation memory software. In *Proceedings of the 27th International Conference on Translating and the Computer (TC27)*. AsLib.

Planas, E., Josselin-Leray, A., & Picton, A. (2016). Knowledge Rich Contexts: a useful feature for translators. In *Next Generation Translation Tools Workshop*. Swansea.

Popovic, M., Avramidis, E., Burchardt, A., Hunsicker, S., Schmeier, S., Tscherwinka, C., … Uszkoreit, H. (2014). Involving language professionals in the evaluation of machine translation. *LREC 2014*, Reykjavik, Iceland. *48*(4), 541–559.

Popovic, M., Lommel, A. R., Burchardt, A., Avramidis, E., & Uszkoreit, H. (2014). Relations between different types of post-editing operations, cognitive effort and temporal effort. *The Seventeenth Annual Conference of the European Association for Machine Translation (EAMT 14)*, 191–198.

Potet, M., Besacier, L., Blanchon, H., & Azouzi, M. (2012). Towards a Better Understanding of Statistical Post-Edition Usefulness. In *Proceedings of the 9th International Workshop on Spoken Language Translation (IWSLT)*. Hong Kong, China.

Pym, A. (2002). Bert Esselink, A Practical Guide to Localization (2000). *Target*, *14*(1), 189–193. https://doi.org/10.1075/target.14.1.17pym.

Pym, A. (2004). *The Moving Text: Localization, translation, and distribution*. Amsterdam/Philadelphia: John Benjamins.

Pym, A. (2010). *Exploring Translation Theories*. Abingdon & NY: Routledge.

Pym, A. (2011a). Translation research terms: a tentative glossary for moments of perplexity and dispute. *Translation Research Projects 3, 3*, 75–99.

Pym, A. (2011b). What technology does to translating. *Translation and Interpreting*, *3*(1), 1–9.

Pym, A. (2013). Translation Skill-Sets in a Machine-Translation Age. *Meta: Journal des Traducteurs, 58*(3), 487. https://doi.org/10.7202/1025047ar.

Pym, A. (2015). Where Translation Studies Lost the Plot: Relations with Language Teaching (pp. 1–13). Talk given at Rikkyo University, Tokyo (December 14, 2015) - Version 1.2.

Pym, A. (2016). *A Typology of Translation Solution Types*. Retrieved 7 January 2017 https://www.academia.edu/28663257/A_typology_of_translation_solution_types.

Rabadán, R. (2008). Refining the idea of 'applied extensions'. In A. Pym (Ed.), *Beyond Descriptive Translation Studies. Investigations in homage to Gideon Toury* (pp. 103–118). Amsterdam/Philadelphia: John Benjamins.

Rabadán, R. (2010). Applied Translations Studies. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies - volume I.* Amsterdam/Philadelphia: John Benjamins.

Reinke, U. (2013). State of the Art in Translation Memory Technology. *Translation: Computation, Corpora, Cognition*, *3*(1), 27–48.

Resnik, P., & Lin, J. (2010). Evaluation of NLP Systems. In A. Clark, C. Fox, & S. Lappin (Eds.), *The Handbook of Computational Linguistics and Natural Language Processing* (pp. 271–296). Chichester: Wiley-Blackwell.

Risku, H. (2016). Situated learning in translation research training: Academic research as a reflection of practice. *Interpreter and Translator Trainer*, *10*(1), 12–28.

Risku, H., Milosevic, J., & Pein-Weber, C. (2016). Writing vs. translating: dimensions of text production in comparison. In R. Muñoz Martínez (Ed.), *Reembedding Translation Process Research* (pp. 47–68). John Benjamins.

Robin, E. (2016). The Translator as Reviser. In I. Horváth (Ed.), *The Modern Translator and Interpreter* (pp. 45–56). Budapest: Eötvös University Press.

Rubino, R., & Huet, S. (2012). Statistical Post-Editing of Machine Translation for Domain Adaptation. *EAMT 2012: Proceedings of the 16th Annual Conference of the European Association for Machine Translation*, (May), 221–228.

Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. *Pearson* (3rd edition). New Jersey: Prentice Hall.

Safaba. (2017). The data dilution effect. Retrieved from http://www.machinetranslation.net/quick-guide-to-machine-translation/statistical-machine-translation-data-dilution-effect.

Sager, J. (1993). *Language engineering and translation: Consequences of automation*. Amsterdam/Philadelphia: John Benjamins.

Samuelsson-Brown, G. (1993 & 2010). *A practical guide for translators* (1st ed. & 5th edition). Multilingual Matters.

Sanchis-Trilles, G., Alabau, V., Buck, C., Carl, M., Casacuberta, F., García-Martínez, M., … Vidal, E. (2014). Interactive Translation Prediction versus Conventional Post-editing in Practice: A Study with the CasMaCat Workbench. *Machine Translation*, *28*(3–4), 217–235. https://doi.org/10.1007/s10590-014-9157-9.

Sarmento, L., Maia, B., Santos, D., Pinto, A., & Cabral, L. (2006). Corpógrafo V3 From Terminological Aid to Semi-automatic Knowledge Engineering. In N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odjik, & D. Tapias (Eds.), *5th International Conference on Language Resources and Evaluation (LREC'06)*. Genoa, Italy.

Schmitt, G. N. (1998). Design and construction as computer-augmented intelligence processes. In *Proceedings of the Third Conference on Computer Aided Architectural Design Research in Asia* (pp. 11–20). Osaka, Japan.

SDL. (2016). The history of SDL language solutions. Retrieved 14 January 2016, from http://www.translationzone.com/about/history.html.

SDL Trados. (2017). Useful review features in SDL Trados Studio. Retrieved 20 January 2017, from https://www.youtube.com/watch?v=nVfAHmCbjJ4.

Séguinot, C. (2000). Management Issues in the Translation Process. In S. Tirkkonen-Condit & R. Jääskeläinen (Eds.), *Tapping and mapping the processes of translation and interpreting* (pp. 143–148). Amsterdam/Philadelphia: John Benjamins.

Sennrich, R. (2016). How Grammatical is Character-level Neural Machine Translation? Assessing MT Quality with Contrastive Translation Pairs. Retrieved 5 November 2016 http://arxiv.org/abs/1612.04629.

Sennrich, R., Haddow, B., & Birch, A. (2015a). Improving Neural Machine Translation Models with Monolingual Data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.* (pp. 86–96). ACL.

Sennrich, R., Haddow, B., & Birch, A. (2015b). Neural Machine Translation of Rare Words with Subword Units. *CoRR*.

Shapira, D., & Storer, J. A. (2001). Edit Distance with Move Operations. *Journal of Discrete Algorithms*, *5*. https://doi.org/10.1007/3-540-45452-7_9.

Shreve, G., & Angelone, E. (Eds.). (2010). *Translation and Cognition*. Amsterdam/Philadelphia: John Benjamins.

Shuttleworth, M., & Cowie, M. (1997). *Dictionary of Translation Studies*. Manchester: St. Jerome Publishing.

Simard, M., & Foster, G. (2013). PEPr: Post-Edit Propagation Using Phrase-based Statistical Machine Translation. In *Proceedings of the XIV Machine Translation Summit* (pp. 191–198). Nice, France.

Simard, M., & Langlais, P. (2001). Sub-sentential exploitation of translation memories. *Machine Translation Summit VIII*, 335–339.

Simianer, P., Karimova, S., & Riezler, S. (2016). A Post-editing Interface for Immediate Adaptation in Statistical Machine Translation. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, 16–20.

Chan, S. (Ed.). (2015). *The Routledge Encyclopedia of Translation Technology*. Abingdon and New York: Routledge. https://doi.org/10.4324/9781315749129.ch3.

Snell-Hornby, M. (1988). *Translated Studies: an Integrated Approach*. Amsterdam/Philadelphia: John Benjamins Publishing.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. *Proceedings of AMTA 2006*, (August), 223–231. https://doi.org/10.1.1.129.4369.

Snover, M., Madnani, N., Dorr, B., & Schwartz, R. (2008). TERp system description. *MetricsMATR Workshop at AMTA.*

Snover, M., Madnani, N., Dorr, B., & Schwartz, R. (2009). TER-Plus: Paraphrase, semantic, and alignment enhancements to Translation Edit Rate. *Machine Translation*, *23*(2–3), 117–127. https://doi.org/10.1007/s10590-009-9062-9.

Somers, H. (1999). Review Article: Example-based Machine Translation. *Machine Translation*, *14*(2), 113–157. https://doi.org/10.1023/A:1008109312730.

Somers, H. (Ed.). (2003). *Computers and Translation*. Amsterdam/Philadelphia: John Benjamins.

Specia, L., Raj, D., & Turchi, M. (2010). Machine translation evaluation versus quality estimation. *Machine Translation*, *24*(1), 39–50.

Specia, L., & Scarton, C. (2016). Quality Estimation of Machine Translation: Recent advances, challenges and software. In *Tutorial on Quality Estimation*. Alicante.

Specia, L., Shah, K., de Souza, J. G. C., & Cohn, T. (2013). QuEst - A Translation Quality Estimation Framework. In *Proceedings of the 51th Conference of the Association for Computational Linguistics - Demo sessions* (pp. 79–84).

Specia, L., Turchi, M., Wang, Z., Shawe-Taylor, J., & Saunders, C. (2009). Improving the Confidence of Machine Translation Quality Estimates. In P*roceedings of the twelfth machine translation Summit,* (pp 136–143).

Stajner, S., Querido, A., Rendeiro, N., Rodrigues, J., & Branco, A. (2016). Use of Domain-Specific Language Resources in Machine Translation. In *LREC2016: 10th Language Resources and Evaluation Conference* (pp. 592–598). Portoroz, Slovenia.

Stoll, C.-H. (1988). Translation tools on PC. In C. Picken (Ed.), *Translating and the Computer 9*. London: AsLing.

Sugden, D. (1985). Machine Aids to Translation: Automated Language Processing System (ALPS). *Meta: Journal des Traducteurs*, *30*(4), 403.

TAUS. (2010). Post-editing in practice. Retrieved 15 March 2017, from https://www.taus.net/think-tank/reports/postedit-reports/postediting-in-practice.

TAUS. (2016a). Post-editing. Retrieved 15 March 2017, from https://www.taus.net/knowledgebase/index.php/Post-editing.

TAUS. (2016b). *TAUS Quality Dashboard*.

Teixeira, C. (2014). *The Impact of Metadata on Translator Performance: How Translators Work with Translation Memories and Machine Translation*. Universitat Rovira i Virgili. https://doi.org/10.13140/RG.2.1.2190.2887.

Teixeira, C., & O'Brien, S. (2016). Recent advances in translation tools from a cognitive ergonomics perspective. In *EST Congress 2016*. Aarhus.

Temizöz, Ö. (2014). *Postediting Machine Translation Output and its Revision: Subject Matter Experts versus Professional Translators*. Universitat Rovira i Virgili.

Tiedemann, J. (2008). Prospects and Trends in Data-Driven Machine Translation. *Resourceful Language Technology*.

Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'2012)* (pp. 2214–2218).

Tirkkonen-Condit, S. (2000). Uncertainty in translation processes. In S. Tirkkonen-Condit & R. Jääskeläinen (Eds.), *Tapping and mapping the processes of translation and interpreting* (pp. 123–142). Amsterdam/Philadelphia: John Benjamins.

Tirkonnen-Condit, S., & Jääskeläinen, R. (Eds.). (2000). *Tapping and Mapping the Processes of Translation and Interpreting*. Amsterdam/Philadelphia: John Benjamins.

Toral, A., & Sánchez-Cartagena, V. M. (2017). A Multifaceted Evaluation of Neural versus Phrase-Based Machine Translation for 9 Language Directions. *CoRR*, (i).

Torregrosa, D., Pérez-Ortiz, J. A., & Forcada, M. (2016). Ranking suggestions for black-box interactive translation prediction systems with multilayer perceptrons. In *Proceedings of AMTA 2016 - MT Researcher's Track*. Austin, Texas.

Toury, G. (1995). *Descriptive Translation Studies – and beyond*. Amsterdam/Philadelphia: John Benjamins.

Trochim, J. (2006). The Research Methods Knowledge Base. Retrieved 20 January 2017, from http://www.socialresearchmethods.net/kb/index.php.

Trujillo, A. (1999). *Translation engines: techniques for machine translation*. Berlin: Springer Verlag.

Turchi, M., Goutte, C., & Cristianini, N. (2012). Learning Machine Translation from In-domain and Out-of-domain Data. In *Proceedings of the 16th EAMT Conference* (pp. 305–312). Trento, Italy.

Turchi, M., Negri, M., & Federico, M. (2015). MT Quality Estimation for Computer-assisted Translation: Does it Really Help? *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 530–535.

Turian, J. P., Shen, L., & Melamed, I. D. (2003). Evaluation of Machine Translation and its Evaluation. In *Proceedings of MT Summit IX*. New Orleans.

Tyers, F. M., Sánchez-Martínez, F., Ortiz-Rojas, S., & Forcada, M. (2010). Free/open-source resources in the Apertium platform for machine translation research and development. *The Prague Bulletin of Mathematical Linguistics*, (93), 67–76.

Vasconcellos, M. (1987a). A Comparison of MT Postediting and Traditional Revision. In K. Kummer (Ed.), *Proceedings of the 28th Annual Conference of the American Translators Association*. Albuquerque: Learned Information, Inc.

Vasconcellos, M. (1987b). Post-editing on-screen: machine translation from Spanish into English. In C. Picken (Ed.), *Proceedings of the conference Translating and the Computer 8* (pp. 133–146). London: Aslib.

Vashee, K. (2016a). The Concept of MT Maturity. Retrieved 5 November 2016, from http://kv-emptypages.blogspot.pt/2016/04/the-concept-of-mt-maturity.html.

Vashee, K. (2016b). The Google Neural Machine Translation Marketing Deception. Retrieved 30 December 2016 from http://kv-emptypages.blogspot.com.br/2016/09/the-google-neural-machine-translation.html?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+EmptyPages+%28eMpTy+Pages%29.

Vermeer, H. J. (1986). *Esboços de uma Teoria da Tradução*. Porto: ASA.

Vieira, L. N., & Specia, L. (2011). A Review of Translation Tools from a Post-Editing Perspective. *Proceedings of the Third Joint EM+/CNGL Workshop Bringing MT to the Users: Research Meets Translators (JEC'11)*.

Vinay, J. P., & Darbelnet, J. (1977). *Stylistique Comparée du Français e de l'Anglais - Méthode de Traduction - Nouvelle édition revue et corrigée*. Paris: Didier.

Wagner, E. (1985). Post-editing Systran - A Challenge for Commission Translators. *Terminologie et Traduction*, *3*.

Wallis, J. (2006). *Interactive Translation vs Pre-translation in the Context of Translation Memory Systems: Investigating the effects of translation method on productivity, quality and translator satisfaction*. University of Ottawa.

Wallis, J. (2008). Interactive Translation vs. Pre-Translation in TMs: A Pilot Study. *Meta: Journal des Traducteurs*, *53*(3), 623–629.

Way, A. (2009). A Critique of Statistical Machine Translation. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, (8), 1–26.

Way, A. (2010a). Machine Translation. In A. Clark, C. Fox, & S. Lappin (Eds.), *The Handbook of Computational Linguistics and Natural Language Processing*. Chichester: Wiley-Blackwell.

Way, A. (2010b). Panning for EBMT gold, or 'remembering not to forget'. *Machine Translation*, *24*(3–4), 177–208. https://doi.org/10.1007/s10590-010-9085-2.

Way, A., & Hearne, M. (2011). On the Role of Translations in State-of-the-Art Statistical Machine Translation. *Linguistics and Language Compass*, *5*(5), 227–248.

Weaver, W. (1949). *Translation*. Retrieved 30 December 2017 from http://www.mt-archive.info/Weaver-1949.pdf.

Wiggins, D. (2011). Analysis of the Shutdown Announcements of the Google Translate API. Retrieved 5 November 2016, from http://kv-emptypages.blogspot.pt/2011/06/analysis-of-shutdown-announcements-of.html.

Wilks, Y. (2009). *Machine translation: its scope and limits*. Springer.

Wisniewski, G., Pécheux, N., & Yvon, F. (2015). Why Predicting Post-Edition is so Hard? Failure Analysis of LIMSI Submission to the APE Shared Task. In *Proceedings of the 10th Workshop on Statistical Machine Translation* (pp. 222–227).

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., … Dean, J. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *ArXiv E-Prints*, 1–23.

Wuebker, J., Green, S., DeNero, J., Hasan, S., & Luong, M.-T. (2016). Models and Inference for Prefix-Constrained Machine Translation. In *ACL 2016* (pp. 66–75).

Yamada, M. (2011). The effect of translation memory databases on productivity. *Translation Research Projects 3*, 63-70.

Zanettin, F. (2013). Corpus Methods for Descriptive Translation Studies. *Procedia - Social and Behavioral Sciences*, *95*, 20–32.

Zaretskaya, A., Pastor, G. C., & Seghiri, M. (2015). Integration of machine translation in CAT tools: State of the art, evaluation and user attitudes. *SKASE Journal of Translation and Interpretation*, *8*(1), 76–88.

Zeiler, M. D. (2012). ADADELTA: An Adaptive Learning Rate Method. *CoRR*.

Zerfass, A. (2002a). Comparing Basic Features of Translation Memory Tools. *Multilingual Computing and Technology - Language Technology Supplement*, *13*(7.

Zerfass, A. (2002b). Evaluating translation memory systems. In *LREC 2002: Language Resources in Translation Work and Research*. Las Palmas.

Zetzsche, J. (2015). Are we really so far behind? *TAUS Review*, (5), 25–26.

Zetzsche, J. (2016a). Neural Machine Translation. *The ATA Chronicle*.

Zetzsche, J. (2016b). The translator's perspective. *TAUS Review*, (6), 26–28.

Zetzsche, J. (2016c). Uplifting developments. *The Tool Box Journal*, (16-11–267).

Zhang, J., Utiyama, M., Sumita, E., Zhao, H., Neubig, G., & Nakamura, S. (2016). Learning local word reorderings for hierarchical phrase-based statistical machine translation. *Machine Translation*, *30*(1–2).

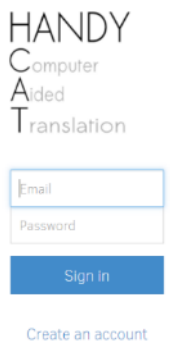
APPENDICES

Appendix 1: HandyCAT Quick Start Guide

(version porto_v0.1)

HANDYCAT – QUICK START GUIDE

HandyCAT version porto_v0.1



HandyCAT has a very simple interface, with some complex features beneath it. In this version, we have inserted a new editing mode, and this is the first time this is tested. The development time was very short, so please disregard the less mature interface issues.

HandyCAT is not a commercial tool. It was developed at Dublin City University by Chris Hokamp and it is available on GitHub, for research and development purposes. For the tests, we will be using a temporary server installed at TIPS, which will only be online for the duration of the workshop.

This guide presents a step-by-step introduction to HandyCAT, as it was setup for the tests during the Workshop “Tradução Automática Interativa e Pós-edição”.

You may follow this guide during the demonstration stage, by performing the sequence of operations highlighted in green text.

In this guide, all keyboard keys appear in **bold**, and on-screen buttons and new terms appear between “quotes”, at least for the first occurrence.

Starting a project

In the opening screen, insert your username and the password that was given to you.

Then click on the “Sign In” button.

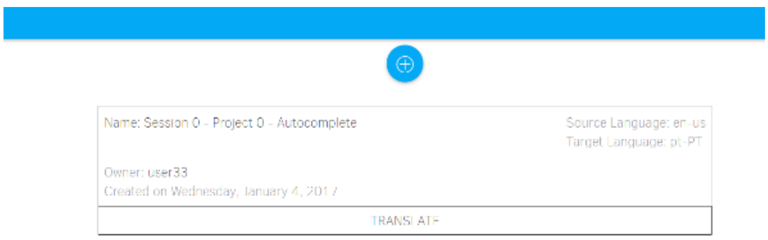
Click on the “plus” sign in the centre, to open the list of available projects: 

From the popup window, choose the project you want to work on:

Start by selecting the work session <u>in Autocomplete mode</u> in the “**Demo**” group:

“**1: Session 0 – Project 0 – Autocomplete**”.

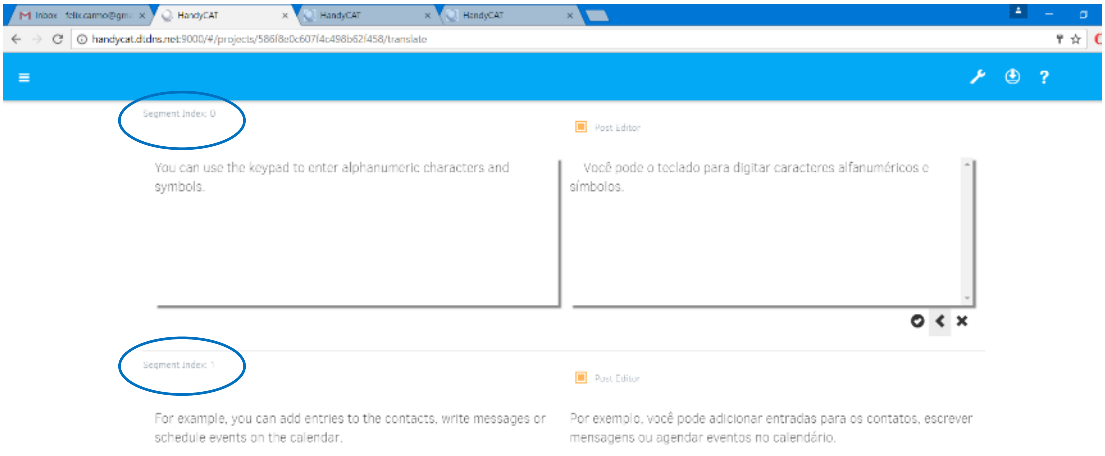
When you click on the name of the project, a new window opens, showing your projects, each in a different table. Click on the “Translate” button, below the first project (it becomes green when you hover over it with the mouse).



First look at the editor

The editor interface is very clean, with two panes for each segment: on the left, the “source window”, and on the right, the “target window”.

Please note that each segment is numbered, above the source window:



All editing is done inside the “target window”.

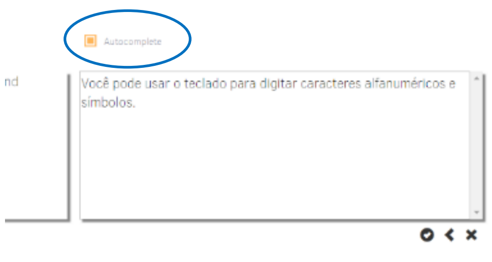
When you open a project, each target window already contains the “translation hypothesis”, which is the translation suggestion created previously by a Machine Translation system. We call this the “MT hypothesis”. Our work is to edit and correct this MT hypothesis.

Once you click inside the target window, it enters “edit mode” and three buttons appear below the window. We will analyse these later in context.

At the top bar, on the left, there is a “menu” button and, on the right, three buttons. We will only use the “menu” button, and this will be explained later.

Confirm that the project opened in the correct mode.

This appears above the “target window” of each segment, on the right.



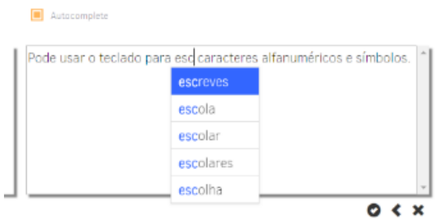
Now, let us look at how each editing mode works.

Autocomplete mode

This mode has no restrictions in terms of the usual navigation and editing in a text box. You can select and delete a word, drag and drop, move back and forward with the arrow keys, and so on.

When you start typing, a popup menu comes up with predictive writing suggestions, based on the letters you type. Check those suggestions and analyse how this interactive mode works. At the end, we want to know your opinion about it and how it compares to the "Post-editor" mode.

Test this mode, by following the instructions below to write or edit a couple of words on the first segments of the file.



If you want to accept the typing suggestion that is highlighted in the popup menu, press **Enter**.

You can scroll down the list of suggestions (with the mouse or the arrow keys) to choose a different suggestion.

To ignore the suggestions, you may keep on typing.
Check that the suggestions are updated as you type more characters. So, even if the first list of suggested words does not include the word you want, this may appear later.

Once you have finished typing a word and you press the **spacebar**, the popup menu will close.

And at any time you want to close the popup menu, press **Escape**.

When you have finished editing the segment, you want to save it. To do that, press **Ctrl+Enter**.

In this mode, the three buttons below the target window may be used to perform the following actions:

"Confirm" button: confirms the segment and saves the changes. The same as pressing **Ctrl+Enter**.

Please note that the translation you saved replaces the MT hypothesis. This means that once you confirmed a segment, you cannot retrieve its MT hypothesis.

"Back" button": in Autocomplete mode, it cancels all your actions.

It retrieves the MT hypothesis, or the translation that you confirmed previously.

To cancel your actions one by one in Autocomplete mode, press **Ctrl+Z** instead.

"Clear" button: it cleans the target window.

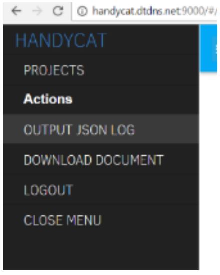
After the target window is empty, you may reinsert the MT hypothesis or the translation you saved by clicking on the "Back" button.

Make a couple of random edits in different segments and test the action of these buttons.

Saving your work

Once you finished each work session, click on the "Menu" button at the left top corner of the HandyCAT window.

Then click on "OUTPUT JSON LOG".



This will save a file with the name: "edit-log.json" to the "Downloads" folder.

PLEASE NOTE that while doing the experiments, you should not close the browser window, or click on "LOGOUT".

To start a new session, click on "PROJECTS". This will take you back to the list of projects. Click on the plus sign at the centre, to see the list of available projects. Choose the next session of work.

Choose the project in "Post-editor mode" in the "**Demo**" group:

"**0: Session 0 - Project 0 - Post Editor**".

The newly selected project will be the first in the page with your projects. Check that you are in the right project before starting your work session.

Post-editor mode

This mode is very restricted in terms of navigation and editing, and you may feel uncomfortable with it at first. You will notice that you cannot type over the MT hypothesis, delete words, or use any of the usual navigation keys inside the target window.

This was designed as a research interface, which forces users to think about each action separately, and be aware of the contexts in which the 4 editing actions are applicable.

All actions involve two-steps:

- selecting a word, and then
- choosing an action from the popup menu that appears.

Let us begin by the “select” step.

In the open segment, select one word.

You will see a black box around it. This is the “edit box”. All actions are applied to the content of the edit box.

You will also see a popup menu with the 4 editing actions, but disregard it for now. Do not select any of the actions for now. Instead, press **Escape**.

Escape cancels and closes the “edit box”.

Now, click and select several words, one at a time.

Notice that the colours of the edit boxes around the words change as you hover over them or select them:

- Clear blue – word over which the mouse is hovering
- Dark blue – word currently selected
- Black – words previously selected

Please also notice that spaces and punctuation may be selected, just like words.

Now, select a full sequence of words, by clicking on each word, then the space after it, and only then the next word.

Please note that selecting more than one word is especially useful with the “move” action.

You will also notice that there are extra spaces at the beginning and the end of each segment, while you are editing a segment. These were inserted to enable you to work beyond the first and the last character in each segment. Disregard these, as they disappear once you confirmed the segment.

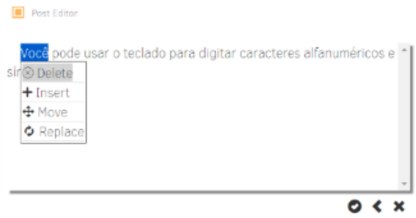
To cancel the selection, you press **Escape**.

Test making different selections and then cancelling them.

How to apply the four editing actions

Let us now see how each action button in the popup menu works, one by one.

Delete – Use this action button in the popup menu to delete words in positions where you will not insert other words.



Click on the word you want to delete. The popup menu comes up. Click on "Delete".

When you delete a word, spaces are adjusted, so you do not need to worry about deleting a space. To delete more than one word, you only need to click on each word, and spaces between them are also deleted.

Test deleting more than one word.

Finalising each segment

Please note that these actions are applied for any editing action or sequence of editing actions performed.

If you selected the wrong word or words, press **Escape** to cancel the selection, before selecting an action button in the popup menu.

When you have finished editing the segment, press **Ctrl+Enter** to save it, or use the "Confirm" button.

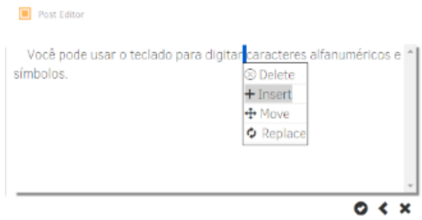
If you have applied the editing action to the wrong word or words, you need to use the "Back" button: (below the target window).

<u>Please note that, in Post-editor mode, **Ctrl+Z** does not work and the "Back" button backtracks one editing action at a time.</u>

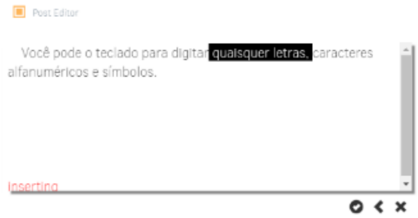
If you want to restart editing the segment, click on the "Clear" button below the target window. You may then retrieve either the MT hypothesis or the translation you saved, by clicking on the "Back" button.

Test these finalisation actions and buttons with deleted words.

Insert – Use this action button to insert words in positions where they are missing.



Click on a space between words, or on the space at the beginning or at the end of a sentence. The popup menu comes up. Click on "Insert". An edit box appears in that space and you enter "insert mode".



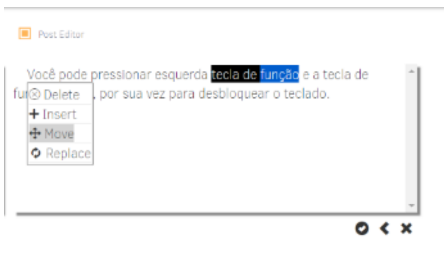
When "insert mode" is active, the word "inserting" appears in red at the bottom of the target window, and the cursor flashes inside the black edit box. You may start typing one or more words, insert spaces and punctuation, delete characters and move freely inside the edit box.

Do not click in the edit box after you selected "Insert". This will disable "insert mode".

To close "insert mode", you need to press **Escape**.

Test inserting words and closing "insert mode".

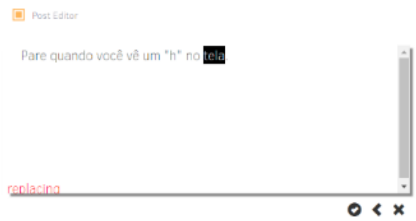
Move – This action is useful when we have the right words in the wrong position in the MT hypothesis. (We could "delete" and "insert", but we may do this in one action.)



Click on a word – the popup menu appears. Click on "Move". Click on a space between words and the selected phrase appears in the correct position.

Test moving one word and then moving a sequence of words. Please note that you need to select the spaces between the words.

Replace – This is used to make small edits to words, like adding an "s", or inserting a capital letter, or simply writing a different word in the same position as a word in the MT hypothesis.



Click on a word. The popup menu appears. Click on "Replace". You enter "replace mode": the word "replacing" appears at the bottom of the window and the cursor flashes inside the edit box. Start typing, delete characters, and use the arrow keys freely to move inside the edit box.

Please note that:

a) This action only works with one word at a time – you cannot "replace" several words;
b) you cannot navigate after the last character in the black box, or you will lose the cursor;
c) do not click in the edit box while in "replace mode", as this will disable it.

In all these cases, press **Escape** to close "replace mode" and restart selecting words.

Test replacing and editing words.

Once you have finished the editing session in this mode, save the log following the instructions above.

Appendix 2: MateCAT processing of source texts

Date:	Tue, 01 Nov 16 20:14:36
Project:	Create_XLIFF
Language direction:	en-US > pt-PT

File:	**TEXT A - Mobile phone instructions**		**TEXT B - Marketing questionnaire**		**TEXT C - Product catalogue**		**TEXT D - Instructions manual**		**4 files**	
Match Types	Words	Percent	Words	Percent	Words	Percent	Words	Percent	Words	Percent
New words	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Context Match	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Repetitions	**0**	**0.00**	**19**	**2.26**	**32**	**3.29**	**10**	**1.99**	**61**	**2.00**
Format Change	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Internal Match	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
100%	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
100% Public TM	8	1.09	37	4.41	3	0.31	13	2.58	61	2.00
95% - 99%	**19**	**2.59**	**0**	**0.00**	**2**	**0.21**	**19**	**3.78**	**40**	**1.31**
85% - 94%	**8**	**1.09**	**30**	**3.58**	**0**	**0.00**	**47**	**9.34**	**85**	**2.79**
75% - 84%	**15**	**2.05**	**31**	**3.69**	**43**	**4.41**	**0**	**0.00**	**89**	**2.92**
50% - 74%	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
MT	683	93.18	722	86.05	894	91.79	414	82.31	2713	88.98
Total Payable	574	78.31	631	75.21	752.7	77.28	377.7	75.09	2335.4	76.60
Total	**733**	**100.00**	**839**	**100.00**	**974**	**100.00**	**503**	**100.00**	**3049**	**100.00**

Source: https://www.matecat.com/analyze/Create_XLIFF/481455-787fb959b42f

Productivity analysis:
MateCAT considers a metric of 3000 words/day (8 hours/day).
The total wordcount is 3049 words and the total "weighted" wordcount is 2335 words ("Total payable" row above).
Weighted wordcounts are explained here:
https://www.matecat.com/support/introducing-matecat/matecat-calculates-payable-words/
According to this, using MateCAT to PE this project would yield a gain of over 2 hours in productivity.

Breakdown of weighted wordcounts:	
No TM match	100%
Machine Translation	80%
Lower fuzzy ranges (50-74%)	100%
Higher fuzzy ranges (75-99%)	**60%**
100% match	30%
Repetition	30%
Context match	0%

TEXT A – INSTRUCTIONS – MOBILE PHONE

Segment ID	Source segment	Target MT hypotheses
1	You can use the keypad to enter alphanumeric characters and symbols.	Você pode usar o teclado para digitar caracteres alfanuméricos e símbolos.
2	For example, you can add entries to the contacts, write messages or schedule events on the calendar.	Por exemplo, você pode adicionar entradas para os contatos, escrever mensagens ou agendar eventos no calendário.
3	You can scroll through this list and select the character you want.	Você pode percorrer esta lista e selecione o caractere desejado.
4	Stop when you see an "h" on the screen.	Pare quando você vê um "h" na tela.
5	After entering the first letter, you can directly press another key (except Send) to enter the next letter.	Depois de inserir a primeira letra, você pode pressionar diretamente outra tecla (exceto Enviar) para inserir a próxima letra.
6	If the next desired letter is on the same key as the current one, wait until the cursor appears to the right of the current letter and then you can enter the next one.	Se a próxima letra desejada estiver na mesma tecla que a actual, aguarde até que o cursor aparece à direita da letra atual e, em seguida, você pode digitar o próximo.
7	You can access the Settings screen by selecting Settings from the main menu.	|||UNTRANSLATED_CONTENT_START|||You can access the Settings screen by selecting Settings from the main menu.|||UNTRANSLATED_CONTENT_END|||
8	Phone settings	Configurações de Telefone...
9	You can change the settings of the phone.	Você pode alterar as configurações do telefone.
10	Date and Time:	Data e hora:
11	To set the system date and time format of the phone.	Para definir a data do sistema e formato da hora do telefone.
12	Profiles:	Perfis:
13	To select different profiles for the phone to suit different environmental situations.	Para selecionar diferentes perfis para o telefone para atender às situações ambientais diferentes.
14	Phone language:	Idioma do telefone:
15	To set the phone language.	Para definir o idioma do telefone.
16	This option is used to set the phone language to be the same as that used in your SIM card.	Esta opção é usada para definir o idioma do telefone a ser o mesmo que o utilizado no seu cartão SIM.
17	You can also select Phone settings / Profiles to set the sounds for the phone.	Você também pode selecionar Configurações do telefone / perfis para definir os sons para o telefone.
18	Keypad lock period:	período de bloqueio do teclado:
19	You can set the keypad lock period.	Você pode definir o período de bloqueio do teclado.
20	If there is no operation on the phone within a preset period of time, the keypad is locked automatically.	Se não houver nenhuma operação no telefone dentro de um período de tempo predefinido, o teclado é bloqueado automaticamente.
21	You can press left function key and right function key in turn to unlock the keypad.	Você pode pressionar esquerda tecla de função e a tecla de função direita, por sua vez para desbloquear o teclado.

22	Shortcuts:	Atalhos:
23	To assign the frequently used functions to the scroll keys as shortcut keys.	Para atribuir as funções utilizadas com frequência às teclas de deslocação como teclas de atalho.
24	Auto On and OFF:	Auto On e OFF:
25	To set your phone to power on or off automatically.	Para configurar o telefone para ligar ou desligar automaticamente.
26	In the standby mode, you can press and hold to switch the current profile to the silent profile.	No modo de espera, você pode pressionar e segurar para mudar o perfil atual para o perfil silencioso.
27	If the current profile is offline, you cannot use this function.	Se o perfil atual está offline, você não pode usar esta função.
28	Display settings	Configurações de exibição
29	Then you can change the call barring password.	Em seguida, você pode alterar o barramento de chamadas.
30	The old call barring password is provided by your network operator if it is the first time for you to change the password.	A senha de restrição de chamadas de idade é fornecido pelo operador de rede se é a primeira vez para você mudar a senha.
31	Call waiting	Chamada em espera
32	This option is used to enable or disable the call waiting function.	Esta opção é usada para ativar ou desativar a função de chamada em espera.
33	The call waiting function is network-dependent.	A função de chamada em espera é dependente da rede.
34	For more information, contact your local Vodafone customer care center.	Para mais informações, contacte o centro de atendimento ao cliente da Vodafone local.
35	Save unknown number	Guardar número desconhecido
36	This function is used to prompt whether to save the number after a call ends.	Esta função é usada para perguntar se pretende guardar o número depois de uma chamada termina.
37	Auto redial	Perguntas freqüentes sobre rediscagem automática
38	This function is used to auto redial the dialed number if the dialing failed.	Esta função é usada para auto remarcar o número marcado caso a marcação falhou.
39	Answer mode	Resposta
40	It is used to set the mode for answering an incoming call.	Ele é usado para definir o modo de atendimento de uma chamada recebida.
41	Three options are available:	Estão disponíveis três opções:
42	Any key, Send key and Slide up.	Qualquer tecla, Tecla de envio e Deslize para cima.
43	Auto Answer	Atendimento Automático
44	This option is used to select auto answering for incoming calls.	Esta opção é usada para selecionar auto atendimento para as chamadas recebidas.
45	Four options are available:	Estão disponíveis quatro opções:

TEXT B – QUESTIONNAIRE – MARKETING

Segment ID	Source segment	Target MT hypotheses
1	Which of the following types of devices do you OWN AND USE?	Qual dos seguintes tipos de dispositivos possui ou usa?
2	Please select all that apply	Selecione todas as opções que se aplicam
3	And which of the following types of devices do you currently USE TO ACCESS THE INTERNET at home, work, or elsewhere (including devices you do not personally own, at work or in a public place e.g. library or internet café)?	E qual dos seguintes tipos de dispositivos que você usa atualmente para acessar a Internet em casa, no trabalho ou em outros lugares (incluindo os dispositivos que você não possui, pessoalmente, no trabalho ou em um ou seja, bibliotecas lugar público ou internet café)?
4	We would now like to ask you about your spending habits.	Gostaríamos agora de lhe perguntar sobre seus hábitos.
5	Please be assured that all information provided will be anonymous and used solely for the purposes of this research and no personally identifiable information will be made available to third parties.	Tenha certeza de que todas as informações fornecidas serão anônimas e utilizados exclusivamente para os fins desta pesquisa e nenhuma informação pessoal será disponibilizada a terceiros.
6	This is the insert marker used for the Q6 question	Este é o marcador de inserção utilizados para a questão Q6
7	Please remember to enter your answers in thousands of <tag>.	Por favor, lembre-se de inserir as respostas em milhares de <tag>.
8	For example, if your total spend is 12 111 <tag>, please type '12' in the box.	Por exemplo, se o seu gasto total é de 12 111 <tag>, por favor, escreva '12' na caixa.
9	This is the insert marker used for the Q6 question	Este é o marcador de inserção utilizados para a questão Q6
10	Please type in your total spend to the nearest one thousand (1,000) <tag>.	Por favor, digite o seu gasto total para a mais próxima mil (1.000) <tag>.
11	Please type in the closest full number to your total spend in <tag>.	Por favor, digite o número total mais próximo do seu gasto total em <tag>.
12	For example, if your total spend is 12.99, please type in 13.	Por exemplo, se seu gasto total é de 12,99, por favor digite 13.
13	Thinking about each of the categories below, please estimate your total spend in each category over the past 12 months:	Pensando em cada uma das categorias abaixo, por favor, estimar o seu gasto total em cada categoria ao longo dos últimos 12 meses:
14	A) For purchases made online.	A) Para compras feitas online.
15	B) For purchases made offline.	B) Para compras feitas offline.
16	This includes food, clothing and other items but excludes rent or mortgage payments, utility bills (gas, electric, council tax) and large one off payments (e.g. buying a new car).	Isso inclui alimentos, roupas e outros itens, mas exclui aluguel ou hipoteca pagamentos, contas de serviços públicos (gás, energia elétrica, imposto municipal) e grandes um off pagamentos (por exemplo, a compra de um carro novo).
17	If you're unsure, please provide your best guess.	Se não tiver certeza, por favor fornecer o seu melhor palpite.
18	If you don't have any spending for the following categories, enter nothing or 0.	Se você não tem qualquer despesa para as seguintes categorias, digite nada, ou 0.
19	By online we mean any purchase made on the internet via a desktop / laptop / notebook / Smartphone or Tablet.	Por on-line que qualquer compra feita na internet através de um desktop / laptop / notebook / Smartphone ou Tablet.

20	By offline we mean any purchase made in a physical store.	Por desligada queremos dizer qualquer compra feita em uma loja física.
21	Please provide an answer.	Por favor, forneça uma resposta.
22	The spending across all devices should be equal to the total amount shown	Os gastos em todos os dispositivos deve ser igual ao valor total mostrado
23	You indicated that your total annual online spend in the last 12 months is as follows.	Você indicou que seus gastos on-line total anual nos últimos 12 meses é a seguinte.
24	How does this split by device?	Como é que esta divisão por dispositivo?
25	Total online spend in the last 12 months	gastos online total nos últimos 12 meses
26	Spend on purchases made via a desktop / laptop / notebook	Gastar em compras feitas através de um desktop / laptop / notebook
27	Spend on purchases made via a Smartphone	Gastar em compras feitas através de um Smartphone
28	Spend on purchases made via a Tablet	Gastar em compras feitas através de um Tablet
29	Spend on purchases made via other devices (e.g. Games consoles, Smart TVs)	Gastar em compras feitas através de outros dispositivos (por exemplo, consolas de jogos, Smart TVs)
30	The total should be equal to 100%	O total deve ser igual a 100%
31	Of all the money you spend online currently, approximately what proportion do you spend in each of the following ways?	De todo o dinheiro que você gasta on-line atualmente, aproximadamente qual a proporção que você gasta em cada uma das seguintes maneiras?
32	Please provide your best guess if you are not sure	Por favor, forneça o seu melhor palpite se você não tem certeza
33	Please enter the percentage (%) of your current spending that falls into each category	Por favor, indique a percentagem (%) de seus gastos corrente que cai em cada categoria
34	On dedicated shopping websites via an internet browser	Em sites de compras dedicados através de um navegador de internet
35	Through a dedicated shopping App on a smartphone or tablet	Através de um dedicado App compras em um smartphone ou tablet
36	Through social media sites (e.g. buyable pins on Pinterest, In-tweet purchases on Twitter, WeChat purchases etc.)	Através de sites de mídia social (por exemplo, pinos compráveis no Pinterest, compras In-tweet no Twitter, WeChat compras etc.)
37	Other	Outros
38	Total amount entered	montante total inscrito
39	Please enter your spending in all categories requested.	Por favor, indique os seus gastos em todas as categorias solicitadas.
40	The amount allocated in the columns 3 and 4 should be equal to 100.	O montante atribuído nas colunas 3 e 4 deve ser igual a 100.
41	ONLINE % of spend - last 12 months	% ONLINE de gastos - últimos 12 meses
42	OFFLINE % of spend – last 12 months	% OFFLINE dos gastos - últimos 12 meses
43	ONLINE % of spend – next 12 months	% ONLINE de gastos - próximos 12 meses
44	OFFLINE % of spend – next 12 months	% OFFLINE dos gastos - próximos 12 meses
45	How do you think the overall amount of your online spending will change over the next 12 months?	Como você acha que a quantidade total de seus gastos on-line irá mudar ao longo dos próximos 12 meses?

TEXT C – PRODUCT CATALOG – OFFICE SUPPLIES

Segment ID	Source segment	Target MT hypotheses
1	PC and plastic surface cleaning wipes	toalhetes de limpeza de PC e de superfície de plástico
2	Individually packed for extra convenience	Embalado individualmente para maior comodidade
3	Duo detachable tearaway packaging	embalagem tearaway destacável Duo
4	Anti-static coating after cleaning	revestimento anti-estático após a limpeza
5	Each pack contains 20 pairs	Cada embalagem contém 20 pares
6	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza
7	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.
8	The <tag> Anti-Static Cleaning Cloths are ideal for fast cleaning of telephone sets, keyboards and other hard surfaces.	A <tag> Anti-estático panos de limpeza são ideais para limpeza rápida dos aparelhos telefónicos, teclados e outras superfícies duras.
9	With its anti-static properties the wipes ensures that dust does not immediately settle again after cleaning.	Com suas propriedades anti-estáticas os toalhetes garante que a poeira não resolve volte imediatamente após a limpeza.
10	The easy-to-use wipes are individually packed which makes it easier to carry around in your bag, case or pocket, and prevents the wipes from drying out if left unused over time.	Os toalhetes de fácil utilização são embalados individualmente, o que torna mais fácil de transportar no seu saco, caso ou bolso, e impede que os toalhetes de secar se não forem utilizadas ao longo do tempo.
11	Each package contains 5 pre-moist wipes.	Cada pacote contém 5 toalhetes pré-húmidas.
12	Individually packed pairs for added convenience	pares embalados individualmente para maior comodidade
13	Designed for cleaning telephone sets, keyboards and other hard surfaces	Projetado para aparelhos telefónicos de limpeza, teclados e outras superfícies duras
14	Pre-moist wipes	toalhetes pré-húmidas
15	Wipes are 100% recyclable	Wipes são 100% recicláveis
16	5 wipes per pack	5 toalhetes por pacote
17	Anti-Static Cleaning Cloth for Telephones, Keyboards and Hard Surfaces	Pano antiestático de limpeza para telefones, teclados e Hard Surfaces
18	Thoroughly clean telephone sets, keyboards and other hard surfaces with the <tag> Anti-Static Cleaning Cloths.	conjuntos completamente limpas de telefone, teclados e outras superfícies duras com a <tag> Anti-estático panos de limpeza.
19	This premium quality <tag> transport and bundling tape comprises a very powerful synthetic rubber adhesive that will bond well to all types of smooth or uneven surfaces including steel, plastic and cardboard.	Esta qualidade superior <tag> transporte e agregação de fita compreende um adesivo de borracha sintética muito poderosa que irá ligar bem a todos os tipos de superfícies lisas ou irregulares, incluindo aço, plástico e papelão.
20	This very reliable and high performing fibreglass packaging tape features glass filaments laminated onto polyester film.	Este desempenho fita de embalagem de fibra de vidro muito confiável e de alta possui filamentos de vidro laminado em película de poliéster.
21	Used primarily to seal heavy duty cartons, securely fasten pallets or to bundle loose items together, <tag> tape is a durable, multipurpose and secure solution that is tear resistant.	Usado principalmente para selar embalagens pesados, prenda paletes ou para agrupar os itens soltos juntos, <tag> fita é uma solução durável, polivalente e seguro que é resistente ao desgaste.
22	It is also ideal for transport securing, fastening components and end-tabbing metal bands and coils.	Também é ideal para proteger o transporte, componentes de fixação e bandas de metal-tabulações finais e bobinas.

23	Each transparent roll of tape measures 19 mm x 50 m.	Cada rolo de fita adesiva transparente mede 19 mm x 50 m.
24	Used for sealing heavy duty cartons, fastening pallets or bundling of heavy items	Usado para selar embalagens pesados, paletes de fixação ou a agregação de itens pesados
25	Durable, tear-resistant fibreglass backing material	, Material de fibra de vidro revestimento protetor resistente ao desgaste durável
26	Very high tensile strength	Muito alta resistência à tracção
27	Very low elongation	Alongamento muito baixo
28	High-quality synthetic rubber adhesive	adesivo de borracha sintética de alta qualidade
29	Use the heavy duty <tag> 04590 Monofilament Tape to securely seal boxes, fasten pallets or bundle loose items with its tear-resistance and high tensile strength.	Use o serviço pesado Tape <tag> 04.590 monofilamento para selar firmemente caixas, paletes prender ou agrupar os itens soltos com a sua lágrima-resistência e alta resistência à tração.
30	The <tag> Pack 'n Go hand dispenser is easy to use and is ideal for the casual packager.	A <tag> Pack 'n dispensador de mão Go é fácil de usar e é ideal para o empacotador casual.
31	Its sturdy metal frame and sharp cutter is complemented with a flexible tape-pressing wiper and a moulded plastic hand grip.	Sua armação de metal resistente e cortador afiado é complementada com um limpador de pára-pressionando fita flexível e um aperto de mão de plástico moldado.
32	Easy to hold with the ergonomically designed grip, this <tag> hand dispenser is an ideal accompaniment for securely sealing shipping cartons and general packages.	Fácil de segurar com o aperto ergonomicamente projetado, este distribuidor <tag> mão é um acompanhamento ideal para a vedação segura caixas de transporte e pacotes gerais.
33	Each dispenser is able to accept tape rolls to a maximum width of 50 mm and maximum length of 66 m.	Cada embalagem é capaz de aceitar rolos de fita para uma largura máxima de 50 mm e comprimento máximo de 66 m.
34	Tape dispenser for packing	dispensador de fita para a embalagem
35	Seal shipping cartons and general packages quicker and easier	caixas de transporte Seal e pacotes gerais mais rápidas e mais fáceis
36	Solid metal frame	estrutura de metal sólido
37	Ergonomically designed hand grip	aperto de mão ergonomicamente projetado
38	Sharp cutter with replaceable blade	cortador afiado com lâmina substituível
39	Rubber roller and flexible tape-pressing wiper makes it easy to use	rolo de borracha e limpador de prensagem fita flexível faz com que seja fácil de usar
40	Featuring an easy-to-use grip and solid metal frame, the <tag> Pack 'n Go Hand Tape Dispenser is the ideal accompaniment for all your tape packaging needs.	Apresentando um aperto fácil de usar e estrutura de metal sólida, o <tag> Pack 'n Go Mão Tape Dispenser é o acompanhamento ideal para todas as suas necessidades de embalagem de fita.
41	Pack 'n Go Hand Tape Dispenser Blue 171 x 68 x 115 mm	Pack 'n Go Mão Tape Dispenser azul 171 x 68 x 115 mm
42	Featuring an easy-to-use grip and solid metal frame, the <tag> Pack 'n Go Hand Tape Dispenser is the ideal accompaniment for all your tape packaging needs.	Apresentando um aperto fácil de usar e estrutura de metal sólida, o <tag> Pack 'n Go Mão Tape Dispenser é o acompanhamento ideal para todas as suas necessidades de embalagem de fita.
43	The <tag> hand dispenser features an easy-to-use grip, fully adjustable tension control and a rubber roller for easy application of all standard packaging tapes.	O distribuidor <tag> mão apresenta um aperto fácil de usar, controle de tensão totalmente ajustável e um rolo de borracha para facilitar a aplicação de todas as fitas de embalagem padrão.
44	It also comes with 2 rolls of polypropylene packaging tape to make one handy and economical package.	Ele também vem com 2 rolos de fita de embalagem de polipropileno para fazer um pacote acessível e econômica.

TEXT D – INSTRUCTION MANUAL – INDUSTRIAL EQUIPMENT

Segment ID	Source segment	Target MT hypotheses
1	Preface	Prefácio
2	This document provides the safety requirements for all <tag> products, including:	Este documento fornece os requisitos de segurança para todos <tag> produtos, incluindo:
3	Some information in this document does not apply to all <tag> equipment.	Algumas informações contidas neste documento não se aplica a todos <tag> equipamento.
4	Refer to the information that applies to the <tag> equipment at your facility and disregard those items that do not apply.	Consulte as informações que se aplica ao <tag> equipamentos em suas instalações e desconsiderar os itens que não se aplicam.
5	The user guides for <tag> products are provided as separate documents.	Os guias do usuário para <tag> produtos são fornecidos como documentos separados.
6	For safety information about products not covered in this document, call the Technical Support Center for <tag> products at	Para obter informações de segurança sobre produtos não abrangidos neste documento, ligue para o Centro de Suporte Técnico para <tag> produtos em
7	<tag> if calling within the United States	<Tag>, se chamando dentro dos Estados Unidos
8	<tag> if calling outside the United States	<Tag>, se chamando fora dos Estados Unidos
9	Defined Terms	Termos Definidos
10	The following table identifies the defined terms used throughout this document for certain products and product features.	A tabela a seguir identifica os termos definidos utilizados ao longo deste documento para determinados produtos e recursos do produto.
11	Product/Feature	Característica do produto
12	Defined Term	Termo definido
13	Equipment Markings and Symbols	À marcação dos equipamentos e símbolos
14	Several symbols are used to convey visual messages in this manual and on the equipment.	Vários símbolos são usados para transmitir mensagens visuais neste manual e no equipamento.
15	The symbols follow the International Organization for Standards (ISO) format.	Os símbolos seguem a Organização Internacional de Normalização (ISO) formato.
16	General hazard—Loss, injury, or damage may occur.	perigo da perda geral, podem ocorrer lesões ou danos.
17	Refer to the accompanying documents or the adjacent text.	Consulte os documentos de acompanhamento ou o texto adjacente.
18	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.
19	Refer servicing to qualified service personnel.	Consulte os serviços de manutenção com pessoal de manutenção qualificado.
20	Tip Over Hazard—Moving equipment requires special attention to avoid tipping it over and injuring an individual.	Derrubar sobre os perigos Movendo equipamento requer atenção especial para evitar a sua queda e ferindo uma pessoa.
21	On Position—Number 1 or letter I. Indicates the switch position for turning on the power supply.	Em Posição-número 1 ou a letra I. Indica a posição do interruptor para ligar a fonte de alimentação.
22	For push-button switches, the following symbol may also be used:	Para interruptores de premir, pode também ser utilizado o seguinte símbolo:
23	Off Position—Number 0 or letter O. Indicates the switch position for turning off the power supply.	Off Position-número 0 ou letra O. Indica a posição do interruptor para desligar a fonte de alimentação.
24	For push-button switches, the following symbol may also be used:	Para interruptores de premir, pode também ser utilizado o seguinte símbolo:
25	Directive 2002/96/EC—Waste electrical and electronic equipment; not for municipal waste.	Directiva 2002/96 / EC-Resíduos de equipamentos eléctricos e electrónicos; não para resíduos urbanos.

26	Directive 2002/95/EC—Restriction of the use of certain hazardous substances in electrical and electronic equipment.	Directiva 2002/95 / EC-Restrição do uso de determinadas substâncias perigosas em equipamentos eléctricos e electrónicos.
27	Maximum Weight—Maximum weight allowed for drawer contents.	peso-Peso Máximo máximo permitido para conteúdo da gaveta.
28	Do not overload the drawer.	Não sobrecarregue a gaveta.
29	Terminology	Terminologia
30	The following terms are used throughout this document to identify <tag> components.	Os seguintes termos são utilizados ao longo deste documento para identificar <tag> componentes.
31	Station—Identifies a unit with or without drawers, doors, <tag> technology, or a computer-based system with a screen and keyboard.	Station-Identifica uma unidade com ou sem gavetas, portas, <tag> tecnologia, ou de um sistema baseado em computador com uma tela e teclado.
32	Main—Identifies a station that has a computer system and screen for managing and tracking the items in an auxiliary station as well as in a main station.	Main-Identifica uma estação que tem um sistema de computador e tela para gerenciar e rastrear os itens em uma estação auxiliar, bem como em uma estação principal.
33	The main may be an open wire frame or closed metal frame construction.	O principal pode ser uma armação de arame aberta ou construção estrutura metálica fechada.
34	Auxiliary—Identifies a station that can be added to a main station to increase capacity.	Auxiliar-Identifica uma estação que pode ser adicionado a uma estação principal para aumentar a capacidade.
35	The auxiliary unit does not have a computer system.	A unidade auxiliar não tem um sistema de computador.
36	Conventions	Convenções
37	This document uses the following conventions:	Este documento utiliza as seguintes convenções:
38	Text	Texto
39	The names of document titles, cross-references, and text that requires emphasis are formatted in <pt40>italics.</pt40>	Os nomes dos títulos de documentos, referências cruzadas, e um texto que exige ênfase são formatados em <Italic>itálico.</Italic>
40	The names of buttons, menu commands, options, icons, file names, and folders are formatted in <pt42>bold</pt42>.	Os nomes dos botões, comandos de menu, opções, ícones, nomes de arquivos e pastas são formatados em <Bold>negrito.</Bold>
41	User input is formatted in <pt44>Courier bold</pt44>.	A entrada do usuário é formatado em <pt44>Courier negrito</pt44>.
42	Programming code is formatted in <pt46>Courier fixed width</pt46>.	código de programação é formatado em <pt46>largura fixa Courier</pt46>.
43	Icons	Ícones
44	This document uses the following symbols:	Este documento utiliza os seguintes símbolos:
45	Notes contain supplementary information or emphasize a point or procedure.	As notas contêm informações complementares ou enfatizar um ponto ou procedimento.

Appendix 3: Questionnaire 1

Workshop on Interactive Machine Translation and Post-Editing - Part 1

This questionnaire is composed of two parts: Part 1, which you should complete and submit as early as possible, but no later than January 5th, and Part 2, which will be completed at the workshop. Part 1 shouldn't take more than 10 minutes to complete.

Most questions do not have an open answer. Please choose the option that best describes your opinion, and then use the comments field at the end of this part of the questionnaire to explain some of the options you have chosen, or simply add your comments.

None of these answers will be traceable back to you. So, please give your opinions in full honesty.

Please check the final instructions at the end of this part of the questionnaire, which will allow us to join the two parts of the questionnaire that you filled in, without having to request any personal details.

*Obrigatório

Before we begin

The questions in this part will be used to associate your professional profile to the results of the practical experiments of the workshop.

1.

1 - For how long have you been a translator? *

Please insert the number of years only, even if it is an approximation.

....

2.

2 - Do you work as a freelancer, or are you employed at a translation agency, or a different company? *

Please register as "freelancer" if most of the translation work you do is at home, if you own your company but you work alone, or if you work for a single client, but you work from home and you have no established work hours.

Marcar apenas uma oval.

- Freelancer
- At a translation agency
- At a company not in the translation sector

3.

3 - Classify each of the tasks below in terms of the frequency in which you perform them. *

Marcar apenas uma oval por linha.

	Very frequent	Frequent	Irregular	Rare	Very rare
Translation into your mother tongue	◯	◯	◯	◯	◯
Translation into a second language	◯	◯	◯	◯	◯
Revision (of translations by other translators)	◯	◯	◯	◯	◯
Post-editing (of machine translation)	◯	◯	◯	◯	◯
Other language tasks	◯	◯	◯	◯	◯
Management tasks	◯	◯	◯	◯	◯

4.

4 - How would you describe your relationship with Information Technology for Translation? *

Marcar apenas uma oval.

◯ Very good - I have always enjoyed learning how to work with new tools and I like exploring new ways of doing my work

◯ Good - I can use the tools that I need in my daily life

◯ Reasonable - I use the tools because I must, but I don't like it when I am forced to change the way I work

◯ Bad - Tools are always changing and I do not have the time to adapt to changes

5.

5 - Do you prefer to type your translation from scratch or to type over the source text? *

Marcar apenas uma oval.

◯ I prefer to type from scratch - my target window is always empty when I start

◯ I prefer to type over the source - my target window contains the source text before I start

6.

6 - Do you prefer that your CAT tool presents several suggestions (from the Translation Memory, the termbase, autosuggest/predictive writing, etc.) as you are typing your translation, or that your editing window is as clean as possible? *

Marcar apenas uma oval.

◯ I prefer a clean editing enviroment, because suggestions usually interfere with my work

◯ I like it when my CAT tool gives me suggestions, but sometimes they are too intrusive, or not useful at all

◯ I always take advantage of all the suggestions my CAT tool gives me

7.
7 - Do you use the "predictive writing", "auto-suggest" or "auto-complete" features that CAT tools usually include? *
Marcar apenas uma oval.

- Very frequently
- Not usually
- Never

8.
8 - Do you use a Machine Translation system to help you translate? *
Marcar apenas uma oval.

- Very frequently *Passe para a pergunta 9.*
- Only in specific situations *Passe para a pergunta 9.*
- Never *Passe para a pergunta 10.*

Reasons for using machine translation as support

9.
8 - b) Please describe the situations in which you use machine translation as a support.

Post-editing

This section focuses on post-editing (correcting a text that has been previously translated by a Machine Translation system).

10.
9 - Please choose the expression that best describes your experience in post-editing. *
Please note that, from this point onwards, the questionnaire is different if you consider that you have experience in post-editing or not.
Marcar apenas uma oval.

- I have a lot of experience in post-editing
- I have some experience in post-editing
- I know what post-editing is, but I have never done it professionally *Passe para a pergunta 19.*

Experience with post-editing

11.

10 - How many words, on average, can you produce in post-editing projects per hour? *

Please give only a number, even if it is an approximation.

12.

11 - How many words do you do in post-editing projects, on average, per month? *

Think about how many words you post-edited last year.
Marcar apenas uma oval.

- No more than 1000 words
- Between 1000 and 10,000 words
- Between 10,000 and 25,000 words
- Between 25,000 and 50,000 words
- Above 50,000 words

13.

12 - How would describe your reaction to post-editing job assignments? *

Marcar apenas uma oval.

- I do not like post-editing - I cannot think like I am used to
- I like post-editing - I see things from a different perspective

14.

13 - How would you describe your speed and productivity in post-editing, as compared to translating? *

Marcar apenas uma oval.

- I work slower, because I do not trust the suggestions and I have to check everything
- I work faster, because I only worry about real errors

15.

14 - Does the fact that post-editing is paid at a lower rate affect your dedication to these projects? *

Marcar apenas uma oval.

- Yes
- No

16.

15 - According to industry standards, translation should always be followed by a revision. What about post-editing? Does it produce "enough" quality to dispense revision? *

Please select the sentence that depicts more accurately your view on this subject, even if you do not do post-editing yourself.
Marcar apenas uma oval.

- Post-editing never produces a target text that dispenses revision
- A post-editing process rarely produces a target text that dispenses revision
- Sometimes, a post-editing process produces enough quality for the target text to be used with no revision
- Post-editing is already a type of revision, so it does not require another revision
- Post-editing should always produce the same quality as translation, so it does not depend on the process whether the target text dispenses revision or not

17.

16 - Present a description of how you see the difference between translation and post-editing. Think of your experience and the tools that you use for each task. *

18.

17 - Please use the space below to add any comments that you may have at this stage, namely concerning any of the answers that you gave above.

Passe para a pergunta 27.

No experience with post-editing

19.

10 - How many words, on average, can you produce in post-editing projects per hour? *

If you do not do post-editing, please insert "0".

20.
11 - How many words do you do in post-editing projects, on average, per month? *

Marcar apenas uma oval.

- I do not do post-editing

21.
12 - How would describe your reaction to post-editing job assignments? *

Marcar apenas uma oval.

- I think I would like to do post-editing
- I think I wouldn't like to do post-editing

22.
13 - How would you describe your speed and productivity in post-editing, as compared to translating? *

Marcar apenas uma oval.

- I think I would work faster
- I think I would work slower

23.
14 - Does the fact that post-editing is paid at a lower rate affect your dedication to these projects? *

Marcar apenas uma oval.

- I think I would dedicate myself in the same way
- I think low-paid tasks cannot require the same dedication

24.
15 - According to industry standards, translation should always be followed by a revision. What about post-editing? Do you think it may produce "enough" quality to dispense revision? *

Please select the sentence that depicts more accurately your view on this subject, even if you do not do post-editing yourself.
Marcar apenas uma oval.

- Post-editing never produces a target text that dispenses revision
- A post-editing process rarely produces a target text that dispenses revision
- Sometimes, a post-editing process produces enough quality for the target text to be used with no revision
- Post-editing is already a type of revision, so it does not require another revision
- Post-editing should always produce the same quality as translation, so it does not depend on the process whether the target text dispenses revision or not

25.

16 - Present a description of how you see the difference between translation and post-editing. Think about what you know and heard about post-editing. *

26.

17 - Please use the space below to add any comments that you may have at this stage, namely concerning any of the answers that you gave above.

Before you submit

Thank you very much for your participation in this part of the questionnaire. Before you close and submit the questionnaire, please insert a "code" below, so that we will be able to correctly link the two parts of the questionnaire to the same participant.

27.

Personal code to write in Part 2 *

Please insert some word, number, or any other string that is not personal and does not allow us to identify you, but which you may remember at the workshop. Do not use words that are too common and other people may use as well, such as "workshop" or "post-editing". Take note of your "code", since, at the beginning of the workshop, you can only start the second part of the questionnaire after you inserted this. We will use only this code to join the two parts of each questionnaire.

Appendix 4: Presentation for the workshop

WORKSHOP
Tradução Automática Interativa e Pós-Edição

] Faculdade de Letras do Porto, Janeiro de 2017 [

Agenda of the workshop

Presentation

- Part 1 — 30 minutes
  - Discussion
- Part 2 — 15 minutes
  - Discussion
- Questionnaire part 2 a) – 4 questions

- *Coffee break*

Practical experiments (4 sessions)

Autocomplete mode	– text 1	15 minutes
Post-editor mode	– text 2	15 minutes
Autocomplete mode	– text 3	15 minutes
Post-editor mode	– text 3	15 minutes

- Questionnaire part 2 b) – 12 questions

Presentation

Part 1 – Current technologies

- Computer-Aided Translation (CAT) tools
- Machine Translation (MT)
- Translation and post-editing
- Future perspectives
- Research and references

Part 2 – A new view on translation technologies

- Post-editing is different from translation
- Post-editing is defined by 4 actions
- CAT tools support to 4 actions
- Testing these concepts

CATs

Computer-Aided Translation tools

- State-of-the-art
- Editing environments
- Interaction & Learning

CAT tools – First generation

Editors with different forms of leveraging content from Translation Memories (TMs) and termbases

- Matches from the TM
  - Full matches = same content
  - Fuzzy matches = similar content
  - Concordances
- Terminology databases
- Corpora
  - Muses (memoQ)

CAT tools – Second generation

Editors with different forms of leveraging content beyond Translation Memories (TMs)

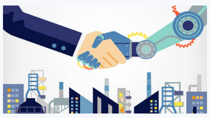
- Collaborative environments
  - SDL Language Cloud / Groupshare; memoQ Language Terminal
- Machine translation
  - APIs to external providers
  - SDL LanguageCloud & SDL AdaptiveMT
- Predictive writing
  - SDL Studio Auto-suggest; memoQ predictive typing
- Fuzzy match composition
  - Transit, Déjà Vu, memoQ
  - SDL Studio – UpLIFT (Fragment recall)

440

CATs as editing environments

- Source window –> read the source
- Target window
  - Copy source –> overwrite or edit
  - Clean window –> write everything
  - Windows/panes with suggestions –> use shortcuts to copy
  - Pop-ups with contextual information –> select to use
  - Search Windows –> shortcuts to use
- Machine Translation
  - Pre-translation inserted –> edit

How to improve interaction and support editing?



MT

Machine Translation

- Different generations of MT
- Machine Learning in SMT
- Neural Machine Translation
- Current MT tools
- Interactive Machine Translation

441

Different generations of MT

- Rule-Based machine Translation (RBMT)
  - Syntactical rules
  - Synthesis -> Transfer -> Generation
- Example-Based Machine Translation (EBMT)
  - Phrase alignments in TMs
  - Recreating translations from phrase alignments
- Statistical Machine Translation (SMT)
  - Google Translate, Bing, Moses, and everyone else…
  - Triple stage model:
    - Identifying recurring pairs in each language (Language model)
    - Identifying probable alignments (Translation model)
    - Identifying most likely units that are aligned in two languages (Decoder)

Machine Learning in SMT

- Training and testing
  - Corpora are used to train (learn) language and translation models
  - Results are evaluated against a "golden reference"
- Researchers try to improve the way systems learn
  - Better mathematical models
  - Better evaluation systems
- Quality Estimation of Machine Translation
  - Estimating whether quality is good enough for editing
  - Estimating editing required

The new kid on the block: Neural Machine Translation

- Google, Systran, etc.
- It creates matrices of words and their surroundings (embeddings)
- Predicts the next word, based on vectors that simulate brain waves…
- 85% improvement…

BUT…

- Evaluations of lexical similarity… comparing with other evaluations… only in some language pairs…
- It takes weeks to train a NMT model (SMT needs hours)
- More grammatically correct than faithful (more fluent than adequate)
  - Why?
- The decoder is totally black-box
  - unpredictable and unexplainable
  - impossible to retrain and tune
- One step back into "autonomous" systems

Interactive Machine Translation

- CasMaCAT (research tool)
  - *Don't waste time deleting wrong translations. It is more productive to start editing the wrong text while the system makes news translation proposals to the right of you typing.*
- Lilt (commercial tool)
  - Adaptive MT – the hypothesis is updated as you type along
- Simianer (Univ. Heidelberg) - 2016
  - Alignment-based: the user deletes, inserts, moves and replaces words, and corrects the alignments to the source

In common:

- Learning to improve the MT system through new alignments
- It is the MT system that translates…
- BUT in most systems (exception: Simianer) the translator writes the whole translation from left to right…

Overview of current MT tools

- CAT tools that link to MT engines and present suggestions
  - Studio's AdaptiveMT, memoQ MT engines
- Free and generic
  - Google Translator; Bing Translator; Babylon, etc.
- Software + services commercial solutions
  - Omniscien Technologies (former Asia Online)
  - Systran – Hybrid, Neural MT…
  - MateCAT, Lilt…

Overview of current MT tools

- "Do Your Own MT engine"
  - Moses for Mere Mortals (PT team at DGT); Slate Desktop (commercial)
- "We build the MT system for you…"
  - KantanMT (Dublin); Tauyou (Spain); PangeaMT (Spain)
- "Anyone can post-edit"
  - Unbabel (PT); many others…

Best tool awards

- Honorary Award
- Best predictive writing
- Best black box system
- Best user interactivity
- Best young player
- BFF with MT

- RBMT (still kicking after all those years)
- Lilt
- KantanMT
- Google Translate
- SDL AdaptiveMT
- Translators and translation industry

The future

- CATs = TM+MT
- More process control

CATs = TM+MT

- Full matches – from the TM
  - High quality – no editing
  - No payment
- Fuzzy matches – built from TM/MT
  - Editing
- New words – pre-translated by MT
  - Editing
- So, the analysis of "fuzzies" or "new" segments, against the TM, will not matter
- All segments require "editing" only
- We do not have to "translate" anything

More control over the processes

- Online tools with time logging
  - Productivity control & evaluation
- QA, sometimes replacing revision:
  - Linguistic sign-off, Quality evaluation, etc…
  - Dynamic Quality Framework (TAUS)
  - LQA models
  - Classification of errors for low-quality spotting
- Tools that record and replay translators' work
  - iOmegaT and other research tools
- Human computation…

- Knowledge is power
- Better technology, better support

The future

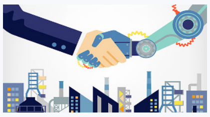
Knowledge is power

- Translation resources (TMs, termbases, etc.) do not contain or describe the fundamentals of translation decisions
- Machines do not decide as humans do BUT they are great decision supports
- Language is still very much unknown and impossible to reproduce by machines
- Machines are built by humans to do what humans want them to do

What do translators want machines to do for them?

Better technology, better support

- Improvements in interactivity, suggestions, online learning, etc.
- Measuring everything by **time** highlights the complexity of the tasks
- Payment must compensate effort -> Time is a fair measure
- The complexity of translation will be recognized, contrasting with the simpler situations in which post-editing is adequate
- Autonomous MT and NLP technologies will continue to struggle with language



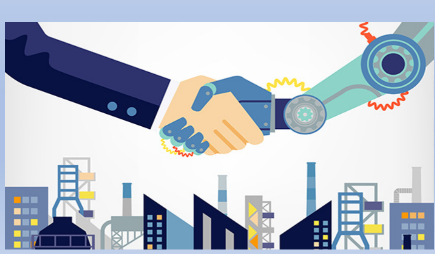
R & R Research and References

Research and references

- Descriptive Translation Studies: Gideon Toury, Juan Sager, Alan Melby and Anthony Pym
- Machine Translation history: John Hutchins and Yorick Wilks
- Post-editing: Krings, Ignacio Garcia and Sharon O'Brien
- Machine Translation (SMT, EBMT and NMT): Brown, Phillip Koehn and Andy Way
- Quality Estimation of Machine Translation: Lucia Specia, Ondrej Bojar, Alon Lavie
- New generation CAT+MT tools: Moses, MateCAT, CasmaCAT, Lilt
- Translation Process Research: Fábio Alves, Michael Carl, PACTE group (UAB)

FIM DA PARTE 1

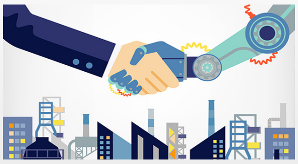
PERGUNTAS? COMENTÁRIOS?



WORKSHOP
Tradução Automática Interativa e Pós-Edição

PARTE 2

] Faculdade de Letras do Porto, Janeiro de 2017 [



Part 2

A new view on translation technologies

- Post-editing is different from translation
- Post-editing is defined by 4 actions
- CAT tools should support these 4 actions
- Testing these concepts

The post-editing process

Post-editing is the correction of machine-generated translation output to ensure it meets a level of quality negotiated in advance between client and post-editor.

(TAUS/CNGL, 2013)

- A revision process, no translation required, because this has been done by the machine
- No specific techniques
- Defined by small tasks at the text level
- Action (not post-edition)

Post-editing as 4 editing actions

4 editing actions
- Deleting
- Inserting
- Moving
- Replacing

In the literature:
- Translation Studies research (Toury, Pym, etc.)
- Translation Process Research (Carl, Alves, etc.)
- Machine Translation research:
  - Edit distance
  - Translation Edit Rate (quality metric)
  - Quality Estimation of Machine Translation
  - etc.

Post-editing redefined

A process applied to machine-translated text, in order to improve its quality, composed of 4 editing tasks (delete, insert, move and replace), within a specific threshold.

We need a measurable threshold to distinguish Editing from Translation

- Let us use the "fuzziness threshold":
  - Below 25% edited words – Editing
  - More than 25% edited words – Translation

We all know that Translation is so much more than editing…

Support to 4 editing actions

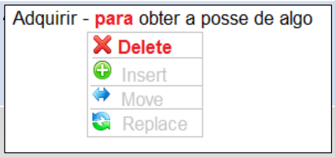
- If we only delete, insert, move and replace, these are very repetitive actions, which make them ideal to be learnt and reproduced by machines.
- Can Machine Learning learn to post-edit, but, instead of trying to do it autonomously, help us do it?

These are the two main hypotheses that I wanted to test in my project.

- Barcelona (Oct 2016): the rationale of the project
- London (Nov 2016): KATs (Knowledge-Assisted Translation tools)

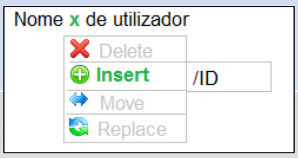
Examples of application - Delete

SOURCE	MT HYPOTHESIS	POST-EDITED
Align - to place something in an orderly position in relation to something else	Alinhar - **para** colocar algo em uma posição ordenada em relação a outra coisa	Alinhar - colocar algo em uma posição ordenada em relação a outra coisa
Allocate - to divide something between different people or projects	Alocar - **para** dividir algo entre diferentes pessoas ou projetos	Alocar - dividir algo entre diferentes pessoas ou projetos
Acquire - to obtain possession of something	Adquirir - **para** obter a posse de algo	Adquirir - obter a posse de algo



Examples of application - Insert

SOURCE	MT HYPOTHESIS	POST-EDITED
Patient Name/ID	Nome do paciente	Nome/**ID** do paciente
Item Name/ID	Nome do item	Nome/**ID** do item
User Name/ID	Nome x de utilizador	Nome/**ID** de utilizador



Examples of application - Move

SOURCE	MT HYPOTHESIS	POST-EDITED
VEC 1 controller + (RD) wire	**1 Controlador VEC** + (RD)	Fio + (RD) do **Controlador VEC 1**
VEC 1 controller – (BL) wire	**VEC 1 controlador** - (BL)	Fio - (BL) do **Controlador VEC 1**
VEC 1 controller pin 7 (BK) wire	Controlador VEC 1 fio do pino 7 (BK) Delete Insert **Move** Replace	Fio do pino 7 (BK) do **Controlador VEC 1**

Examples of application - Replace

SOURCE	MT HYPOTHESIS	POST-EDITED
Assess - to examine something in order to judge or evaluate it	Avaliar - examinar algo para **juiz** ou avaliar	Avaliar - examinar algo para **ajuizar** ou avaliar
Act - to do something to change a situation	**Ato** - fazer algo para mudar uma situação	**Atuar** - fazer algo para mudar uma situação
Users must be set up and maintained at the console.	Os utilizadores têm de estar **configurado** ser **configurado e mantido** na consola. Delete Insert Move **Replace** configurada e mantida / configurados e mantidos	Os utilizadores têm de estar **configurados e mantidos** na consola.

A call to reality

- Highly productive translation or post-editing processes are very fast and it is difficult to separate each action
- In each segment, we do several actions, recursively and cumulatively
- Some actions are simpler than others
  - e.g. replacing to correct an agreement issue vs replacing wrong terminology
- We need a very sophisticated system to learn the 4 editing actions
  - AdaptiveMT
  - Online learning models
  - EBMT
  - QEMT
  - …

Testing the concepts

- HandyCAT – an open and evolving translation editor
- Short time to develop a learning system and a mature interface
- Two alternative editing modes:
  - Autocomplete – predictive writing
  - Post-editor – conditioned to the 4 editing actions
- Main objective:
  - Check whether the concept of the 4 actions is useful for professional translators

Features of this version of HandyCAT

- Word-based (dictionary) -> Complex phrase selection
- Token-based -> spaces and punctuation need to be selected specifically
- No MT engine (we simulated MT with MateCAT)
- No Machine Learning -> no editing suggestions
- No clean interface for translation from scratch
- No Translation Memory or concordance feature

BUT it may include all that: Hokamp (2015)
AND it has some very interesting implementations in this interface

Measuring the results

- HandyCAT saves a log that registers the actions that we perform
  - This log contains:
    - The username that you were given (random – not-traceable)
    - Times that you spent on each segment
    - The initial translation by the MT system
    - In Post-editor mode: the number of actions (delete, insert, move and replace)
  - We will use this log to:
    - Compare the two editing modes, in terms of time for each segment
    - Analyse the frequency of each action
    - Identify any other statistically relevant conclusion when crossing the data
  - These details will be complemented with the final part of the questionnaire

We are NOT measuring

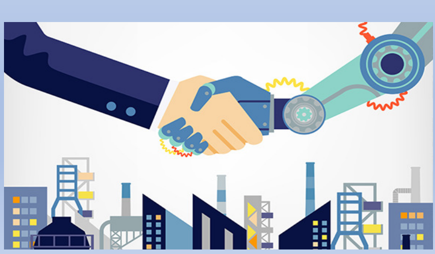
- Quality
  - We do not have a standard against which to measure your translations
  - We do not have a scale of correct/incorrect edit actions
- Productivity
  - We will not measure your productivity against your answers in the questionnaire

This is an open, exploratory work and we will only analyse:

- What the data shows us, in terms of trends
- Your impressions and opinions in the two parts of the questionnaire

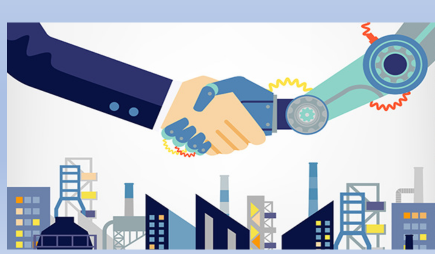
FIM DA APRESENTAÇÃO

PERGUNTAS? COMENTÁRIOS?



AO TRABALHO!

Que as máquinas não dormem…



Obrigado e

Bom Ano de 2017!!

U.PORTO

Centro de Linguística da Universidade do Porto

LIAAD INESCTEC

DCU

Appendix 5: Questionnaire 2

Workshop on Interactive Machine Translation and Post-Editing - Part 2

This is the second part of the questionnaire. The purpose of this part is to register opinions and feedback on the discussions and experiments that make up the workshop.

Please do your best to answer each question clearly and use the fields for comments to add any extra information or feedback you may want to address to us.

*Obrigatório

At the workshop

At the beginning of the presentation, please insert the details in the two fields below. The next section of the questionnaire should be completed only after the presentation.

1. **Before we begin, please write below the code that you inserted at the end of Part 1 of the questionnaire.** *

2. **Now, please insert here your username for the tests, which was given to you at the beginning of the workshop.** *
   This username will be used to link your answers in the questionnaire to the results of the experiments with HandyCAT.

After the presentation

This is a short section, for you to register your impressions about the presentation that composes the first part of the workshop.

3. **1 - Do you think that the redefinition of post-editing is clear and useful?** *

4.
2 - Do you agree with the 25% editing effort threshold to define whether a segment was translated or edited? *

Marcar apenas uma oval.

- I totally agree
- I am inclined to agree
- I don't know
- I am inclined to disagree
- I disagree completely

5.
3 - Do you think that this technological environment will be positive for your work and profession? *

Marcar apenas uma oval.

- Yes, this will have a positive impact on the profession
- It may have a positive impact on the profession
- I don't know
- I think it may have a negative impact on the profession
- I am sure this will have a negative effect on the profession

6.
4 - Please register here any further questions or comments about the issues discussed in the presentation.

....

....

....

....

....

Experiments with HandyCAT

7.
1 - Before we move on to the practical experiments, take a look at the source texts that we are about to post-edit and classify them according to a scale of complexity, especially considering the perspective of machine translation. *

Marcar apenas uma oval por linha.

	Very complex	Fairly complex	Not so complex
Text A - Mobile phone instructions			
Text B - Marketing questionnaire			
Text C - Product catalog			
Text D - Instructions manual			

8.

Once you have answered this question, please leave this browser window open, and open the HandyCAT tab in your browser. If you close the window by accident, you will need to restart Part 2 of the questionnaire.

You don't need to write anything in this "question". When you come back to the questionnaire, after the experiments, just click on "Seguinte"/"Next".

After the experiments

Please remember that HandyCAT is not fully developed. The full implementation of the concepts presented here is very complex, involving the most advanced technologies available, in a context of interactivity that has never been tested before.

So, please do not evaluate the concepts and objectives of this project based on usability glitches and problems of implementation. Focus on the actions that you performed and the adjustment of the interface to these actions.

9.

2 - Please classify the texts you post-edited in terms of a global assessment of the quality of the initial machine translation hypotheses. *

Marcar apenas uma oval por linha.

	Very high quality	High quality	Average quality	Low quality
Text A - Mobile phone instructions				
Text B - Marketing questionnaire				
Text C - Product catalog				
Text D - Instructions manual				

Comparing the two editing modes

10.

3 - Please compare the two editing modes available in HandyCAT, in terms of the following features. *

Marcar apenas uma oval por linha.

	Autocomplete mode	Posteditor mode
I work faster in this mode		
It is easier to work in this mode		
This mode is better for translation work		
This mode is better for post-editing work		
This mode is better for editing longer sequences		
This mode is better for editing just a few characters		
This mode allows me to plan and reflect about what to do		
This mode is more intrusive		

11.

4 - Which actions do you feel that you made more often in all work sessions? *

Marcar apenas uma oval.

- Deleting
- Inserting
- Moving
- Replacing

12.

5 - In which of the modes did you feel that you made more edits to the translation hypotheses? *

Marcar apenas uma oval.

- I clearly made more changes in Autocomplete mode
- I think I made most changes in Autocomplete mode
- I think I made most changes in Post-editor mode
- I clearly made more changes in Post-editor mode

Evaluating Post-editor mode

13.

6 - Post-editor mode covers more actions than Autocomplete mode. In which actions did you most feel that Post-editor mode should present suggestions for editing, as a useful support? *

Marcar tudo o que for aplicável.

- [ ] Deleting
- [ ] Inserting
- [ ] Moving
- [ ] Replacing

14.

7 - Please complete the following sentence: "Post-editor mode may be very useful if..." *

....

....

....

....

....

15.

8 - Classify each of the four editing actions of the Post-editor mode in terms of their utility in an ideal system that could present previous suggestions for their application. *

Marcar apenas uma oval por linha.

	Very useful	Useful	Does not bring anything new	It hinders our work
Deleting				
Inserting				
Moving				
Replacing				

Closing thoughts

16.

9 - Please comment on how much the approaches to post-editing presented in this workshop allowed you to see this task in a different perspective. *

17.

10 - Do you agree that "The four editing actions are an interesting form to learn or teach post-editing."? *

Marcar apenas uma oval.

- I agree completely
- I am inclined to agree
- I don't know
- I am inclined to disagree
- I disagree completely

18.

11 - Please tell us which features you would like to see in a post-editing tool. You may refer to the features that were tested in this workshop or to other that were not included in this workshop. *

19.
12 - Please present a global evaluation of HandyCAT. *

20.
13 - Please give us your global opinion about the workshop. *

Thank you very much for your participation.

This research project will soon be closed, so we really appreciate any feedback that you may want to send us.

If you would like to know more about any of the subjects addressed in this presentation and research project, please feel free to send us an email message asking for more details, be it the main references in this area, or any of the materials that were produced in this project.

21.
If you have anything else to add, please take note of it below.

Appendix 6: Sample log

```
"document": {

## Start of session in PE mode

"Session 0 - Project 0 - Post Editor": {
    "segments": {
      "0": [
        {
          "time": 1483789238192,
          "user": {
            "name": "user-02"
          },
          "project": {
            "name": "Session 0 - Project 0 - Post Editor",
            "_id": "5870d4b948980efc50f1ca7e"
          },
          "action": "change-segment",
          "data": {
            "segmentId": 0,
            "currentValue": "Você pode usar o teclado para digitar caracteres alfanuméricos e símbolos.",
            "configuration": {
              "target": {
                "activeComponent": "postEditor",
                "defaultComponent": "postEditor",
                "components": [
                  {
                    "directiveName": "postEditor",
                    "textName": "Post Editor"
                  }
                ]
              }
            }
          }
        },
        {
          "time": 1483789319850,
          "user": {
            "name": "user-02"
          },
          "project": {
            "name": "Session 0 - Project 0 - Post Editor",
            "_id": "5870d4b948980efc50f1ca7e"
          },
          "action": "postEditor.move",
          "data": {
            "originalTarget": "Você pode usar o teclado para digitar caracteres alfanuméricos e símbolos.",
            "newTarget": "usar o teclado VocêVocê podepode para digitar caracteres alfanuméricos e símbolos."
          }
        },
        {
          "time": 1483789349150,
          "user": {
            "name": "user-02"
          },
```

```
        "project": {
          "name": "Session 0 - Project 0 - Post Editor",
          "_id": "5870d4b948980efc50f1ca7e"
        },
        "action": "postEditor.delete",
        "data": {
          "originalTarget": "Você pode usar o teclado para digitar caracteres alfanuméricos e símbolos.",
          "newTarget": "pode usar o teclado para digitar caracteres alfanuméricos e símbolos."
        }
      },
      {
        "time": 1483789371525,
        "user": {
          "name": "user-02"
        },
        "project": {
          "name": "Session 0 - Project 0 - Post Editor",
          "_id": "5870d4b948980efc50f1ca7e"
        },
        "action": "postEditor.replace",
        "data": {
          "originalTarget": "pode usar o teclado para digitar caracteres alfanuméricos e símbolos.",
          "newTarget": "Pode usar o teclado para digitar caracteres alfanuméricos e símbolos."
        }
      },
      {
        "time": 1483789383408,
        "user": {
          "name": "user-02"
        },
        "project": {
          "name": "Session 0 - Project 0 - Post Editor",
          "_id": "5870d4b948980efc50f1ca7e"
        },
        "action": "segment-complete",
        "data": {
          "segmentId": 0,
          "previousValue": "Você pode usar o teclado para digitar caracteres alfanuméricos e símbolos.",
          "newValue": "Pode usar o teclado para digitar caracteres alfanuméricos e símbolos.",
          "configuration": {
            "target": {
              "activeComponent": "postEditor",
              "defaultComponent": "postEditor",
              "components": [
                {
                  "directiveName": "postEditor",
                  "textName": "Post Editor"
                }
              ]
            }
          }
```

```
              }
            }
          ],
          ## Start of session in AC mnode

            "Group 1 - Session 1": {
        "segments": {
          "0": [
            {
              "time": 1483790104481,
              "user": {
                "_id": "5870d4aa48980efc50f1ca7d",
                "name": "user-02"
              },
              "project": {
                "name": "Group 1 - Session 1",
                "_id": "5870d7e148980efc50f1ca9b"
              },
              "action": "change-segment",
              "data": {
                "segmentId": 0,
                "currentValue": "Qual dos seguintes tipos de dispositivos possui ou
                usa?",
                "configuration": {
                  "target": {
                    "activeComponent": "typeaheadEditor",
                    "defaultComponent": "typeaheadEditor",
                    "components": [
                      {
                        "directiveName": "typeaheadEditor",
                        "textName": "Autocomplete"
                      }
                    ]
                  }
                }
              }
            },
            {
              "time": 1483790110957,
              "user": {
                "_id": "5870d4aa48980efc50f1ca7d",
                "name": "user-02"
              },
              "project": {
                "name": "Group 1 - Session 1",
                "_id": "5870d7e148980efc50f1ca9b"
              },
              "action": "change-segment",
              "data": {
                "segmentId": 0,
                "currentValue": "Qual dos seguintes tipos de dispositivos possui ou
                usa?",
                "configuration": {
                  "target": {
                    "activeComponent": "typeaheadEditor",
                    "defaultComponent": "typeaheadEditor",
                    "components": [
```

```json
                    {
                      "directiveName": "typeaheadEditor",
                      "textName": "Autocomplete"
                    }
                  ]
                }
              }
            }
          },
          {
            "time": 1483790152800,
            "user": {
              "_id": "5870d4aa48980efc50f1ca7d",
              "name": "user-02"
            },
            "project": {
              "name": "Group 1 - Session 1",
              "_id": "5870d7e148980efc50f1ca9b"
            },
            "action": "segment-complete",
            "data": {
              "segmentId": 0,
              "previousValue": "Qual dos seguintes tipos de dispositivos possui ou usa?",
              "newValue": "Qual dos seguintes tipos de dispositivos POSSUI E USA?",
              "configuration": {
                "target": {
                  "activeComponent": "typeaheadEditor",
                  "defaultComponent": "typeaheadEditor",
                  "components": [
                    {
                      "directiveName": "typeaheadEditor",
                      "textName": "Autocomplete"
                    }
                  ]
                }
              }
            }
          }
        ],
        "1": [
          {
            "time": 1483790152807,
            "user": {
              "_id": "5870d4aa48980efc50f1ca7d",
              "name": "user-02"
            },
            "project": {
              "name": "Group 1 - Session 1",
              "_id": "5870d7e148980efc50f1ca9b"
            },
            "action": "change-segment",
            "data": {
              "segmentId": 1,
              "currentValue": "Selecione todas as opções que se aplicam",
              "configuration": {
                "target": {
```

```
                      "activeComponent": "typeaheadEditor",
                      "defaultComponent": "typeaheadEditor",
                      "components": [
                        {
                          "directiveName": "typeaheadEditor",
                          "textName": "Autocomplete"
                        }
                      ]
                    }
                  }
                }
              },
              {
                "time": 1483790166184,
                "user": {
                  "_id": "5870d4aa48980efc50f1ca7d",
                  "name": "user-02"
                },
                "project": {
                  "name": "Group 1 - Session 1",
                  "_id": "5870d7e148980efc50f1ca9b"
                },
                "action": "segment-complete",
                "data": {
                  "segmentId": 1,
                  "previousValue": "Selecione todas as opções que se aplicam",
                  "newValue": "Selecione todas as opções que se aplicam",
                  "configuration": {
                    "target": {
                      "activeComponent": "typeaheadEditor",
                      "defaultComponent": "typeaheadEditor",
                      "components": [
                        {
                          "directiveName": "typeaheadEditor",
                          "textName": "Autocomplete"
                        }
                      ]
                    }
                  }
                }
              }
            ],
            "2": [
              {
                "time": 1483790166189,
                "user": {
                  "_id": "5870d4aa48980efc50f1ca7d",
                  "name": "user-02"
                },
                "project": {
                  "name": "Group 1 - Session 1",
                  "_id": "5870d7e148980efc50f1ca9b"
                },
                "action": "change-segment",
                "data": {
                  "segmentId": 2,
                  "currentValue": "E qual dos seguintes tipos de dispositivos que você
```

```
            usa atualmente para acessar a Internet em casa, no trabalho ou em
            outros lugares (incluindo os dispositivos que você não possui,
            pessoalmente, no trabalho ou em um ou seja, bibliotecas lugar público
            ou internet café)?",
            "configuration": {
              "target": {
                "activeComponent": "typeaheadEditor",
                "defaultComponent": "typeaheadEditor",
                "components": [
                  {
                    "directiveName": "typeaheadEditor",
                    "textName": "Autocomplete"
                  }
                ]
              }
            }
          }
        },
        {
          "time": 1483790322077,
          "user": {
            "_id": "5870d4aa48980efc50f1ca7d",
            "name": "user-02"
          },
          "project": {
            "name": "Group 1 - Session 1",
            "_id": "5870d7e148980efc50f1ca9b"
          },
          "action": "segment-complete",
          "data": {
            "segmentId": 2,
            "previousValue": "E qual dos seguintes tipos de dispositivos que você
            usa atualmente para acessar a Internet em casa, no trabalho ou em
            outros lugares (incluindo os dispositivos que você não possui,
            pessoalmente, no trabalho ou em um ou seja, bibliotecas lugar público
            ou internet café)?",
            "newValue": "E qual dos seguintes tipos de dispositivos usa atualmente
            para aceder à Internet em casa, no trabalho ou noutros lugares
            (incluindo os dispositivos que não possui, pessoalmente, no trabalho
            ou num lugar público, por exemplo, bibliotecas ou internet café)?",
            "configuration": {
              "target": {
                "activeComponent": "typeaheadEditor",
                "defaultComponent": "typeaheadEditor",
                "components": [
                  {
                    "directiveName": "typeaheadEditor",
                    "textName": "Autocomplete"
                  }
                ]
              }
            }
          }
        }
      ],
```

Appendix 7: Answers to Questionnaires

Q1.1 - For how long have you been a translator?

Up to 1 year	3
From 1 year to 5 years	12
From 6 years to 10 years	13
From 11 years to 15 years	8
From 16 years to 20 years	3
20 years and over	11

Q1.2 - Do you work as a freelancer, or are you employed at a translation agency, or a different company?

At a company not in the translation sector	3
At a translation agency	18
Freelancer	29

Q1.3 - Classify each of the tasks below in terms of the frequency in which you perform them.

	Very frequent	Frequent	Irregular	Rare	Very rare
Translation into L1	42	2	3	2	1
Translation into L2	8	5	13	9	15
Revision (of translations)	18	16	9	3	4
Post-editing	10	11	10	7	12
Management tasks	13	4	12	4	17
Other language tasks	10	14	11	9	6

Q1.4 - How would you describe your relationship with Information Technology for Translation?

Bad - Tools are always changing and I do not have the time to adapt to changes	1
Reasonable - I use the tools because I must, but I don't like it when I am forced to change the way I work	4
Good - I can use the tools that I need in my daily life	16
Very good - I have always enjoyed learning how to work with new tools and I like exploring new ways of doing my work	29

Q1.5 - Do you prefer to type your translation from scratch or to type over the source text?

I prefer to type from scratch - my target window is always empty when I start	16
I prefer to type over the source - my target window contains the source text before I start	34

Q1.6 - Do you prefer that your CAT tool presents several suggestions (from the Translation Memory, the termbase, autosuggest/predictive writing, etc.) as you are typing your translation, or that your editing window is as clean as possible?

I prefer a clean editing enviroment, because suggestions usually interfere with my work	5
I like it when my CAT tool gives me suggestions, but sometimes they are too intrusive, or not useful at all	31
I always take advantage of all the suggestions my CAT tool gives me	14

Q1.7 - Do you use the "predictive writing", "auto-suggest" or "auto-complete" features that CAT tools usually include?

Never	5
Not usually	24
Very frequently	21

Q1.8 - Do you use a Machine Translation system to help you translate?

Never	13
Only in specific situations	24
Very frequently	13

Q1.9 - Please choose the expression that best describes your experience in post-editing.

I know what post-editing is, but I have never done it professionally	12
I have some experience in post-editing	25
I have a lot of experience in post-editing	11

Q1.10 - How many words, on average, can you produce in post-editing projects per hour?

>3500 words	0
3000 words	2
2500 words	0
2000 words	4
1500 words	1
1000 words	5
1000-500 words	14
500-100 words	10
0 words	14

Q1.11 - How many words do you do in post-editing projects, on average, per month?

I do not do post-editing	14
No more than 1.000 words	7
Between 1.000 and 10,000 words	18
Between 10,000 and 25,000 words	10
Between 25,000 and 50,000 words	1

Q1.12 - How would describe your reaction to post-editing job assignments?

I do not like post-editing - I cannot think like I am used to	12
I think I wouldn't like to do post-editing	*6*
I think I would like to do post-editing	*8*
I like post-editing - I see things from a different perspective	24

Q1.13 - How would you describe your speed and productivity in post-editing, as compared to translating?

I work slower, because I do not trust the suggestions and I have to check everything	13
I think I would work slower	*7*
I think I would work faster	*7*
I work faster, because I only worry about real errors	23

Q1.14 - Does the fact that post-editing is paid at a lower rate affect your dedication to these projects?

Yes	14
I think I would dedicate myself in the same way	*9*
I think low-paid tasks cannot require the same dedication	*5*
No	22

Q1.15 - According to industry standards, translation should always be followed by a revision. What about post-editing? Does it produce "enough" quality to dispense revision?

Post-editing never produces a target text that dispenses revision	14
A post-editing process rarely produces a target text that dispenses revision	12
Post-editing is already a type of revision, so it does not require another revision	0
Sometimes, a post-editing process produces enough quality for the target text to be used with no revision	11
Post-editing should always produce the same quality as translation, so it does not depend on the process whether the target text dispenses revision or not	13

Q2.A.2 -Do you agree with the 25% editing effort threshold to define whether a segment was translated or edited?

I totally agree	6
I am inclined to agree	31
I don't know	7
I am inclined to disagree	5
I disagree completely	1

Q2.A.3 -Do you think that this technological environment will be positive for your work and profession?

Yes, this will have a positive impact on the profession	10
It may have a positive impact on the profession	33
I don't know	7
I think it may have a negative impact on the profession	0
I am sure this will have a negative effect on the profession	0

Q2.1 - Before we move on to the practical experiments, take a look at the source texts that we are about to post-edit and classify them according to a scale of complexity, especially considering the perspective of machine translation.

	Not so complex	Fairly complex	Very complex
Text A - Mobile phone instructions	37	13	
Text B - Marketing questionnaire	17	27	6
Text C - Product catalog	16	20	14
Text D - Instructions manual	22	20	8

Q2.2 - Please classify the texts you post-edited in terms of a global assessment of the quality of the initial machine translation hypotheses.

	Low quality	Average quality	High quality	Very high quality
Text A - Mobile phone instructions	1	26	19	4
Text B - Marketing questionnaire	3	30	13	4
Text C - Product catalog]	9	32	8	1
Text D - Instructions manual	3	28	16	3

Q2.3 - Please compare the two editing modes available in HandyCAT, in terms of the following features.

	Autocomplete	Posteditor
This mode is more intrusive	17	33
This mode is better for post-editing work	20	30
This mode is better for editing just a few characters	20	30
This mode allows me to plan and reflect about what to do	23	27
I work faster in this mode	38	12
It is easier to work in this mode	40	10
This mode is better for editing longer sequences	41	9
This mode is better for translation work	46	4

Q2.4 - Which actions do you feel that you made more often in all work sessions?

Moving	1
Deleting	5
Inserting	5
Replacing	39

Q2.5 - In which of the modes did you feel that you made more edits to the translation hypotheses?

I clearly made more changes in Autocomplete mode	8
I think I made most changes in Autocomplete mode	22
I think I made most changes in Post-editor mode	17
I clearly made more changes in Post-editor mode	3

Q2.6 - Post-editor mode covers more actions than Autocomplete mode. In which actions did you most feel that Post-editor mode should present suggestions for editing, as a useful support? (choose all that apply)

Deleting	5
Inserting	25
Moving	1
Replacing	40

Q2.8 - Classify each of the four editing actions of the Post-editor mode in terms of their utility in an ideal system that could present previous suggestions for their application.

	Moving	Replacing	Inserting	Deleting
Very useful	10	19	22	29
Useful	25	19	14	10
Does not bring anything new	10	9	11	9
It hinders our work	5	3	3	2

Q2.10 - Do you agree that "The four editing actions are an interesting form to learn or teach post-editing."?

I disagree completely	0
I am inclined to disagree	4
I don't know	4
I am inclined to agree	25
I agree completely	17

Q1.8 - b) Please describe the situations in which you use machine translation as a support.

Always, unless not permitted by the source file
When there is enough stored material to apply the machine translation correctly.
When I have trouble finding a solution (for example, word order)
Technical translation and post-editing projects.
index
To find suggestions about multi-word units of translation
When I am running out of time, when I don't understand the meaning of a sentence, when I am not satisfied with my translation and want a
I have tested systran and one or two other systems
Always
When the deadlines are too tight or it is a customer requirement
Very large project with a very tight deadline
For instances, when I don’t know the meaning of several terms in a sentence or when I realize that I’m spending too much time to understand the context in a sentence.
Only when such requirement is defined by the client.
I use machine translation as a support because I receive machine translated texts from clients.
When required by clients
Repetitive work
In case of very technical terms, but only with words or expressions. Almost never whole sentences.
I currently use machine translation when the clients require us (me and the agency where I work) to do so. The type of translation is mostly IT-related (software, hardware).
I often receive the jobs pretranslated by my client's MT system.
when the cat tool provide it
When requested to do so by our client.
Translation of instruction manuals, product launch packages, technical documents
Translation into a second language, English, basically. Technical, legal, institutional texts
As a base for my translation, keeping the words I agree with and changing the style to match my own.
For all my translation work
When requested by the client.
To look up a word or expression in Google Translate; when I use MateCat and I'm pressed for time, I usually activate the Machine Translation
Manuals, repetitive text.
*In specific themes, when translating into a foreign language.
Very simples translations, like names of countries
translation of a list of countries
My agency works on several machine translated embedded projects
Looking for specialized vocabulary
*To speed up translation
Specific expressions, but not technical terminology.
Any translation work with over 1000 words

Q1.16 - Present a description of how you see the difference between translation and post-editing. Think of your experience and the tools that you use for each task.

Post-editing implies working on a predefined text, which you need to check for correctness.
because I know/work with at least 4 languages, I do know that sometimes a translated word does not exactly mean what the source language word meant... so when I think about translation of novels, poems etc. I can hardly imagine letting the machine do the translation for me. I will agree that in some other cases, "machine translated textes" are suitable, but if I were a decision maker in a translation firm, I would always prefer "regular" translation and revision of the texts instead of doing only post-editing.
Very useful in mechanical text (e.g. annual reports), possibly in creative texts as it may offer original and new vocabulary
I'm a bit suspicious about post-editing for, on the rare occasions I've had to deal with it, I had to spend a greater effort on it than on translation
The main difference for me is that in translation I always try to write in a more natural way, whereas in post-editing I try to make minimum changes to the suggested MT. I do this because 1) it is the only way to work faster and 2) it is the best way to help the process of MT, so that my edited translation can feed the databases of MT and be used in other instances.
In a translation task, the translator can take benefit of the TM, termbase supplied by the cliente and also of another resources of CAT tools such as predictive writing. In post-editing, besides the resources already mentioned, the translator can also use the suggestions made by machine
sintax complexity
Translation is a computer-aided human performance; post-editing is human-aided computer performance.
In post-editing, your line of thoughts will be influenced by what is already written.
Theoretically, post-editing should be better than translation, productivity wise. This is true after some time of feeding the machine translation engine with trusted (manual) translations, i.e. for the machine translation to have quality, its output has to be thoroughly edited and reviewed for some time. Eventually, the machine will be "smart" enough to produce quality translations on its own, with minimum revision, thus saving the translator's time and money. I believe that, with a fairly good machine translation, the translator saves a lot of time just by not having to type most of the text. However, if the machine translation is bad, it is better to delete it and start from scratch.

*Answers with star were translated from Portuguese.

Quite different Modus operandi and required mindset
I use the same CAT tools in both tasks
Completely different
Post-editing is a language service, not a translation job, although it is more appropriately done by translators.
Post-editing is part of an automated translation process, not requiring as much time as that necessary to translate a text from scratch and allowing a higher productivity rate.
In my opinion, the main difference between translation and post-editing is that the second one is made by a machine translation. Therefore, a machine translation can't interpret and translate the meaning of, for instances, an idiomatic expression or a double meaning in a specific sentence.
Translation involves total freedom and full responsibility; with post-editing there is shared responsibility and a certain amount of "restrictions". The way the brain works on each task must be different too, but I haven't yet understood how.
The tools I use for both translation and post-editing are the same. I think there is a big difference in terms of sentence structures, the way you organize your ideas and the text you edit. This results in 'raw' target texts in post-editing. As for terminology, I think there are bigger risks in post-editing too, especially if there will be no revision.
Post-editing is most useful in repetitive, regular jobs, such as instruction manuals, regular technical publications, etc. Translation is needed in cases where more attention needs to be paid to style (eg. marketing) or interpretation (eg. law), for instance.
I consider post-editing more like a revision than a translation.
Translation and post-editing are very different processes. At the beginning, even when using CAT Tools, you have to do nearly everything from scratch. Over time you start to build up translation memories and termbases and in turn the translation process becomes faster. On the other hand, when doing post-editing you start with a "complete" translation that was done based upon previous texts that were fed into the Machine Translation software. Although the translation process may be faster, the post-editing process requires a lot more attention, because you need to spot possible mistakes and mistranslated terms, which may be tricky. Also, when using most Machine Translation and Post-Editing software you come across many syntax mistakes that can be easily avoided and would not occur in the "normal" translation process. From my experience, Post-editing is more fruitful when used to translate very technical texts with a lot of recurring and specialized terminology. When dealing with marketing texts, ads, etc. that require creativity and some adapting to the target culture you are better off with "traditional" translation. In conclusion, post-editing is advantageous in terms of speed and in the fact that you can continue feeding the software. As you do so, you can keep bettering the quality of its initial output. It is best suited for technical texts with recurring terminology and will undoubtedly be one of the main tools for translation in the near future.
Post-editing has potential to actually increase productivity especially in large projects, but I believe this might be a tool with which some clients/companies will try to take advantage of the lower rates regardless of the machine translation quality. From my experience, the machine translation quality can really influence the productivity, as well as the simplicity and correctness of the SOURCE text.
Post-editing, as opposed to translation, seems to be some sort of "budget" translation service and the amount o work needed in order to achieve a satisfactory target text depends a lot on how well tuned is the MT system. Translation, on the other hand, depends a lot more on the translator's experience and therefore, it requires a bit more experience and time, unless it's a type of source text that the translator is familiar with.
I don't think the tools required for each task to be that different. However, I certainly use a different approach (whether intentionally or not): when translating, I create the translation from scratch, allowing only the interference from my references (translation memories, termbases, auto-suggest dictionaries, etc.). In post-editing, I rely more on the machine translation, adapting my wording to the already existing text.
Post-editing is to correct translation made by machine translation.
Post-editing targets a specific market and adapts information to culture or public, sometimes even with a commercial goal, whilst translation intents to pass on the source message and to make sure the message is understood and clear.
Post-editing requires a mental translation and a corresponding correction or not on na actual translation. Post-editing allows some distance from the produced translation, sometimes similar to revision, which can be helpful in the process.
Post-editing can be misleading and more unreliable than translation. As there is already a translation suggestion available, it is very easy to leave that suggestion as it is and not to perform research to confirm if the suggestion is correct.
Translation involves some sort of human-based intervention. Post-editing is the result of a machine-based approach to the translation process.
I know very little about post-editing.
Post-editing uses all of the machine translation tools, therefore you are not "working alone" on a translation, this is valid for the positive and negative aspects.
Regarding tools there is no difference for me because I receive the pre-translated files for post-editing from the client. So I have no experience with machine translation engines, and use a regular CAT tool for post-editing. The most significant difference between translation and post-editing is productivity, post-editing allows an increase in word output of up to 4 times.
There is no "soul" in post-editing, comparing with translation. The language's nuances don't exist.
Post-editing helps but not 100%
To me, post-editing is still quite rare so I have not yet developed a systematic approach to it. Translation requires more input on my part and more creativity whereas post-editing requires more attention and can become quite dull.
.
It all depends on how we define translation and post-editing, on whether our definition is narrow and rigid or ample and flexible. The key difference, in my opinion, is that post-editing is always a human 'reworking' (with or without specific tools) of a target-language version produced by a machine. The human 'reworking' can involve all aspects of the text at both macro and micro-level and it is an eminently pragmatic process whose final aim is to ensure that the target language version can be used effectively for the purpose for which it is being produced. Thus, post-editing is always a human action/intervention on 'text' produced by a machine. Translation, on the other hand, can be seen as the product of human endeavour assisted by technology (such as CAT tools) or the product of machine translation systems. However, it is becoming increasingly more difficult to keep these concepts (and their respective terms) separate. It is probably best to place them on a continuum of activities aimed at the production of a text that serves a specific purpose and audience.

*Answers with star were translated from Portuguese.

Very bad original text
* The same tools. In some texts, PE is indeed faster than translating and editing is closer to a revision. But in specific texts the quality produced by machine translation is very low.
Translation is made by humans, post-editing by machines.
Translate is create and post-editing is basically correct what has been translated by a machine that doesn't have the ability to think like we do.
We use the same software (SDL Trados Studio) for both tasks but we add SDL Language Cloud in post-editing in order to make our work more efficient while maintaining an acceptable reliability of the automated translation
Translation implies some creativity and allows critical thinking, while post-editing is mostly a guided exercise of informative writing.
*Post-editing is done over a translation that was produced before, usually by a machine translation system, with the objective of obtaining a functional product in the target language. My experience includes two different types of work: revising texts in Word which had been translated by a machine translation system, and revising texts in an online platform which offers machine translation functionalities.
Although I have no professional experience, I have discussed post-editing with fellow coworkers who have. My idea of post-editing is that it is not comparable to translation - at least not yet. It is usually a work that must be completed in shorter deadlines and in which the post-editor cannot have an "obligation" of adapting or localising the content into a target culture, as machine translation does not cover cultural aspects, idiomatic
In post-editing, the MT output quality is extremely important as it is more likely to influence the quality of the final text when compared to
Post-editing requires the translator to edit merely the parts of the segment that actually need changes, regardless of his/her personal preference, whereas translation involves the act of creating the content entirely from scratch and based on the translator's personal preferences.
When post-editing, I always check carefully every segment with MT, since I don't fully trust the results. Sometimes I erase the whole segment and translate it from scratch. However, I'm actually surprised with how well it works in some projects and it saves some time. When translating, I take more time since I have to translate all from scratch, check TMs and other resources more carefully.
Translation is supposedly done from scratch, while in PE you quickly edit the text looking for specific problems. My experience is limited, so I can't mention tools and workflow
Post-editing has real potencial in terms of quality and, compared to translation, in terms of consistency and productivity. However, it still has a long way to go in terms of categorisation of the different "service" levels it requires.

Q1.17 - Please use the space below to add any comments that you may have at this stage, namely concerning any of the answers that you gave

Question 12 - There are only two options - like or dislike - which do not reflect my opinion. I understand post-editing can speed up the translation process for certain text types (namely technical) but have seen it applied to marketing content and in this case it's simply used to reduce what translators get paid for word.
I consider myself a very open-minded person and I am usually a computer-friendly person, but for some specific tasks, like translations, I think a trained human being is better prepared for the subtilities of texts/words and that at the end, this would buy us considerable time
Note that I have no direct experience in post-editing machine translation
Fortunately, technology is always improving, so I hope to get a new light on post-editing with this workshop and start working on it with better
Since I am not a very tech-savy person, although I have been using CAT tools for quite some time, post-editing seems to have a "techonologic streak" to significant for my personal taste/profile.
The reaction to post-editing job assignments and the speed and productivity in post-editing, as compared to translating, both depend on the quality of the machine translation results. If it is not good, I do not like post-editing and I work slower. The fact that post-editing is paid at a lower rate affect the dedication to these projects if the machine translation results provided are not good at all and they require more effort post-editing than translating from scratch. If the mahine translation results are good enough, sometimes, a post-editing process produces enough quality for the target text to be used with no revision.
In the Portuguese language, another aspect that might be negative regarding post-editing is the possible "mixture" of the Portuguese variants (European and Brazilian, for instance), resulting in more modifications and terminology differences. I don't think that everything about post-editing is bad, but I think that when a company chooses to use machine translation and assign a post-editing job/project, it should prepare it beforehand as good as possible and/or be opened to feedback from translators/post-editors/translation agencies in order to improve the process as a whole.
N/A
Although I am yet not experienced in post-editing, I believe it is an inevitable process and that it can be improved over time, so as to become the next step in computer-aided translation.
Question 16 tries to problematize the concepts of translation and post-editing but it is not clear whether these two terms refer to the social/cognitive/etc. activities of translating and post-editing or rather to the products of those activities, or to both. I therefore found it difficult to give what I would consider a satisfactory answer.
Some answers are too restrictive in their scope.
*Some questions did not include enough levels of response.
Q 12 and Q 13: I've answered "I don't like post-editing" and "I work slower" based on the (many) bad experiences. My answers are the exact opposite when Post-edition TMs and AT logic are working as they're supposed to.

*Answers with star were translated from Portuguese.

Q2.A1 - Do you think that the redefinition of post-editing is clear and useful?

Yes, but it needs to be more specific as to the changes made.
YES
Extremely clear and useful to decide on exactly what I am doing/have done
It is clear and useful, but incomplete, because it sets the tone on machine translation instead of human translation.
The definition is clear and useful, but it may raise some questions, namely: there might be more actions involved in the process (rephrasing or partially changing words); and what if only 20% of the text is changed, but those 20% are made up of complex terminology, that take hours of research? Or what if there are more than 25% changes in the text, but the changes are mainly preferential?
Yes, it is clear and useful but I have to try it first.
yes
It is useful to add the 4 editing tasks, but the level of quality negotiated in advance is missing.
It gives you a general idea, but it is still somewhat subjective and depends on the actual situation.
Yes.
Yes, though imcomplete.
yes
Yes
Not entirely.
The redefinition is clear and useful.
Yes
Yes
Yes.
Somehow
no, I think it might be necessary to improve it furthermore the definition
Yes.
Yes. But the need for a revision after post-editing is arguable and depends on the characteristics of each project.
Yes, because it establishes some sort of metrics and allows some differentiation between post-editing and translation tasks.
Generally speaking, yes. I believe it presents a globally clearer definition of what I believe post-editing to be, even though it cannot embrace all the nuances of this task.
yes
It is definitely useful but I still believe it needs to be more accurate.
Useful yes, but not clear enough. There is a grey area when defining it against revision or translation.
Yes
*Yes, although in some cases it creates some conditionalisms.
sim
Yes
Yes
Yes
yes
Yes, however I need more information on the matter.
*It may be a bit incomplete.
I don't think the word threshold makes it a clear definition. The rest of the definition is fairly clear, but I think its usefulness could be improved.
Absolutely
*In a way, it leaves the distinction between translation and revision open. The two processes are a bit mixed.
yes
yes
Yes
Yes
*I think it is not sufficient to describe all tasks performed during post-editing.
Yes.
Yes
Yes
It's clear, however it might not be as useful as one would like, since the MTs are often not as good and the project requires full translation (even if the client considers it a post-editing job).
Yes. Current definitions for light and full PE are unclear as for the translators tasks, the redefinition would seem to solve this.
I think it's definitely useful but still unclear.

Q2.A.4 - Please register here any further questions or comments about the issues discussed in the presentation.

These changes caused by these technologies still need further analysis and consideration. The industry is not prepared for this yet. There is a danger that they are used merely to lower the price paid for translation.
Still depends on the source text - some texts I would not dream of machine translating because I need to view the whole block, not just isolated sentences and I want the result to be harmonious, not repetitive
n.a.
I seriously doubt that a quantitative threshold can establish a difference between translation and post-editing.
the assessment of the time spent while doing the post-editing should be discussed

*Answers with star were translated from Portuguese.

The impact of these technological advances and all the research done around translation, machine translation and post-editing may be positive one the profession. In my opinion, this environment may be advantageous in two ways:
1- It may indeed help the translator work faster while maintaining the quality of the translation, thanks to the new tools at their disposal.
2- All the efforts put into MT systems may show researchers and users in general that the quality of the translations produced by these systems is not comparable with the one that is produced by translators. This will mean that people will look at translation as a complex profession that need professional and specialized human translators.

Further discussions should be made to reach a more concrete definition of the post-editing efforts, as well as the budgeting parameters, etc.

The presentation was clear and concise and covers many of the issues related to the MT work environment. I think it was well prepared and is quite relevant for the professional market.

Technological environment as always adavntages

no question

*As per question 3, it may have a positive impact and be useful in certain types of text. However, clients may use this notion of post-editing to pay all texts as post-editing instead of translation, even when the machine translation results do not permit post-editing and require that the whole translation is deleted and rewritten.

None

*Clear presentation, including the main trends in this line of research.

Q2.7 - Please complete the following sentence: "Post-editor mode may be very useful if..."

the user practises enough.

The machine translation is of high quality, requiring minor changes to words (number/gender kind of changes) and does not require a lot of moving

the text is highly technical and the memory is specific to the context

The commands on the interface are refined.

the actions can be more automatized, for example, by using keyboard shortcuts to select words, to perform the actions and also to move inside the words that need to be replaced/edited.

it also displays suggestions for replacing words.

listing words

if it were more intuitive.

if there are only minor changes to be made to the translation.

You don't have to hit Escape every time you edit a word, or if you could use the keyboard instead of the mouse to select words/actions.

The text results from a specialized field MT with good corpora and training

----

we are to translate very, very short sentences.

only a few changes are require.

There are only minor changes to make to a text.

the machine translation has high quality.

the changes required are mostly connected with changing word order/placing.

... the MT text being edited presents a fairly high quality level, implying just a few changes.

it presented suggestions/auto-replacement of words in upper/lowercase according to source text; there was no need to press "esc" after each action; it learned

the sentences are short/small amount of segments to be translated

the translation is a high-quality on that need only a few changes in terms of terminology or grammatical structures or style.

The text to post-edit is made of short and relatively simple sentences.

The translation hypothesis has some level of quality and sentence sequence is almost correct

...it uses an improved UI. It is certainly useful in terms of not making too many changes and making the translator (editor?) think about what really needs to be changed, but it is quite slow to use as it requires too many "clicks" to make a change. If these changes can be implemented (non-requirement of the ESC key, insertion of shortcut keys, a different approach to spaces, etc.), it is certainly a "tool" worth giving a try.

The source text is not very technical

it allows to think in a larger segment scale

we are doing a final revision with minor edits as number and gender concordance.

it was easier to select words for deletion/replacing. It may also be very useful if the actions to perform were more automated.

dealing with shorter less complex texts, involving little / superficial intervention at text level

After I practice a long time

The system learns from your actions

the MT output has high quality

we are working in an area that requires a specific terminology

it allows to insering

post-editing technical translations already reviewed.

*Above all, it should be combined with other functionalities, such as auto-suggestions of corrections. After the translator is already fanmiliarised with the editing mode, and with faster controls available, it may indeed be useful in post-editing assignments.

the quality of the machine translation if already very good.

the suggestions are more intuitive.

*if the machine translation suggestions are very good and only require minor changes

the hypotheses are very good

The actions were faster and easier to do (the four of them)

the "move" action has a drag & drop feature instead of point & click to a certain section of the segment

It does not exist...

*If it identifies the edits that are necessary in the next segments and these are applied automatically

*Answers with star were translated from Portuguese.

a) it is possible to replace more than one word and b) if it is possible to mouse click on the part of the selected word/expression that we want to change while in replace mode (instead of having to get there just by using arrow keys).
the text produced by the machine translation has a lot of replacing changes to be done as it helps in making small adjustments to single words.
it is made more user-friendly and when the post-editor is accustomed to the way of working of this mode.
One could change one one character in some words instead of rewriting the whole word.
there are suggestions for replacement, similar to predictive typing
I need to assess the amount of changes I make and reflect upon my post-editing process.

Q2.9 - Please comment on how much the approaches to post-editing presented in this workshop allowed you to see this task in a different perspective.

They allowed us to start thinking about the changes to come to this industry and ways to deal with them.
Did not really change my ideas about what the task requires.
I had never broken up these 4 actions, usually I only think of replace and move
Post-editing will play a more important role in the future and cannot be ignored.
The presentation helped to clarify some concepts and also to learn more about the latest technologies in MT. it also raised an interesting question about the relationship between post-editing and revision/proofreading and how this may affect the post-editing process.
The approaches offer the translator more options to post-editing a text already translated by MT.
didn't change my perspective
The 4 tasks was something new.
It didn't change my perspective on post-editing, although it made it clear for me that there are numerous tools. I still think it all depends on the quality of the translation (and, in some cases, the quality of the source text).
Considering that I already have some experience in post-editing, I would say my perspective changed from a theoretical point of view, meaning it made me think of post-editing in a more cerebral way.
It was an objective approach, I think, that helps me to keep in the course of better understanding the process
It showed me that Post-editing is far more complex and that are many different approaches
Reasonably
It definitely made me understand the large extent to which post-editing is different from revision.
This workshop made me consider the use of machine translation as a pre-translation phase (I do not usually use machine translation), as it will allow me to produce a higher output a the end of the day.
In my opinion, the four editing actions are a useful way to post-edit. However, I think that at a first sight that post-editing work takes longer to perform. Perhaps with some practice this approach could be useful.
Testing HandyCat Post-editor mode gave the nice experience of not feeling so "lost" among the "chaos" of a translation to be post-edited.
This made me realize the main actions I perform while post-editing, something many of us really don't realize while doing actual work. But my main views on post-editing remain basically the same.
Actions that were made unconsciously are now clear and if they can be automatically performed by the machine, it can be of great help for translators.
I might consider it as a useful tool in the future. I will although need more time to learn to use it better
The approach resented during the workshop helped me better understand where the research on MT and post-editing is heading and what the main focus is. On the other hand, I found the hypothesis of post-editing based only on four operations very intriguing and interesting. After testing it out during the practical part, I found it may be a good way to work (in terms of post-editing) if developed further.
It made me learn about other possibilities to approach post-editing.
My perspective on post-editing was already very similar to the one shown in this workshop. The most significant change that was presented was relative to the future of translation itself, since there's a clear trend to introduce more and more MT in the process.
I believe the main difference was in the way it made me realize how some changes are not actually "required".
It does not because I am use to do post-editing jobs
This worshop was very useful in the sense that allows me to reflect on the changes I introduce to the AT text.
It did. we forget that a big part of our work as translator is in fact post-editing and processes and client demands are changing at a somewhat fast pace. We need to reflect more about the new processes and tools for translation and post-editing, in a way that we can improve the productivity and quality or our work.
It was interesting to have a "technological" approach, an explanation on why there is a higher demand of post-editing projects and how it works.
Very useful since brought the whole process into a new perspective. Focus on the process and also on decision-making processes
This workshop was very useful, even if I am not fluent in English language.
Very much. It allowed me a better understanding of the task and a hands-on approach.
The need to choose an action, forced me to consider if I was making a preferential change or a necessary change, which is one of the main characteristics of post-editing - avoid unnecessary changes.
seeing the 4 tasks as a complement to help machine translation do do a better work.
Interresting
It could be helpful, and the knowledge transmitted could be helpful as well. But it does require always the human factor.
*I did not have a very strong notion of what post-editing is, so this was a great opprotunity to gain a new perspective over the world of translation.
It gave me a clearer perspective of what the language industry means by post-editing.
Quite a lot.
*It has reinforced a previous notion that post-editing will become more common when clients are translation companies. I feel that in the future I will do more post-editing, and that "pure" translation can only be done for direct customers.
My perspective hasn't changed.
I understood better how its works in terms of logic and knowledge and it seems now more interesting to me
If used with adequate caution and on less intricate texts, it can improve productivity and make post-editing a more enjoyable task. A lot needs to be done on the interface and functionality of the software but I see a lot of potential for this tool.
Not so different from my previous work.
*It has allowed me to understand how this task will take on a greater importance in translators' work and the tools that are available. It was possible to see the direction the industry is taking.

*Answers with star were translated from Portuguese.

I believe they can be helpful in making the post-editing work more "mechanic" for the translator, but I also think it could be slightly more user-friendly to make our work faster.
I had never thought about how we could divide post-editing in these 4 actions.
It did not really change my perspective. However, I think it was very useful to get an overview of how we got here, what are the challenges and opportunities that lie ahead in the future, and how systems could improve the way post-editing is done.
It is, in fact, very useful when handed properly.
My work with PE is very limited, I was never asked to work in a different environment (it's usually a CAT tool), and for sure never limited as far as typing goes. When you have this limit, the changes you make are also influenced, being limited to the least needed.
I have never really thought about post-editing in these terms (reduced to 4 simple actions) although they are not clearly defined yet.

Q2.11 - Please tell us which features you would like to see in a post-editing tool. You may refer to the features that were tested in this workshop or to oth

The features tested seemed to be a useful approach. For technical translations, terminology tools are still essential.
I would like to see autocomplete included in post-editor mode, in the "inserting" action
Spellcheck, grammar, double words and spaces
Auto-suggest could be of some interest.
A blend of the features presented in this workshop would be quite useful.
The most useful feature would be predictive writing-
predective
No idea.
If I change a word from masc to fem, or from sing to plural, it would be useful to have like an autocorrect that automatically changes the sentence in those terms.
Auto-suggest should definitely be included, as well as a feature that could allow us to change the suffix of a word in order to switch number/gender/tense, etc. A translation memory is also useful to propagate full matches and fuzzy matches.
A mix of auto-suggest and the actions above, plus a good terminology search tool.
All those tested, The replace option is a bit tricky.
Free insert and delete
Predictive phrases.
Propagation of changes implemented to other segments (for instance, if one deletes a word more than twice, that should be propagate to other segments where the word appears).
I'm not quite sure, but maybe the four editing actions.
I'm afraid I don't know enough about the matter.
Predictive writing is a useful help which should be present in post-editing tool, although I don't usually use it. I also found replacement interesting.
Present suggestions/auto-replacement of words in upper/lowercase according to source text; no need to press a key after each action; learn automatically that frequently erased words are to be erased automatically in future segments (eg.: "você" always erased); articles added automatically to words
the tools (deleting/inserting/replacing/moving) should be improved.
- Word suggestions (including synonyms) - Termbase integration for researching alternative terms - Delete feature - Move feature (although copy-paste may still be more efficient)
I think that the four features might be useful, but they should not be the only solution. Possibly integrated within the normal capability of just writing the text. Also, the Replace feature could have a "sub-feature" with words suggestions (plural form, female form, etc).
Some sort of consistency in the way some terms are translated and especially consistency on product names and brands.
I don't know if there is something much different to insert in such a tool. What I believe is that these features can be presented in more intuitive, clean, user-friendly ways.
It should understand the translation field and provide term bases by it self. That will be awesome.
syntactic features
rewrite for multiples words
It would be useful not to have Brazilian Portuguese in texts that are being translated into European Portuguese. Another interesting feature would be to include a post-analysis functionality to understand how much of the original machine translation text was changed.
Concordance, phraseology, terminology
Don't have an opinion about this meter.
An interesting feature would be a combination of auto-suggest and the 4 steps
All tested features are mandatory, it would also be good to have a quick text feature in the insert function
replace
Corrections
The features are helpful, but should be more user-friendly, namely the "Move" option. It should also include a propagation tool.
*As commented before, an auto-suggestion feature.
I would like to see some of the 4 features conflated into one. For example, you do not need Insert and Replace. There could be only 1 function called Insert/Replace.
Drag and drop
*An autosuggest feature, that learns from the corrections that are made.
Post.editing analysis of changes made
the four features of this tool are very useful but the way of working with them is rather slow, so maybe in terms of time it won't be saving a lot (when compared to translation)
Replace
None whatsoever.
*The most relevant would be to reproduce the Replace function in the next segments that present similar contexts, and the ability to apply it to more than on word.

*Answers with star were translated from Portuguese.

In auto-complete mode, I think it would be useful to simple replace the entire word we chose to change, instead of adding the different part of the suggested word to the actual word.
The features tested would be useful.
Since I have no relevant experience in post-editing in a professional context, it is difficult for me to say... I would only be able to answer this question if I had more hands-on and everyday experience on post-editing.
It should include an option to change characters.
Global changes option, since many segments are repetitions. The lack of mouse use when editing was especially frustrating and harmful to my LER
I love the idea of a way of counting the changes made and of tracking the time so that the post-editing work can be paid accordingly.

Q2.12 - Please present a global evaluation of HandyCAT.

The tool seems to be useful to deal with post-editing but still requires development.
I think it has all the makings to become a good tool for the professional.
The Autocomplete is great, Post-editing is too clumsy, hinders my thinking
The Autocomplete feature needs a larger database and the Post Editor commands could be refined.
As is, the process of post-editing with this tool is a bit slow, because we need to click too many times to perform an action (click to select, to choose action and also to escape). Nevertheless, the logic behind it seems to be useful and could prove quite useful for post-editing in the future.
Like the name suggests, it is a handy tool that translators and post-editors can use to improve their work.
easy to use and friendly user
Interesting.
I preferred the autocomplete mode, it gave me more freedom to alter anything. However, I sometimes felt that the suggestions were not at all accurate.
It seemed like a "handy" tool for post-editing. Simple interface, useful tools.
Promissing
Globally good.
It's a clean CAT tool - useful, but it needs improvment
The tool is good.
The autocompletion mode is, in my opinion, more useful and allows post-editors to work faster than the post-editor mode. Consideration should be taken to make replace action simpler to use than it is currently.
My experience in the autocomplete mode was not the best. In this mode every time I entered a new word the following word were "eaten". So, my work was not productive, because I had to entered every word.
A very pleasant surprise.
Allows fast post-editing in autocomplete mode. Post-editor mode is interesting but hinders productivity.
Interesting and promising perspective
If/when improved and with more time to test it, it might be helpful while doing some technical translations
I think HandyCat is an interesting tool to play around with and if developed further and tweaked a bit, may be useful for professional post-editing work. One of the main flaws is the the tool treats spaces in between words.
It is a simple tool (in a good way) but the use of the "Esc" button might present a problem for new users.
It surely needs a lot of work in order to be useful or productive enough to the professional translator, autocomplete mode is quite redundant and ineffective.
It's an interesting concept, but it requires a bit more developing, mainly in terms of improving the UI, so that the translators can test it in a more user-friendly environment, thus giving a more definitive opinion.
I thin it is much user friendly and not complicated at all. I enjoyed!
Very good approach and very well thought.
simple interface, minor bugs.
It is a practical tool which can still be improved. The editor is easy to use, but it lacks some useful functionalities (like QA functions, propagation of repetitions and a mode to see the complete text more easily).
Needs a bit of practice, somehow intrusive, needs cognitive approach. Can be time-consuming. Very useful afterwards, should you get used to it.
Very satisfied
It is an interesting post-editing tool. I felt that with the post-editor mode, you can pay more attention to detail, because you get more time to analyze the wording as it takes more time to edit it.
I found it pretty useful, and much better during use than while reading about the 4 functions. On the post-editor, the tags view needs some improvement because once a segment is opened for editing, tags disappear, both in source and target segments.
very good
Very good
3
Com alguns ajustes, ferramenta muito útil. Função de eliminar não me parece muito necessária.
Good and easy to use after a brief training session. Pressing the Esc button all the time is a bit of a nuisance, though. Also, I believe the four functions could be reduced to 3.
Excellent aid.
*Appropriate for the purposes of the workshop, but I wouldn't like to work with such an application professionally.
It is just a tool for a research project. I would never use it in another environment.
the global evaluation in very positive
Good
Useless.
*Average
Very positive and potentially helpful.
Globally, seems to be a useful tool. It would help to solve some issues in Pos-editor, such as the spacing problems at the end of the sentences.
Seems very interesting in its essence, and I believe that it can develop into a very useful tool.
Overall, it is a useful tool although it may require some extra features.

*Answers with star were translated from Portuguese.

Interesting tool. It needs more speed and integration with other functions.
Not enough elements to evaluate. It has been a useful tool to analyse the post-editing process.

Q2.13 - Please give us your global opinion about the workshop.

Very interesting insight into new technologies and new approaches.
I enjoyed the workshop, since it gave me an overall perspective of PE. I also enjoyed working with this tool and I think I may include some of what I have learned in future projects (not a complete change of approach to these tasks, but maybe a fine-tuning of what I do)
Fabulous introduction to the subject - I would like to participate in more!
The workshop has brought me a glimpse of the future in translation.
it was quite clear and useful.
The topics of the workshop were covered in a very interesting way and the approach gave me a different perspective of post-editing.
excelent!
Very good
For me, being fairly new to the profession, it was very interesting to learn about all the tools available and their differences. I do still have some issues with MT. Though I think it is a useful tool, I regard it only as a tool. Sometimes, it only means more work...
It was very enriching from a theoretical point of view. It was also good to put our knowledge into practice.
Very good
Very interesting
I really liked it, it was clear and useful!
Excellent work. Well done!
Very useful and interesting. Thank you.
At overall, it focuses every aspect of machine translation and post-editing, giving me the opportunity to learn more about the theme.
A good opportunity to recap, update and know more about post-editing.
It was worth attending it. Alhtough my views on post-editing didn't change much, such initiatives are always useful to make us aware about what we do everyday.
Interesting, well structured, needed more time
it was very interesting and useful; some interesting input was given; it might have changed my perspective of post-editing
I really liked the workshop. I found it interesting and very informative. Nonetheless, I would have liked that the part about the future of machine translation and the new tools (and how they work) had been discussed a bit more in depth.
It was helpful and presented new ways of looking at Machine Translation and all that it implies. I would do it again and I hope there will be more similar workshops.
An intelligent approach on post-editing and machine translation, especially for professionals who might not be as convinced as needed that this might be the future of translation. The presented information, concepts and the discussions on some subjects were clear and informative. Great job!
Quite interesting in terms of thinking about the future of our industry. I found it useful to become aware with the concept of the 4 operations, and I would like to see it implemented in a more "finished" way.
It was o good way to spend a Saturday morning, listen to professionals and meet new perspectives.
The workshop has the ambition of foreseeing post-editing work as a future corporate strategy for professional translators.
Very useful, offered valuable insights in this are, always changing and difficult to keep up with, and opened the curiosity follow up more closely this theme and news on it.
Very good and informative :)
Very useful, well-organised, clear and professional. Should be implemented at agency-level, freelancers and training sessions. Good luck on your research.
Very satisfied
A very thorough and complete workshop, giving us an insight of the state of the art of machine translation and post-edition tools and teaching us how to use them to our advantage.
It was very enriching, and I learned a lot about new MT processes. I had never thought about a tool specific for post-editing, but while doing the same text in this tool (after using autocomplete) I realized I was more focused on real and necessary changes.
Excellent work
Excelent
Quite informative, explicit, practical and useful for future post-editing jobs.
*Quite informative.
I enjoyed it very much. It was extremely well organized in the pre-workshop phase, with all the information we needed. The practical part of the workshop (the actual working with Handycat) was a bit rushed. It should have taken double the time. If you want people to reflect on their experience and produce valuable opinions (data you will use in your project) you should give them more time.
Very user-friendly. Felix is an excellent speaker.
*Very interesting for the new perspectives and to make people aware of the need to adapt to the future. Whether we like it or not, we have to adapt to technologies and to what these bring us, no matter if it is good or not so good.
Good, although I feel a little bit disappointed as I thought we would learn how to work (translate and post-edit) with Matecat.
very good, well planned, interesting and useful
Excellent
Very interesting.
*Quite positive
Also very positive and helpful, and it gave me a bigger insight on post-editing work.
This workshop was extremely useful as as gained some knowledge about what is happening in the industry and post-editing.
Very interesting and insightful (as would be expected from anything coming from Félix). Hope that the info/data provided can be helpful.
It was very useful, since it helped me to get a broader perspective regarding post-editing. Also, it introduced me a few new tools.
Very much interesting. Useful information shared about MT and state-of-the-art technology, NMT especially. A different perspective on PE definition and pratical work were a good combination.
Excellent. Simple, clear, straightforward and practical. It motivates me to reflect upon the post-editing process and to discuss it.

*Answers with star were translated from Portuguese.

Appendix 8: TER alignment report - Simulated edits

DETAILS of TERp Results

Segment ID [hyp][[Text A - Source text for MT.txt][0001]][][] (back to top)
Number of Edits 1.0
Number of Words 11.0
TERp Score 0.0909090909090909091
Original Reference Você pode usar o teclado para digitar caracteres alfanuméricos e símbolos.
Original Hypothesis Você usar o teclado para digitar caracteres alfanuméricos e símbolos.
Alignment
Reference você pode D usar o teclado para digitar caracteres alfanumére ímbolos.
Hyp After Shifts você usar o teclado para digitar caracteres alfanumére ímbolos.
Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0002]][][] (back to top)
Number of Edits 1.0
Number of Words 16.0
TERp Score 0.0625
Original Reference Por exemplo, você pode adicionar entradas para os contatos, escrever mensagens ou agendar eventos no calendário.
Original Hypothesis Por exemplo, você pode ainda adicionar entradas para os contatos, escrever mensagens ou agendar eventos no calendário.
Alignment
Reference por exemplo, você pode I adicionar entradas para os contatos, escreveensagen ou agendareventos no alendário.
Hyp After Shifts por exemplo, você pode ainda adicionar entradas para os contatos, escreveensagen ou agendareventos no alendário.
Reference por exemplo, você pode I adicionar entradas para os contatos, escreveensagen ou agendareventos no alendário.
Hyp After Shifts por exemplo, você pode ainda adicionar entradas para os contatos, escreveensagen ou agendareventos no alendário.
Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0003]][][] (back to top)
Number of Edits 1.0
Number of Words 10.0
TERp Score 0.1
Original Reference Você pode percorrer esta lista e selecione o caractere desejado.
Original Hypothesis Você pode percorrer esta lista e seleccionar o caractere desejado.
Alignment
Reference você pode percorrer esta lista e selecione S o caractere desejado.
Hyp After Shifts você pode percorrer esta lista e seleccionar o caractere desejado.
Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0004]][][] (back to top)
Number of Edits 1.0
Number of Words 8.0
TERp Score 0.125
Original Reference Pare quando você vê um "h" na tela.
Original Hypothesis Pare você quando vê um "h" na tela.
Alignment
Reference pare quando você vê um "h" na tela.
Hyp After Shifts pare quando você vê um "h" na tela.
Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
1	1	2	1.0	você	você

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0005]][][] (back to top)
Number of Edits 1.0
Number of Words 19.0
TERp Score 0.05263157894736842
Original Reference Depois de inserir a primeira letra, você pode pressionar diretamente outra tecla (exceto Enviar) para inserir a próxima letra.
Original Hypothesis Depois de inserir a primeira letra, pode você pressionar diretamente outra tecla (exceto Enviar) para inserir a próxima letra.
Alignment
Reference depois de inserir a primeira letra, você pode pressiona diretamente outra tecla (excetoenviar) para inserir a próxima letra.
Hyp After Shifts depois de inserir a primeira letra, você pode pressiona diretamente outra tecla (excetoenviar) para inserir a próxima letra.
Reference depois de inserir a primeira letra, você pode pressiona diretamente outra tecla (excetoenviar) para inserir a próxima letra.
Hyp After Shifts depois de inserir a primeira letra, você pode pressiona diretamente outra tecla (excetoenviar) para inserir a próxima letra.
Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
6	6	7	1.0	pode	pode

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0006]][][] (back to top)
Number of Edits 1.0
Number of Words 31.0
TERp Score 0.03225806451612903
Original Reference Se a próxima letra desejada estiver na mesma tecla que a actual, aguarde até que o cursor aparece à direita da letra atual e, em seguida, você pode digitar o próximo.
Original Hypothesis Se a próxima letra desejada estiver na tecla que mesma a actual, aguarde até que o cursor aparece à direita da letra atual e, em seguida, você pode digitar o próximo.
Alignment
Reference se a próxima letra desejada estiver na mesma tecla que a actual,aguarde até que o cursor aparece à direita da letra atual e, em seguida você pode digitar o próximo
Hyp After Shifts se a próxima letra desejada estiver na mesma tecla que a actual,aguarde até que o cursor aparece à direita da letra atual e, em seguida você pode digitar o próximo
Reference se a próxima letra desejada estiver na mesma tecla que a actual,aguarde até que o cursor aparece à direita da letra atual e, em seguida você pode digitar o próximo
Hyp After Shifts se a próxima letra desejada estiver na mesma tecla que a actual,aguarde até que o cursor aparece à direita da letra atual e, em seguida você pode digitar o próximo
Reference se a próxima letra desejada estiver na mesma tecla que a actual,aguarde até que o cursor aparece à direita da letra atual e, em seguida você pode digitar o próximo
Hyp After Shifts se a próxima letra desejada estiver na mesma tecla que a actual,aguarde até que o cursor aparece à direita da letra atual e, em seguida você pode digitar o próximo
Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
9	9	6	1.0	mesma	mesma

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0007]][][] (back to top)
Number of Edits 1.0
Number of Words 14.0
TERp Score 0.07142857142857142
Original Reference Você pode acessar as configurações de tela selecionando Configurações a partir do menu principal.
Original Hypothesis Você pode acessar as tela configurações de selecionando Configurações a partir do menu principal.
Alignment
Reference você pode acessar as configurações de tela selecionando configura a partir do menu incipal.
Hyp After Shifts você pode acessar as configurações de tela selecionando configura a partir do menu incipal.
Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
4	4	6	1.0	tela	tela

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0008]][][] (back to top)
Number of Edits 3.0
Number of Words 10.0
TERp Score 0.3
Original Reference Definições do telefone - Pode alterar as definições do telefone.
Original Hypothesis Definições do telefone - Pode alterar as definições.
Alignment
Reference definições do telefone - pode alterar as definições D do D telefone. S
Hyp After Shifts definições do telefone - pode alterar as definições.
Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0009]][][] (back to top)
Number of Edits 3.0
Number of Words 16.0
TERp Score 0.1875
Original Reference Data e Hora - Para definir o formato de data e hora do sistema do telefone.
Original Hypothesis Data e Hora - Para definir o formato de data e do telefone.
Alignment
Reference data e hora - para definir o formato de data e hora D do D sistema D do telefone.
Hyp After Shifts data e hora - para definir o formato de data e do telefone.

Reference data e hora - para definir o formato de data e hora do sistema do telefone.
D D D
Hyp After Shifts data e hora - para definir o formato de data e do telefone.

Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0010]][][] (back to top)
Number of Edits 2.0
Number of Words 14.0
TERp Score 0.14285714285714285
Original Reference Perfis - Para selecionar diferentes perfis para o telefone para atender diferentes situações ambientais.
Original Hypothesis Perfis - Para selecionar diferentes perfis para o telefone do vizinho para atender diferentes situações ambientais.
Alignment
Reference perfis - para selecionar diferentes perfis para o telefone para atende difere situaçõ ambientais.
I I
Hyp After Shifts perfis - para selecionar diferentes perfis para o telefone do vizinho para atende difere situaçõ ambientais.
Reference perfis - para selecionar diferentes perfis para o telefone para atende difere situaçõ ambientais.
I I
Hyp After Shifts perfis - para selecionar diferentes perfis para o telefone do vizinho para atende difere situaçõ ambientais.

Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0011]][][] (back to top)
Number of Edits 4.0
Number of Words 10.0
TERp Score 0.4
Original Reference Idioma do telefone - Para definir o idioma do telefone.
Original Hypothesis Idioma do telefone - Para definir o idioma do telefone como o seu.
Alignment
Reference idioma do telefone - para definir o idioma do telefone.
I I I S
Hyp After Shifts idioma do telefone - para definir o idioma do telefone como o seu.

Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0012]][][] (back to top)
Number of Edits 2.0
Number of Words 18.0
TERp Score 0.1111111111111111
Original Reference Esta opção é utilizada para definir o idioma do telefone como sendo o mesmo utilizado no cartão SIM.
Original Hypothesis Esta opção é utilizada para definir o idioma da chamada como sendo o mesmo utilizado no cartão SIM.
Alignment
Reference esta opção é utilizada para definir o idioma do telefone como sendo o mesmo tilizad no cartão sim.
S S
Hyp After Shifts esta opção é utilizada para definir o idioma da chamada como sendo o mesmo tilizad no cartão sim.
Reference esta opção é utilizada para definir o idioma do telefone como sendo o mesmo tilizad no cartão sim.
S S
Hyp After Shifts esta opção é utilizada para definir o idioma da chamada como sendo o mesmo tilizad no cartão sim.

Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0013]][][] (back to top)
Number of Edits 3.0
Number of Words 16.0
TERp Score 0.1875
Original Reference Você também pode selecionar Configurações do telefone / perfis para definir os sons para o telefone.
Original Hypothesis Você também pode selecionar Configurações do telefone / perfis para definir os sons com a chamada.
Alignment
Reference você também pode selecionar configurações do telefone / perfis para definir os sons para o telefone.
S S S
Hyp After Shifts você também pode selecionar configurações do telefone / perfis para definir os sons com a chamada.
Reference você também pode selecionar configurações do telefone / perfis para definir os sons para o telefone.
S S S
Hyp After Shifts você também pode selecionar configurações do telefone / perfis para definir os sons com a chamada.

Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0014]][][] (back to top)
Number of Edits 1.0
Number of Words 14.0
TERp Score 0.07142857142857142
Original Reference Período de bloqueio do teclado - Pode definir o período de bloqueio do teclado.
Original Hypothesis Período de bloqueio do teclado - Pode definir o bloqueio período de do teclado.
Alignment
Reference período de bloqueio do teclado - pode definir o período de bloquei do eclado.
Hyp After Shifts período de bloqueio do teclado - pode definir o período de bloquei do eclado.

Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
9	9	11	1.0	bloqueio	bloqueio

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0015]][][] (back to top)
Number of Edits 1.0
Number of Words 19.0
TERp Score 0.05263157894736842
Original Reference Se não houver nenhuma operação no telefone dentro de um período de tempo predefinido, o teclado é bloqueado automaticamente.
Original Hypothesis Se não houver nenhuma no telefone operação dentro de um período de tempo predefinido, o teclado é bloqueado automaticamente.
Alignment
Reference se não houver nenhuma operação no telefone dentro de um períodc de tempo definic o teclado é bloqueado automaticamente.
Hyp After Shifts se não houver nenhuma operação no telefone dentro de um períodc de tempo definic o teclado é bloqueado automaticamente.
Reference se não houver nenhuma operação no telefone dentro de um períodc de tempo definic o teclado é bloqueado automaticamente.
Hyp After Shifts se não houver nenhuma operação no telefone dentro de um períodc de tempo definic o teclado é bloqueado automaticamente.

Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
6	6	3	1.0	operação	operação

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0016]][][] (back to top)
Number of Edits 1.0
Number of Words 21.0
TERp Score 0.047619047619047616
Original Reference Você pode pressionar a tecla de função esquerda e a tecla de função direita, por sua vez, para desbloquear o teclado.
Original Hypothesis Você pode pressionar a esquerda tecla de função e a tecla de função direita, por sua vez, para desbloquear o teclado.
Alignment
Reference você pode pressionar a tecla de função esquerda e a tecla de função direita por sua vez, para desblo o eclado.
Hyp After Shifts você pode pressionar a tecla de função esquerda e a tecla de função direita por sua vez, para desblo o eclado.
Reference você pode pressionar a tecla de função esquerda e a tecla de função direita por sua vez, para desblo o eclado.
Hyp After Shifts você pode pressionar a tecla de função esquerda e a tecla de função direita por sua vez, para desblo o eclado.

Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
4	4	7	1.0	esquerda	esquerda

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0017]][][] (back to top)
Number of Edits 1.0
Number of Words 17.0
TERp Score 0.058823529411764705
Original Reference Atalhos - Para atribuir as funções usadas com freqüência às teclas de rolagem como teclas de atalho.
Original Hypothesis Atalhos - Para atribuir as usadas com freqüência funções às teclas de rolagem como teclas de atalho.
Alignment
Reference atalhos - para atribuir as funções usadas com freqüênciàs teclas de rolagem como teclas de atalho.
Hyp After Shifts atalhos - para atribuir as funções usadas com freqüênciàs teclas de rolagem como teclas de atalho.
Reference atalhos - para atribuir as funções usadas com freqüênciàs teclas de rolagem como teclas de atalho.
Hyp After Shifts atalhos - para atribuir as funções usadas com freqüênciàs teclas de rolagem como teclas de atalho.

Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
8	8	4	1.0	funções	funções

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0018]][][] (back to top)
Number of Edits 1.0
Number of Words 14.0
TERp Score 0.07142857142857142
Original Reference Auto On e OFF - Para configurar o telefone para ligar ou desligar automaticamente.
Original Hypothesis Auto On e OFF - Para configurar para ligar o telefone ou desligar automaticamente.
Alignment
Reference auto on e off - para configurar o telefone para ligar ou lesliga aticamente.
Hyp After Shifts auto on e off - para configurar o telefone para ligar ou lesliga aticamente.

Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
7	8	10	1.0	para ligar	para ligar

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0019]][][] (back to top)
Number of Edits 1.0
Number of Words 18.0
TERp Score 0.05555555555555555
Original Reference No modo de espera, você pode pressionar e segurar para mudar o perfil atual para o perfil silencioso.
Original Hypothesis No modo de espera, você e segurar pode pressionar para mudar o perfil atual para o perfil silencioso.
Alignment
Reference no modo de espera, você pode pressionar e segurar para mudar o perfil atual para o perfil silencioso.
Hyp After Shifts no modo de espera, você pode pressionar e segurar para mudar o perfil atual para o perfil silencioso.
Reference no modo de espera, você pode pressionar e segurar para mudar o perfil atual para o perfil silencioso.
Hyp After Shifts no modo de espera, você pode pressionar e segurar para mudar o perfil atual para o perfil silencioso.

Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
5	6	8	1.0	e segurar	e segurar

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0020]][][] (back to top)
Number of Edits 4.0
Number of Words 12.0
TERp Score 0.3333333333333333
Original Reference Se o perfil atual estiver offline, você não poderá usar essa função.
Original Hypothesis Se o perfil atual estiver offline, você essa função não poderá usar.
Alignment
Reference se o perfil atual estiver offline, você não poderá usar essa função.
S S
Hyp After Shifts se o perfil atual estiver offline, você não poderá função essa usar.

Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
9	10	6	1.0	não poderá	não poderá
9	9	10	1.0	essa	essa

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0021]][][] (back to top)
Number of Edits 1.0
Number of Words 15.0
TERp Score 0.06666666666666667
Original Reference Configurações de exibição - Em seguida, você pode alterar a senha de restrição de chamadas.
Original Hypothesis Configurações de exibição - Em seguida, você pode senha de restrição alterar a de chamadas.
Alignment
Reference configurações de exibição - em seguida, você pode alterar a senha de estriçã de hamadas.
Hyp After Shifts configurações de exibição - em seguida, você pode alterar a senha de estriçã de hamadas.

Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
11	12	7	1.0	alterar a	alterar a

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0022]][][] (back to top)
Number of Edits 2.0
Number of Words 24.0
TERp Score 0.08333333333333333
Original Reference A senha de restrição de chamadas de idade é fornecido pelo operador de rede se é a primeira vez para você mudar a senha.
Original Hypothesis A senha de restrição de chamadas de é fornecido pelo operador de rede se é a primeira vez para mudar a senha.
Alignment
Reference a senha de restrição de chamadas de idade é fornecido pelo perado de rede se é a primeira vez para você mudar a nha.
D D
Hyp After Shifts a senha de restrição de chamadas de é fornecido pelo perado de rede se é a primeira vez para mudar a nha.
Reference a senha de restrição de chamadas de idade é fornecido pelo perado de rede se é a primeira vez para você mudar a nha.
D D
Hyp After Shifts a senha de restrição de chamadas de é fornecido pelo perado de rede se é a primeira vez para mudar a nha.

Shifts

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0023]][][] (back to top)
Number of Edits 2.0
Number of Words 18.0
TERp Score 0.1111111111111111
Original Reference Chamada em espera - Esta opção é utilizada para activar ou desactivar a função de chamada em espera.
Original Hypothesis Chamada em espera opcional - Esta opção é utilizada para activar ou desactivar a função de chamada internacional em espera.
Alignment
Reference chamada em espera - esta opção é utilizada para activar ou sactiv a função de chamada em espera.
I I
Hyp After Shifts chamada em espera opcional - esta opção é utilizada para activar ou sactiv a função de chamada internacional em espera.
Reference chamada em espera - esta opção é utilizada para activar ou sactiv a função de chamada em espera.
I I
Hyp After Shifts chamada em espera opcional - esta opção é utilizada para activar ou sactiv a função de chamada internacional em espera.

Shifts

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0024]][][] (back to top)
Number of Edits 2.0
Number of Words 9.0
TERp Score 0.2222222222222222
Original Reference A função de chamada em espera depende da rede.
Original Hypothesis A função de telefone em espera depende da chamada.
Alignment
Reference a função de chamada em espera depende da rede.
S S
Hyp After Shifts a função de telefone em espera depende da chamada.

Shifts

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0025]][][] (back to top)
Number of Edits 4.0
Number of Words 15.0
TERp Score 0.26666666666666666
Original Reference Para obter mais informações, entre em contato com o centro de atendimento ao cliente local.
Original Hypothesis Para obter informações mais, entre em contato com o centro de atendimento ao local cliente.
Alignment
Reference para obter mais informações, entre em contato com o centro de endimen ao cliente local.
S S S S
Hyp After Shifts para obter informações mais, entre em contato com o centro de endimen ao local cliente.

Shifts

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0026]][][] (back to top)
Number of Edits 2.0
Number of Words 21.0
TERp Score 0.09523809523809523
Original Reference Salvar número desconhecido - Esta função é usada para solicitar se deseja salvar o número após a conclusão de uma chamada.
Original Hypothesis Salvar desconhecido - Esta função é usada para solicitar se você deseja salvar o número após a conclusão de uma chamada.
Alignment
Reference salvar número desconhecido - esta função é usada para solicitar se deseja salvar o número após a onclusã de uma hamada.
D I
Hyp After Shifts salvar desconhecido - esta função é usada para solicitar se você deseja salvar o número após a onclusã de uma hamada.
Reference salvar número desconhecido - esta função é usada para solicitar se deseja salvar o número após a onclusã de uma hamada.
D I
Hyp After Shifts salvar desconhecido - esta função é usada para solicitar se você deseja salvar o número após a onclusã de uma hamada.

Shifts

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0027]][][] (back to top)
Number of Edits 2.0
Number of Words 18.0
TERp Score 0.1111111111111111
Original Reference Remarcação automática - Esta função é utilizada para remarcar automaticamente o número discado se a discagem tiver falhado.
Original Hypothesis Remarcação automática - Esta é utilizada para remarcar depressa o número discado se a discagem tiver falhado.
Alignment
Reference remarcação automática - esta função é utilizada para remarcar automaticamente o número discado se a discagem tiver falhado.
D S
Hyp After Shifts remarcação automática - esta é utilizada para remarcar depressa o número discado se a discagem tiver falhado.
Reference remarcação automática - esta função é utilizada para remarcar automaticamente o número discado se a discagem tiver falhado.
D S
Hyp After Shifts remarcação automática - esta é utilizada para remarcar depressa o número discado se a discagem tiver falhado.
Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0028]][][] (back to top)
Number of Edits 2.0
Number of Words 16.0
TERp Score 0.125
Original Reference Modo de resposta - É utilizado para definir o modo para responder a uma chamada recebida.
Original Hypothesis Modo de resposta - É para definir o modo para a responder uma chamada recebida.
Alignment
Reference modo de resposta - é utilizado para definir o modo para esponde a uma chamada recebida.
D
Hyp After Shifts modo de resposta - é para definir o modo para esponde a uma chamada recebida.
Reference modo de resposta - é utilizado para definir o modo para esponde a uma chamada recebida.
D
Hyp After Shifts modo de resposta - é para definir o modo para esponde a uma chamada recebida.

Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
10	10	11	1.0	a	a

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0029]][][] (back to top)
Number of Edits 3.0
Number of Words 12.0
TERp Score 0.25
Original Reference Estão disponíveis três opções: qualquer tecla, tecla Enviar e deslizar para cima.
Original Hypothesis Estão disponíveis quase três opções: qualquer tecla, tecla Enviar para cima.
Alignment
Reference estão disponíveis três opções: qualquer tecla, tecla enviar e leslizar para cima.
I D D
Hyp After Shifts estão disponíveis quase três opções: qualquer tecla, tecla enviar para cima.
Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0030]][][] (back to top)
Number of Edits 3.0
Number of Words 15.0
TERp Score 0.2
Original Reference Atendimento automático - Esta opção é usada para selecionar a resposta automática para chamadas recebidas.
Original Hypothesis Atendimento automático imediato - Esta opção é usada quando quer a resposta automática para chamadas recebidas.
Alignment
Reference atendimento automático - esta opção é usada para selecionar a esposta tomáti para chamada recebidas.
I S S
Hyp After Shifts atendimento automático imediato - esta opção é usada quando quer a esposta tomáti para chamada recebidas.
Reference atendimento automático - esta opção é usada para selecionar a esposta tomáti para chamada recebidas.
I S S
Hyp After Shifts atendimento automático imediato - esta opção é usada quando quer a esposta tomáti para chamada recebidas.
Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0031]][][] (back to top)
Number of Edits 2.0
Number of Words 13.0
TERp Score 0.15384615384615385
Original Reference Estão disponíveis quatro opções: Resposta imediata, Mensagem de texto, Voz gravada e Desactivar.
Original Hypothesis Estão disponíveis mais quatro opções: Resposta imediata, Mensagem de texto, e Voz gravada Desactivar.
Alignment
Reference estão disponíveis quatro opções: resposta imediata, mensagem de texto, voz gravada e sactivar.
I
Hyp After Shifts estão disponíveis mais quatro opções: resposta imediata, mensagem de texto, voz gravada e sactivar.

Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
10	10	12	1.0	e	e

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0032]][][] (back to top)
Number of Edits 4.0
Number of Words 24.0
TERp Score 0.16666666666666666
Original Reference ID do chamador - Esta opção permite que você selecione se deseja exibir seu número de telefone no telefone da pessoa que você chamou.
Original Hypothesis ID do telefone - Esta opção permite se deseja exibir seu número de telefone no telefone da pessoa que você chamou.
Alignment
Reference id do chamador - esta opção permite que você selecione se deseja exibir seu número de telefone no elefone da pessoa que você amou.
S D D D
Hyp After Shifts id do telefone - esta opção permite se deseja exibir seu número de telefone no elefone da pessoa que você amou.
Reference id do chamador - esta opção permite que você selecione se deseja exibir seu número de telefone no elefone da pessoa que você amou.
S D D D
Hyp After Shifts id do telefone - esta opção permite se deseja exibir seu número de telefone no elefone da pessoa que você amou.
Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0033]][][] (back to top)
Number of Edits 5.0
Number of Words 17.0
TERp Score 0.29411764705882354
Original Reference Se seleccionar Auto, se o seu número de telefone será mostrado no telefone chamado depende da rede.
Original Hypothesis Se seleccionar Ontem, se o seu número de telefone será mostrado no telefone chamado depende da rede e de si.
Alignment
Reference se seleccionar auto, se o seu número de telefone será ostrad no elefon chamad depende da rede.
S I I I S
Hyp After Shifts se seleccionar ontem, se o seu número de telefone será ostrad no elefon chamad depende da rede e de si.
Reference se seleccionar auto, se o seu número de telefone será ostrad no elefon chamad depende da rede.
S I I I S
Hyp After Shifts se seleccionar ontem, se o seu número de telefone será ostrad no elefon chamad depende da rede e de si.
Shifts
Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0034]][][] (back to top)
Number of Edits 2.0
Number of Words 15.0
TERp Score 0.13333333333333333
Original Reference Selecionar linha - Esta opção é usada para definir a linha de chamada do telefone.
Original Hypothesis Selecionar telefone - Esta opção é a usada para definir linha de chamada do telefone.
Alignment
Reference selecionar linha - esta opção é usada para definir a linha de chamada do elefone.
S
Hyp After Shifts selecionar telefone - esta opção é usada para definir a linha de chamada do elefone.

Shifts

Start Pos	End Pos	Moved To	Cost	Hyp String	Ref String
6	6	9	1.0	a	a

Phrase Substitutions

Segment ID [hyp][[Text A - Source text for MT.txt][0035]][][] (back to top)
Number of Edits 4.0
Number of Words 18.0
TERp Score 0.2222222222222222
Original Reference Meu número - Esta função pode ser usada para salvar seus próprios números de telefone no cartão SIM.
Original Hypothesis Meu número - Esta função usada para salvar seus próprios números de telefone e contactos no cartão SIM.
Alignment
Reference meu número - esta função pode ser usada para salvar seus róprio números de telefone no cartão sim.

Appendix 9: Compared analysis of TER - Simulated edits

						EXPECTED								POST-EDIT COMPARE REPORT						
	No. Seg.	REFERENCE	LENGTH	ACTIONS	EDITED	Delete	Insert	Replace	Move	EDITing actions	Edits / length	No. Words edited	Edited words / length	Delete	Insert	Subs	Shift	Track changes	Nº of errors	TERP
GROUP 1	1	Você **pode** usar o teclado para digitar caracteres alfanuméricos e símbolos.	**11**	**Delete 1 word**	Você usar o teclado para digitar caracteres alfanuméricos e símbolos.	1	0	0	0	1	9,09%	1	9,09%	1	0	0	0	Você ~~pode~~ usar o teclado para digitar caracteres alfanuméricos e símbolos.	1	9,09%
	2	Por exemplo, você pode adicionar entradas para os contatos, escrever mensagens ou agendar eventos no	**16**	**Insert 1 word**	Por exemplo, você pode **ainda** adicionar entradas para os contatos, escrever mensagens ou agendar eventos no	0	1	0	0	1	6,25%	1	6,25%	0	1	0	0	Por exemplo, você pode ainda adicionar entradas para os contatos, escrever mensagens ou agendar eventos no calendário.	1	6,25%
	3	Você pode percorrer esta lista e **selecione** o caractere desejado.	**10**	**Replace 1 word**	Você pode percorrer esta lista e **seleccionar** o caractere desejado.	0	0	1	0	1	10,00%	1	10,00%	0	0	1	0	Você pode percorrer esta lista e ~~selecione~~seleccionar o caractere desejado.	1	10,00%
	4	Pare **quando** você vê um "h" na tela.	**8**	**Move 1 word 1 position forward**	Pare você **quando** vê um "h" na tela.	0	0	0	1	1	12,50%	1	12,50%	0	0	0	1	Pare ~~quando~~ você quando vê um "h" na tela.	1	12,50%
	5	Depois de inserir a primeira letra, você **pode** pressionar diretamente outra tecla (exceto Enviar) para inserir a próxima letra.	**19**	**Move 1 word 1 position back**	Depois de inserir a primeira letra, **pode** você pressionar diretamente outra tecla (exceto Enviar) para inserir a próxima letra.	0	0	0	1	1	5,26%	1	5,26%	0	0	0	1	Depois de inserir a primeira letra, ~~você~~ pode você pressionar diretamente outra tecla (exceto Enviar) para inserir a próxima letra.	1	5,26%
	6	Se a próxima letra desejada estiver na **mesma** tecla que a actual, aguarde até que o cursor aparece à direita da letra atual e, em seguida, você pode digitar o próximo.	**31**	**Move 1 word 2 positions forward**	Se a próxima letra desejada estiver na tecla que **mesma** a actual, aguarde até que o cursor aparece à direita da letra atual e, em seguida, você pode digitar o próximo.	0	0	0	1	1	3,23%	1	3,23%	0	0	0	1	Se a próxima letra desejada estiver na ~~mesma~~ tecla que mesma a actual, aguarde até que o cursor aparece à direita da letra atual e, em seguida, você pode digitar o próximo.	1	3,23%
	7	Você pode acessar as configurações de **tela** selecionando Configurações a partir do menu principal.	**14**	**Move 1 word 2 positions back**	Você pode acessar as **tela** configurações de selecionando Configurações a partir do menu principal.	0	0	0	1	1	7,14%	1	7,14%	0	0	0	1	Você pode acessar as tela configurações de ~~tela~~ selecionando Configurações a partir do menu principal.	1	7,14%
2	8	Definições do telefone - Pode alterar as definições **do telefone**.	**10**	**Delete 1 phrase (2 words)**	Definições do telefone - Pode alterar as definições.	1	0	0	0	1	10,00%	2	20,00%	2	0	1	0	Definições do telefone - Pode alterar as definições ~~do telefone~~.	3	30,00%
	9	Data e Hora - Para definir o formato de data e **hora do sistema** do telefone.	**16**	**Delete 1 phrase (3 words)**	Data e Hora - Para definir o formato de data e do telefone.	1	0	0	0	1	6,25%	3	18,75%	3	0	0	0	Data e Hora - Para definir o formato de data e ~~hora do sistema~~ do telefone.	3	18,75%
	10	Perfis - Para selecionar diferentes perfis para o telefone para atender diferentes situações ambientais.	**14**	**Insert 1 phrase (2 words)**	Perfis - Para selecionar diferentes perfis para o telefone **do vizinho** para atender diferentes situações ambientais.	0	1	0	0	1	7,14%	2	14,29%	0	2	0	0	Perfis - Para selecionar diferentes perfis para o telefone do vizinho para atender diferentes situações ambientais.	2	14,29%
	11	Idioma do telefone - Para definir o idioma do telefone.	**10**	**Insert 1 phrase (3 words)**	Idioma do telefone - Para definir o idioma do telefone **como o seu**.	0	1	0	0	1	10,00%	3	30,00%	0	3	1	0	Idioma do telefone - Para definir o idioma do telefone como o seu.	4	40,00%
	12	Esta opção é utilizada para definir o idioma **do telefone** como sendo o mesmo utilizado no cartão SIM.	**18**	**Replace 1 phrase (2 words)**	Esta opção é utilizada para definir o idioma **da chamada** como sendo o mesmo utilizado no cartão SIM.	0	0	1	0	1	5,56%	2	11,11%	0	0	2	0	Esta opção é utilizada para definir o idioma ~~do~~da ~~telefone~~chamada como sendo o mesmo utilizado no cartão SIM.	2	11,11%
	13	Você também pode selecionar Configurações do telefone / perfis para definir os sons **para o telefone**.	**16**	**Replace 1 phrase (3 words)**	Você também pode selecionar Configurações do telefone / perfis para definir os sons **com a chamada**.	0	0	1	0	1	6,25%	3	18,75%	0	0	3	0	Você também pode selecionar Configurações do telefone / perfis para definir os sons ~~para~~com ~~o~~a ~~telefone~~chamada.	3	18,75%
	14	Período de bloqueio do teclado - Pode definir o **período de bloqueio** do teclado	**14**	**Move 1 phrase (2 words) 1 position forward**	Período de bloqueio do teclado - Pode definir o bloqueio **período de** do teclado	0	0	0	1	1	7,14%	2	14,29%	0	0	0	1	Período de bloqueio do teclado - Pode definir o bloqueio período de ~~bloqueio~~ do teclado.	1	7,14%

						EXPECTED								POST-EDIT COMPARE REPORT						
	No. Seg.	REFERENCE	LENGTH	ACTIONS	EDITED	Delete	Insert	Replace	Move	EDITing actions	Edits / length	No. Words edited	Edited words / length	Delete	Insert	Subs	Shift	Track changes	Nº of errors	TERP
GROUP	15	Se não houver nenhuma operação **no telefone** dentro de um período de tempo predefinido, o teclado é bloqueado automaticamente.	**19**	**Move 1 phrase (2 words) 1 position back**	Se não houver nenhuma **no telefone** operação dentro de um período de tempo predefinido, o teclado é bloqueado automaticamente.	0	0	0	1	1	5,26%	2	10,53%	0	0	0	1	Se não houver nenhuma ~~operação~~ no telefone operação dentro de um período de tempo predefinido, o teclado é bloqueado automaticamente.	1	5,26%
	16	Você pode pressionar a **tecla de função** esquerda e a tecla de função direita, por sua vez, para desbloquear o teclado.	**21**	**Move 1 phrase (3 words) 1 position forward**	Você pode pressionar a esquerda **tecla de função** e a tecla de função direita, por sua vez, para desbloquear o teclado.	0	0	0	1	1	4,76%	3	14,29%	0	0	0	1	Você pode pressionar a esquerda tecla de função ~~esquerda~~ e a tecla de função direita, por sua vez, para desbloquear o teclado.	1	4,76%
	17	Atalhos - Para atribuir as funções **usadas com freqüência** às teclas de rolagem como teclas de	**17**	**Move 1 phrase (3 words) 1 position back**	Atalhos - Para atribuir as **usadas com freqüência** funções às teclas de rolagem como teclas de atalho.	0	0	0	1	1	5,88%	3	17,65%	0	0	0	1	Atalhos - Para atribuir as ~~funções~~ usadas com freqüência funções às teclas de rolagem como teclas de atalho.	1	5,88%
	18	Auto On e OFF - Para configurar **o telefone** para ligar ou desligar	**14**	**Move 1 phrase (2 words) 2 positions forward**	Auto On e OFF - Para configurar para ligar **o telefone** ou desligar automaticamente.	0	0	0	1	1	7,14%	2	14,29%	0	0	0	1	Auto On e OFF - Para configurar ~~o telefone~~ para ligar o telefone ou desligar automaticamente.	1	7,14%
	19	No modo de espera, você pode pressionar **e segurar** para mudar o perfil atual para o perfil silencioso.	**18**	**Move 1 phrase (2 words) 2 positions back**	No modo de espera, você **e segurar** pode pressionar para mudar o perfil atual para o perfil silencioso.	0	0	0	1	1	5,56%	2	11,11%	0	0	0	1	No modo de espera, você ~~pode pressionar~~ e segurar pode pressionar para mudar o perfil atual para o perfil silencioso.	1	5,56%
	20	Se o perfil atual estiver offline, você **não poderá usar** essa função.	**12**	**Move 1 phrase (3 words) 2 positions forward**	Se o perfil atual estiver offline, você essa função **não poderá usar.**	0	0	0	1	1	8,33%	3	25,00%	0	0	2	2	Se o perfil atual estiver offline, você essa função não poderá usar ~~essa função~~.	4	33,33%
	21	Configurações de exibição - Em seguida, você pode alterar a **senha de restrição** de chamadas.	**15**	**Move 1 phrase (3 words) 2 positions back**	Configurações de exibição - Em seguida, você pode **senha de restrição** alterar a de chamadas.	0	0	0	1	1	6,67%	3	20,00%	0	0	0	1	Configurações de exibição - Em seguida, você pode ~~alterar a~~ senha de restrição alterar a de chamadas.	1	6,67%
	22	A senha de restrição de chamadas de **idade** é fornecido pelo operador de rede se é a primeira vez para **você** mudar a senha.	**24**	**Delete 1 + 1 word (dift positions)**	A senha de restrição de chamadas de é fornecido pelo operador de rede se é a primeira vez para mudar a senha.	2	0	0	0	2	8,33%	2	8,33%	2	0	0	0	A senha de restrição de chamadas de ~~idade~~ é fornecido pelo operador de rede se é a primeira vez para ~~você~~ mudar a senha.	2	8,33%
	23	Chamada em espera - Esta opção é utilizada para activar ou desactivar a função de chamada em espera.	**18**	**Insert 1 + 1 word (dift positions)**	Chamada em espera **opcional** - Esta opção é utilizada para activar ou desactivar a função de chamada **internacional** em espera.	0	2	0	0	2	11,11%	2	11,11%	0	2	0	0	Chamada em espera opcional - Esta opção é utilizada para activar ou desactivar a função de chamada internacional em espera.	2	11,11%
	24	A função de **chamada** em espera depende da **rede.**	**9**	**Replace 1 + 1 word (dift positions)**	A função de **telefone** em espera depende da **chamada.**	0	0	2	0	2	22,22%	2	22,22%	0	0	2	0	A função de ~~chamada~~telefone em espera depende da ~~rede~~chamada.	2	22,22%
	25	Para obter **mais** informações, entre em contato com o centro de atendimento ao **cliente** local.	**15**	**Move 1 + 1 word one pos. fwd (dift positions)**	Para obter informações **mais,** entre em contato com o centro de atendimento ao local **cliente.**	0	0	0	2	2	13,33%	2	13,33%	0	0	4	0	Para obter ~~mais~~ informações mais, entre em contato com o centro de atendimento ao ~~cliente~~ local cliente.	4	26,67%
	26	Salvar **número** desconhecido - Esta função é usada para solicitar se deseja salvar o número após a conclusão de uma chamada.	**21**	**Delete 1 word + Insert 1 word (dift positions)**	Salvar desconhecido - Esta função é usada para solicitar se **você** deseja salvar o número após a conclusão de uma chamada.	1	1	0	0	2	9,52%	2	9,52%	1	1	0	0	Salvar ~~número~~ desconhecido - Esta função é usada para solicitar se você deseja salvar o número após a conclusão de uma chamada.	2	9,52%
	27	Remarcação automática - Esta **função** é utilizada para remarcar **automaticamente** o número discado se a discagem tiver falhado.	**18**	**Delete 1 word + Replace 1 word (dift positions)**	Remarcação automática - Esta é utilizada para remarcar **depressa** o número discado se a discagem tiver falhado.	1	0	1	0	2	11,11%	2	11,11%	1	0	1	0	Remarcação automática - Esta ~~função~~ é utilizada para remarcar ~~automaticamente~~depressa o número discado se a discagem tiver falhado.	2	11,11%

						EXPECTED								POST-EDIT COMPARE REPORT						
	No. Seg.	REFERENCE	LENGTH	ACTIONS	EDITED	Delete	Insert	Replace	Move	EDITing actions	Edits / length	No. Words edited	Edited words / length	Delete	Insert	Subs	Shift	Track changes	Nº of errors	TERP
GROUP 3	28	Modo de resposta - É utilizado para definir o modo para responder a uma chamada recebida.	16	Delete 1 word + Move 1 word one pos. fwd (dift positions)	Modo de resposta - É para definir o modo para a responder uma chamada recebida.	1	0	0	1	2	12,50%	2	12,50%	1	0	0	1	Modo de resposta - É ~~utilizado~~ para definir o modo para ~~responder~~ a responder uma chamada recebida.	2	12,50%
	29	Estão disponíveis três opções: qualquer tecla, tecla Enviar e deslizar para cima.	12	Insert 1 word + Delete 1 phrase (2wd) (dift	Estão disponíveis quase três opções: qualquer tecla, tecla Enviar para cima.	1	1	0	0	2	16,67%	3	25,00%	2	1	0	0	Estão disponíveis quase três opções: qualquer tecla, tecla Enviar ~~e deslizar~~ para cima.	3	25,00%
	30	Atendimento automático - Esta opção é usada para selecionar a resposta automática para chamadas recebidas.	15	Insert 1 word + Replace 1 phrase (2wd) (dift positions)	Atendimento automático imediato - Esta opção é usada quando quer a resposta automática para chamadas recebidas.	0	1	1	0	2	13,33%	3	20,00%	0	1	2	0	Atendimento automático imediato - Esta opção é usada ~~para~~quando ~~selecionar~~quer a resposta automática para chamadas recebidas.	3	20,00%
	31	Estão disponíveis quatro opções: Resposta imediata, Mensagem de texto, Voz gravada e Desactivar.	13	Insert 1 word + Move 1 phrase (2wd) one pos. fwd (dift positions)	Estão disponíveis mais quatro opções: Resposta imediata, Mensagem de texto, e Voz gravada Desactivar.	0	1	10	1	2	15,38%	3	23,08%	0	1	0	1	Estão disponíveis mais quatro opções: Resposta imediata, Mensagem de texto, e Voz gravada ~~e~~ Desactivar.	2	15,38%
	32	ID do chamador - Esta opção permite que você selecione se deseja exibir seu número de telefone no telefone da pessoa que você chamou.	24	Replace 1 word + Delete 1 phrase (3wd) (dift positions)	ID do telefone - Esta opção permite se deseja exibir seu número de telefone no telefone da pessoa que você chamou.	1	0	1	0	2	8,33%	4	16,67%	3	0	1	0	ID do ~~chamador~~telefone - Esta opção permite ~~que você selecione~~ se deseja exibir seu número de telefone no telefone da pessoa que você chamou.	4	16,67%
	33	Se seleccionar Auto, se o seu número de telefone será mostrado no telefone chamado depende da rede.	17	Replace 1 word + Insert 1 phrase (3wd) (dift positions)	Se seleccionar Ontem, se o seu número de telefone será mostrado no telefone chamado depende da rede e	0	1	1	0	2	11,76%	4	23,53%	0	3	2	0	Se seleccionar ~~Auto~~Ontem, se o seu número de telefone será mostrado no telefone chamado depende da rede e de si.	5	29,41%
	34	Selecionar linha - Esta opção é usada para definir a linha de chamada do telefone.	15	Replace 1 word + Move 1 phrase (3wd) one pos. (dift positions)	Selecionar telefone - Esta opção é a usada para definir linha de chamada do telefone.	0	0	1	1	2	13,33%	4	26,67%	0	0	1	1	Selecionar ~~linha~~telefone - Esta opção é a usada para definir ~~a~~ linha de chamada do telefone.	2	13,33%
	35	Meu número - Esta função pode ser usada para salvar seus próprios números de telefone no cartão SIM.	18	Delete 1 phrase (2wd) + Insert 1 phrase (2wd)	Meu número - Esta função usada para salvar seus próprios números de telefone e contactos no cartão SIM.	1	1	0	0	2	11,11%	4	22,22%	2	2	0	0	Meu número - Esta função ~~pode ser~~ usada para salvar seus próprios números de telefone e contactos no cartão SIM.	4	22,22%
	36	Videochamada - Pode alterar as definições da função Video Call.	10	Delete 1 phrase (2wd) + Replace 1 phrase (2wd)	Videochamada - Pode alterar as função Chamada Vídeo.	1	0	1	0	2	20,00%	4	40,00%	2	0	2	0	Videochamada - Pode alterar as ~~definições da~~ função ~~Video~~Chamada ~~Call~~Vídeo.	4	40,00%
	37	Modo de imagem - Para definir o modo de imagem durante uma chamada de	15	Delete 1 phrase (2wd) + Move 1 phrase (2wd) 2	Modo - Para definir de imagem o modo durante uma chamada de vídeo.	1	0	0	1	2	13,33%	4	26,67%	2	0	0	1	Modo ~~de imagem~~ - Para definir ~~o modo~~ de imagem o modo durante uma chamada de vídeo.	3	20,00%
GROUP 4	38	Qualidade da imagem - Para definir a qualidade da imagem para o vídeo.	13	Delete 1 wd + Insert 1 wd + Replace 1 wd	Qualidade da imagem - definir melhor a qualidade da imagem para o filme.	1	1	1	0	3	23,08%	3	23,08%	0	0	2	1	Qualidade da imagem - ~~Para~~ definir melhor a qualidade da imagem para o ~~vídeo~~filme.	3	23,08%
	39	Mudar câmara desligada - Para abrir ou fechar o canal de vídeo durante uma chamada de vídeo.	17	Delete 1 wd + Insert 1 wd + Move 1 wd 5 positions	Mudar câmara - Para abrir depressa ou fechar o de vídeo durante uma chamada canal de vídeo.	1	1	0	1	3	17,65%	3	17,65%	1	1	0	1	Mudar câmara ~~desligada~~ - Para abrir depressa ou fechar o ~~canal~~ de vídeo durante uma chamada canal de vídeo.	3	17,65%
	40	Áudio desligado - Para abrir ou fechar o canal de áudio durante uma chamada de vídeo.	16	Delete 1 wd + Replace 1 wd + Move 1 wd 7 positions	desligado - Para abrir ou activar o de áudio durante uma chamada de vídeo canal.	1	1	0	1	3	18,75%	3	18,75%	2	1	2	0	~~Áudio~~ desligado - Para abrir ou ~~fechar~~activar o ~~canal~~ de áudio durante uma chamada de vídeo canal.	5	31,25%

Appendix 10: Global analysis

nLinU	B.USER	USER	Job.profile	Does.MNG.ta	Does.PE.ta	YearsWork	Experience.profile	AgreePE4Learn	Typing.method	Use.CAT.suggestions	Predictive.writing	Use.MT	LikesDislikes.P	PE.speed	Opinion.on.f	Change.after.w	SPEED.Avg	TER.Avg.R	Editions.S	Editions.A	ZeroEdits.S	ZeroTER.a	Edited.TER	Edited.TER	SPEED.Avg	SPEED.SD
1	U.01	user-01	Freelancer	Rare	Irregular	4	Some_experience	Inclined to agree	Overwriting	Likes_suggestions	No_PW	No_or_rare	Likes_PE	Faster	No_opinion	New_view	34	36	125	4	0,3529	0,1775	0,2744	0,2193	6,8691	3,6348
2	U.02	user-02	Employed	Very_frequen	Irregular	2	No_experience	Inclined to agree	Typing	Likes_suggestions	No_PW	No_or_rare	Likes_PE	Faster	Optimist	No_changes	12	32	101	3	0,2683	0,2278	0,3113	0,2363	4,8029	2,6671
3	U.03	user-03	Freelancer	Very_rare	Very_rare	3	No_experience	Inclined to agree	Overwriting	Uses_suggestions	Uses_PW	Uses_MT	Likes_PE	Faster	Optimist	New_view	35	44	135	2	0,3814	0,1380	0,2231	0,2070	6,9130	3,6741
4	U.04	user-04	Freelancer	Irregular	Rare	3	No_experience	Inclined to agree	Typing	Likes_suggestions	Uses_PW	No_or_rare	Dislikes_PE	Slower	Optimist	New_view	18	29	118	4	0,1842	0,2679	0,3284	0,2210	5,3972	2,6331
5	U.05	user-05	Freelancer	Very_rare	Frequent	1	Some_experience	Inclined to agree	Typing	Likes_suggestions	Uses_PW	No_or_rare	Likes_PE	Faster	Optimist	New_view	20	24	132	4	0,2000	0,2803	0,3503	0,2395	5,4286	2,4562
6	U.06	user-06	Employed	Very_frequen	Frequent	3	Long_experience	Agree	Overwriting	Uses_suggestions	No_PW	Uses_MT	Likes_PE	Faster	Optimist	New_view	13	42	82	2	0,3704	0,1549	0,2461	0,1988	4,8233	2,6465
7	U.07	user-07	Employed	Very_frequen	Irregular	4	Some_experience	Agree	Overwriting	Uses_suggestions	No_PW	Uses_MT	Likes_PE	Faster	Optimist	No_changes	10	38	89	3	0,4694	0,1443	0,2720	0,1524	4,4876	2,6376
8	U.08	user-08	Freelancer	Frequent	Very_rare	3	No_experience	Inclined to agree	Overwriting	Likes_suggestions	No_PW	No_or_rare	Dislikes_PE	Slower	No_opinion	New_view	6	46	104	3	0,3019	0,1509	0,2161	0,1509	4,1453	1,9866
9	U.09	user-09	Employed	Very_rare	Rare	1	No_experience	Inclined to agree	Typing	Uses_suggestions	No_PW	No_or_rare	Dislikes_PE	Slower	Optimist	No_changes	26	10	159	4	0,0488	0,4053	0,4261	0,3187	6,0843	3,4818
10	U.14	user-14	Employed	Very_rare	Irregular	1	Long_experience	Agree	Overwriting	Likes_suggestions	Uses_PW	No_or_rare	Dislikes_PE	Faster	Optimist	New_view	25	20	166	5	0,1316	0,3202	0,3687	0,1594	6,0388	3,5836
11	U.15	user-15	Employed	Very_frequen	Rare	3	Some_experience	Inclined to agree	Overwriting	Likes_suggestions	No_PW	No_or_rare	Dislikes_PE	Slower	Optimist	New_view	15	35	148	4	0,2045	0,2364	0,2972	0,1960	5,0792	2,4514
12	U.16	user-16	Freelancer	Rare	Very_rare	4	Long_experience	Agree	Overwriting	Likes_suggestions	No_PW	Uses_MT	Likes_PE	Faster	Optimist	New_view	7	45	99	4	0,4318	0,1252	0,2203	0,1490	4,2270	2,1252
13	U.17	user-17	Employed	Very_frequen	Very_rare	1	No_experience	Inclined to disagre	Typing	Dislikes_suggestions	No_PW	No_or_rare	Dislikes_PE	Slower	Optimist	New_view	29	30	167	4	0,3881	0,1984	0,3242	0,2280	6,2953	3,5955
14	U.18	user-18	Freelancer	Very_rare	Irregular	4	Some_experience	Inclined to agree	Overwriting	Likes_suggestions	Uses_PW	Uses_MT	Likes_PE	Faster	No_opinion	New_view	41	15	222	6	0,2453	0,2942	0,3897	0,2373	8,4188	4,1371
15	U.19	user-19	Freelancer	Irregular	Irregular	2	Some_experience	Inclined to agree	Overwriting	Uses_suggestions	Uses_PW	No_or_rare	Likes_PE	Faster	Optimist	New_view	40	14	210	5	0,1800	0,3205	0,3909	0,2588	8,3598	7,0368
16	U.20	user-20	Freelancer	Very_rare	Frequent	1	Some_experience	Inclined to agree	Overwriting	Likes_suggestions	No_PW	No_or_rare	Likes_PE	Faster	Optimist	New_view	3	16	100	5	0,2593	0,2809	0,3792	0,2324	3,4533	1,6161
17	U.21	user-21	Freelancer	Frequent	Frequent	4	Some_experience	Agree	Overwriting	Likes_suggestions	No_PW	Uses_MT	Dislikes_PE	Faster	Optimist	New_view	44	31	64	3	0,2813	0,2298	0,3198	0,2744	8,9344	4,6049
18	U.22	user-22	Employed	Very_frequen	Very_frequ	4	Long_experience	Inclined to agree	Overwriting	Likes_suggestions	No_PW	Uses_MT	Dislikes_PE	Slower	No_opinion	No_changes	9	27	147	5	0,0938	0,3018	0,3330	0,1878	4,3068	2,3296
19	U.24	user-24	Employed	Irregular	Very_frequ	3	Long_experience	Inclined to agree	Overwriting	Uses_suggestions	Uses_PW	No_or_rare	Dislikes_PE	Slower	Optimist	New_view	43	11	162	5	0,2000	0,3377	0,4222	0,3215	8,4634	10,0988
20	U.25	user-25	Freelancer	Very_rare	Irregular	4	Some_experience	I don't know	Overwriting	Uses_suggestions	No_PW	No_or_rare	Dislikes_PE	Slower	Optimist	No_changes	33	40	97	4	0,4255	0,1469	0,2558	0,1579	6,8303	7,3146
21	U.26	user-26	Employed	Very_frequen	Irregular	1	Some_experience	Agree	Typing	Likes_suggestions	No_PW	No_or_rare	Likes_PE	Slower	Optimist	New_view	30	1	161	7	0,1154	0,5395	0,6099	0,3035	6,7303	7,1574
22	U.27	user-27	Employed	Rare	Frequent	2	Some_experience	Inclined to agree	Overwriting	Likes_suggestions	No_PW	No_or_rare	Dislikes_PE	Faster	Optimist	New_view	19	17	146	5	0,2000	0,2997	0,3747	0,2353	5,4107	2,4294
23	U.28	user-28	Freelancer	Very_rare	Very_frequ	1	Long_experience	Inclined to disagre	Typing	Likes_suggestions	No_PW	No_or_rare	Likes_PE	Faster	Optimist	No_changes	4	28	62	4	0,3704	0,2069	0,3285	0,2215	3,7818	1,8387
24	U.30	user-30	Employed	Irregular	Very_frequ	2	Long_experience	Agree	Overwriting	Likes_suggestions	Uses_PW	Uses_MT	Likes_PE	Faster	Optimist	New_view	42	6	230	6	0,1111	0,4043	0,4548	0,2509	8,4503	4,2986
25	U.31	user-31	Freelancer	Very_rare	Frequent	1	Some_experience	Inclined to agree	Overwriting	Uses_suggestions	Uses_PW	No_or_rare	Likes_PE	Faster	Optimist	No_changes	14	48	99	3	0,6211	0,0735	0,1939	0,1709	5,0179	2,1041
26	U.32	user-32	Freelancer	Frequent	Rare	2	No_experience	Inclined to agree	Typing	Likes_suggestions	No_PW	No_or_rare	Likes_PE	Slower	Optimist	New_view	27	21	171	5	0,1463	0,3116	0,3650	0,1850	6,1438	2,8791
27	U.33	user-33	Employed	Very_frequen	Very_frequ	2	Some_experience	I don't know	Overwriting	Uses_suggestions	Uses_PW	Uses_MT	Likes_PE	Slower	Optimist	New_view	36	19	173	5	0,1220	0,3245	0,3696	0,2218	7,3942	4,1103
28	U.34	user-34	Employed	Very_frequen	Very_frequ	3	Long_experience	I don't know	Overwriting	Likes_suggestions	No_PW	No_or_rare	Likes_PE	Faster	Optimist	New_view	45	18	237	4	0,0862	0,3385	0,3705	0,2238	9,5452	5,1845
29	U.35	user-35	Freelancer	Very_frequen	Frequent	4	Some_experience	Inclined to agree	Overwriting	Dislikes_suggestions	No_PW	No_or_rare	Likes_PE	Faster	Optimist	New_view	38	12	142	4	0,2222	0,3262	0,4194	0,3556	8,0095	13,1164
30	U.36	user-36	Freelancer	Very_rare	Very_rare	1	No_experience	I don't know	Overwriting	Dislikes_suggestions	No_PW	No_or_rare	Likes_PE	Faster	No_opinion	New_view	2	47	57	3	0,4324	0,1213	0,2137	0,1162	3,4018	2,4418
31	U.37	user-37	Employed	Irregular	Frequent	2	Some_experience	Agree	Overwriting	Uses_suggestions	Uses_PW	No_or_rare	Likes_PE	Faster	Optimist	New_view	23	26	176	4	0,2222	0,2693	0,3462	0,2287	5,9225	3,5918
32	U.38	user-38	Freelancer	Rare	Frequent	2	Some_experience	Agree	Typing	Uses_suggestions	Uses_PW	No_or_rare	Likes_PE	Faster	Optimist	New_view	22	34	144	4	0,3333	0,2054	0,3081	0,2209	5,5398	2,8244
33	U.39	user-39	Freelancer	Very_rare	Very_rare	2	No_experience	Inclined to disagre	Overwriting	Likes_suggestions	No_PW	Uses_MT	Likes_PE	Faster	Optimist	New_view	8	33	144	4	0,2826	0,2223	0,3098	0,2022	4,2909	2,6365
34	U.41	user-41	Freelancer	Very_rare	Very_rare	3	Some_experience	Agree	Overwriting	Likes_suggestions	Uses_PW	No_or_rare	Dislikes_PE	Faster	Optimist	New_view	17	39	95	4	0,2895	0,1830	0,2576	0,2060	5,3785	7,3532
35	U.46	user-46	Employed	Very_rare	Very_rare	1	No_experience	Agree	Typing	Likes_suggestions	Uses_PW	No_or_rare	Likes_PE	Slower	Optimist	New_view	11	41	82	3	0,4815	0,1301	0,2508	0,2325	4,5434	2,3417
36	U.50	user-50	Freelancer	Irregular	Very_rare	4	No_experience	Inclined to agree	Overwriting	Dislikes_suggestions	No_PW	No_or_rare	Dislikes_PE	Faster	Optimist	New_view	5	43	82	3	0,3333	0,1511	0,2267	0,1096	3,7958	2,4744
37	U.51	user-51	Freelancer	Very_rare	Rare	4	Some_experience	Agree	Overwriting	Uses_suggestions	Uses_PW	No_or_rare	Likes_PE	Faster	Optimist	New_view	24	37	90	3	0,3617	0,1745	0,2733	0,2068	5,9557	9,1999
38	U.52	user-52	Freelancer	Irregular	Very_frequ	4	Long_experience	Agree	Typing	Dislikes_suggestions	Uses_PW	No_or_rare	Dislikes_PE	Slower	Optimist	New_view	39	22	226	5	0,1379	0,3135	0,3637	0,2507	8,2780	4,3351
39	U.53	user-53	Employed	Very_frequen	Rare	4	Some_experience	Inclined to agree	Overwriting	Uses_suggestions	Uses_PW	Uses_MT	Dislikes_PE	Slower	No_opinion	No_changes	46	4	192	5	0,2400	0,3747	0,4931	0,3495	9,9702	5,7387
40	U.54	user-54	Employed	Very_frequen	Very_rare	1	No_experience	Agree	Typing	Likes_suggestions	No_PW	No_or_rare	Dislikes_PE	Slower	Optimist	New_view	37	23	156	5	0,3333	0,2340	0,3509	0,2145	7,8775	5,4242
41	U.55	user-55	Employed	Irregular	Very_frequ	1	Some_experience	Inclined to agree	Typing	Likes_suggestions	Uses_PW	Uses_MT	Likes_PE	Slower	Optimist	New_view	21	8	166	6	0,1471	0,3774	0,4425	0,2214	5,5309	3,2431
42	U.57	user-57	Freelancer	Irregular	Very_rare	4	Some_experience	Inclined to agree	Overwriting	Likes_suggestions	Uses_PW	No_or_rare	Likes_PE	Faster	Optimist	New_view	16	9	161	5	0,1500	0,3630	0,4270	0,2403	5,0999	2,5274
43	U.58	user-58	Freelancer	Frequent	Frequent	2	No_experience	Agree	Overwriting	Likes_suggestions	No_PW	No_or_rare	Likes_PE	Faster	Optimist	New_view	47	2	44	4	0,1667	0,4408	0,5289	0,2459	13,7105	15,7455
44	U.60	user-60	Freelancer	Irregular	Very_frequ	1	Long_experience	Agree	Overwriting	Likes_suggestions	No_PW	No_or_rare	Likes_PE	Faster	Optimist	New_view	31	3	162	5	0,1951	0,4170	0,5181	0,3832	6,7497	4,6317
45	U.64	user-64	Employed	Very_frequen	Very_rare	3	No_experience	Agree	Typing	Likes_suggestions	No_PW	No_or_rare	Likes_PE	Faster	No_opinion	No_changes	1	7	19	2	0,1111	0,4008	0,4509	0,3411	2,6329	1,2360
46	U.66	user-66	Employed	Very_rare	Very_frequ	1	Some_experience	Inclined to agree	Overwriting	Likes_suggestions	No_PW	No_or_rare	Dislikes_PE	Slower	Optimist	New_view	32	13	176	5	0,1957	0,3165	0,3935	0,2471	6,7533	4,2876
47	U.68	user-68	Freelancer	Very_rare	Irregular	2	Some_experience	Inclined to agree	Overwriting	Uses_suggestions	Uses_PW	Uses_MT	Likes_PE	Faster	Optimist	New_view	48	5	84	6	0,3333	0,3033	0,4549	0,2974	14,1438	13,0145
48	U.69	user-69	Freelancer	Irregular	Frequent	1	Long_experience	Inclined to agree	Overwriting	Likes_suggestions	Uses_PW	No_or_rare	Dislikes_PE	Slower	Optimist	New_view	28	25	188	4	0,2182	0,2732	0,3495	0,2260	6,2607	3,0873

		Questionnaires		Variables of result				
		Q2.1	Q2.2	Re-edited segments				
Texts	Type/token ratio	Very complex	Low quality	No. segments	No. repetitions	Avg.Duration	Avg.TER	Avg.Speed
B-Questionnaire	0,30	6	3	44	5	00:00:10	18%	00:00:12
C-Catalogue	0,35	14	9	37	10	00:00:11	31%	00:00:13
D-Manual	0,47	8	3	57	10	00:00:09	26%	00:00:14

MANUAL - AC | Segm. 17 | Average TER: 83%

No.LINE	USER	TEXT	SEGMENT	MODE	EVENT	SESSION	SOURCE	REFERENCE	RESULT	TIME	DURATION	Delete	Insert	Replace	Move	EDITS	LENGTH	TERp
N7826	user-60	D-Manual	17	AC	confirm-se	A	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque - perigo de choque eléctricos.	16:57:21	00:00:25	0	3	2	0	5	5	1
N7235	user-54	D-Manual	17	AC	confirm-se	A	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque - Perigo de choque.	11:59:16	00:00:17	0	2	1	0	3	5	0,6
N7061	user-53	D-Manual	17	AC	confirm-se	A	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque-Perigo de choques.	12:00:44	00:00:19	0	0	3	1	4	5	0,8
N6842	user-52	D-Manual	17	AC	confirm-se	A	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque eléctrico - Perigo de choque eléctrico.	12:00:45	00:00:32	0	4	1	0	5	5	1
N6668	user-51	D-Manual	17	AC	confirm-se	A	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque eléctrico.	12:01:09	00:00:35	1	0	1	1	3	5	0,6
N2986	user-21	D-Manual	17	AC	confirm-se	C	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque eléctrico - perigo de choque eléctricos.	17:21:01	00:00:22	0	4	1	0	5	5	1

MANUAL - PE | Segm. 17 | Average TER: 52%

No.LINE	USER	TEXT	SEGMENT	MODE	EVENT	SESSION	SOURCE	REFERENCE	RESULT	TIME	DURATION	Delete	Insert	Replace	Move	EDITS	LENGTH	TERp
N6214	user-41	D-Manual	17	PE	confirm-se	B	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Choque eléctrico - perigo de choque.	17:12:09	00:00:02	0	1	0	0	1	5	0,20
N4313	user-31	D-Manual	17	PE	confirm-se	B	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Choque eléctrico - perigo de choque.	12:10:56	00:00:01	0	1	0	0	1	5	0,20
N4952	user-34	D-Manual	17	PE	confirm-se	B	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo eléctrico - perigo de choque.	12:11:39	00:00:02	0	1	1	0	2	5	0,40
N6060	user-39	D-Manual	17	PE	confirm-se	B	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Choque elétrico: perigo de choque elétrico.	17:11:12	00:00:02	0	1	2	0	3	5	0,60
N6386	user-46	D-Manual	17	PE	confirm-se	B	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque - perigo de choque.	12:12:26	00:00:03	0	2	1	0	3	5	0,60
N5629	user-37	D-Manual	17	PE	confirm-se	B	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque - perigo de choque.	17:10:13	00:00:02	0	2	1	0	3	5	0,60
N3800	user-27	D-Manual	17	PE	confirm-se	B	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque – perigo de choque.	12:16:07	00:00:03	0	2	1	0	3	5	0,60
N5262	user-35	D-Manual	17	PE	confirm-se	B	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque eléctrico - perigo de choque.	17:12:49	00:00:01	0	3	0	0	3	5	0,60
N5836	user-38	D-Manual	17	PE	confirm-se	B	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque eléctrico.	17:09:08	00:00:01	1	0	1	1	3	5	0,60
N4158	user-30	D-Manual	17	PE	confirm-se	B	Shock hazard—Shock hazard.	Choque eléctrico perigo de choque.	Perigo de choque - Choque eléctrico.	12:16:41	00:00:01	0	1	2	1	4	5	0,80

MANUAL - AC | Segm. 10 | Average TER: 69%

No.LINE	USER	TEXT	SEGMENT	MODE	EVENT	SESSION	SOURCE	REFERENCE	RESULT	TIME	DURATION	Delete	Insert	Replace	Move	EDITS	LENGTH	TERp
N7221	user-54	D-Manual	10	AC	confirm-se	A	Product/Feature	Característica do produto	Característica do produto	11:57:17	00:00:05	0	0	0	0	0	3	0,00
N6654	user-51	D-Manual	10	AC	confirm-se	A	Product/Feature	Característica do produto	Característica do produto	11:57:58	00:00:21	0	0	0	0	0	3	0,00
N6499	user-50	D-Manual	10	AC	confirm-se	A	Product/Feature	Característica do produto	Característica do produto	12:00:06	00:00:02	0	0	0	0	0	3	0,00
N2292	user-17	D-Manual	10	AC	confirm-se	C	Product/Feature	Característica do produto	Característica do produto	17:20:47	00:00:06	0	0	0	0	0	3	0,00
N1168	user-07	D-Manual	10	AC	confirm-se	C	Product/Feature	Característica do produto	Característica do produto	12:21:41	00:00:04	0	0	0	0	0	3	0,00
N2972	user-21	D-Manual	10	AC	confirm-se	C	Product/Feature	Característica do produto	Característica/Produto	17:19:12	00:00:13	2	0	1	0	3	3	1,00
N8362	user-69	D-Manual	10	AC	confirm-se	A	Product/Feature	Característica do produto	Produto/Característica	16:54:26	00:00:13	2	0	1	0	3	3	1,00
N7047	user-53	D-Manual	10	AC	confirm-se	A	Product/Feature	Característica do produto	Produto/Característica	11:58:10	00:00:13	2	0	1	0	3	3	1,00
N6828	user-52	D-Manual	10	AC	confirm-se	A	Product/Feature	Característica do produto	Produto/Característica	11:58:29	00:00:19	2	0	1	0	3	3	1,00
N1004	user-06	D-Manual	10	AC	confirm-se	C	Product/Feature	Característica do produto	Produto/Característica	12:20:10	00:00:18	2	0	1	0	3	3	1,00
N0514	user-03	D-Manual	10	AC	confirm-se	C	Product/Feature	Característica do produto	Produto/Característica	12:20:39	00:00:19	2	0	1	0	3	3	1,00
N0267	user-02	D-Manual	10	AC	confirm-se	C	Product/Feature	Característica do produto	Produto/Característica	12:22:02	00:00:23	2	0	1	0	3	3	1,00
N0130	user-01	D-Manual	10	AC	confirm-se	C	Product/Feature	Característica do produto	Produto/Característica	12:20:43	00:00:17	2	0	1	0	3	3	1,00
N7583	user-57	D-Manual	10	AC	confirm-se	A	Product/Feature	Característica do produto	Produto/Função	16:55:50	00:00:24	2	0	1	0	3	3	1,00
N8035	user-66	D-Manual	10	AC	confirm-se	A	Product/Feature	Característica do produto	Produto/funcionalidade	11:58:56	00:00:16	2	0	1	0	3	3	1,00
N7806	user-60	D-Manual	10	AC	confirm-se	A	Product/Feature	Característica do produto	Produto/Funcionalidade	16:54:07	00:00:17	2	0	1	0	3	3	1,00

MANUAL - PE | Segm. 10 | Average TER: 64%

No.LINE	USER	TEXT	SEGMENT	MODE	EVENT	SESSION	SOURCE	REFERENCE	RESULT	TIME	DURATION	Delete	Insert	Replace	Move	EDITS	LENGTH	TERp
N6023	user-39	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Característica do produto	17:05:23	00:00:05	0	0	0	0	0	3	0
N4290	user-31	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Característica do produto	12:07:52	00:00:04	0	0	0	0	0	3	0
N3430	user-25	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Característica do produto	12:11:36	00:00:04	0	0	0	0	0	3	0
N2335	user-17	D-Manual	10	PE	confirm-se	D	Product/Feature	Característica do produto	Característica do produto	17:26:07	00:00:03	0	0	0	0	0	3	0
N1369	user-08	D-Manual	10	PE	confirm-se	D	Product/Feature	Característica do produto	Característica do produto	12:27:18	00:00:04	0	0	0	0	0	3	0
N1206	user-07	D-Manual	10	PE	confirm-se	D	Product/Feature	Característica do produto	Característica do produto	12:26:55	00:00:03	0	0	0	0	0	3	0
N4563	user-32	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Característica por produto	12:11:45	00:00:02	0	0	1	0	1	3	0,3333
N5422	user-36	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Características do produto	17:10:59	00:00:02	0	0	1	0	1	3	0,3333
N5219	user-35	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Características do produto	17:07:25	00:00:01	0	0	1	0	1	3	0,3333
N4735	user-33	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Características do produto	12:11:44	00:00:01	0	0	1	0	1	3	0,3333
N3030	user-21	D-Manual	10	PE	confirm-se	D	Product/Feature	Característica do produto	Característica/ Produto	17:25:56	00:00:01	1	0	1	0	2	3	0,6667
N1047	user-06	D-Manual	10	PE	confirm-se	D	Product/Feature	Característica do produto	produto Característica	12:26:55	00:00:08	1	0	0	1	2	3	0,6667
N5799	user-38	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Produto / Característica	17:03:38	00:00:03	0	0	1	2	3	3	1
N3634	user-26	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Produto / Característica	12:15:47	00:00:09	0	0	1	2	3	3	1
N3262	user-24	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Produto / Característica	12:13:01	00:00:02	0	0	1	2	3	3	1
N6180	user-41	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Produto/ Característica	17:06:58	00:00:04	1	0	1	1	3	3	1
N6361	user-46	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Produto/Característica	12:08:50	00:00:01	2	0	1	0	3	3	1
N4914	user-34	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Produto/Característica	12:08:28	00:00:02	2	0	1	0	3	3	1
N3941	user-28	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Produto/Característica	12:12:58	00:00:02	2	0	1	0	3	3	1
N3767	user-27	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Produto/Característica	12:09:39	00:00:04	2	0	1	0	3	3	1
N0561	user-03	D-Manual	10	PE	confirm-se	D	Product/Feature	Característica do produto	Produto/Característica	12:26:03	00:00:02	2	0	1	0	3	3	1
N0311	user-02	D-Manual	10	PE	confirm-se	D	Product/Feature	Característica do produto	Produto/Característica	12:28:17	00:00:03	2	0	1	0	3	3	1
N0171	user-01	D-Manual	10	PE	confirm-se	D	Product/Feature	Característica do produto	Produto/Característica	12:27:50	00:00:03	2	0	1	0	3	3	1
N4114	user-30	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Produto/Características	12:11:08	00:00:01	2	0	1	0	3	3	1
N0899	user-05	D-Manual	10	PE	confirm-se	D	Product/Feature	Característica do produto	Produto/Função	12:27:43	00:00:02	2	0	1	0	3	3	1
N5598	user-37	D-Manual	10	PE	confirm-se	B	Product/Feature	Característica do produto	Produto/recurso	17:05:16	00:00:02	2	0	1	0	3	3	1

CATALOGUE - AC | Segm. 5 | Average TER: 81%

No.LINE	USER	TEXT	SEGMENT	MODE	EVENT	SESSION	SOURCE	REFERENCE	RESULT	TIME	DURATION	Delete	Insert	Replace	Move	EDITS	LENGTH	TERp
N6582	user-50	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza Wet and Dry	12:21:27	00:00:11	1	0	0	1	2	7	0,2857
N3366	user-25	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	toalhetes de limpeza de superfície Wet and Dry	11:59:28	00:00:10	0	1	0	2	3	7	0,4286
N5728	user-38	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies Wet and Dry	16:52:32	00:00:13	1	2	0	1	4	7	0,5714
N4495	user-32	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza de superfícies Wet and Dry	11:56:57	00:00:24	1	2	0	1	4	7	0,5714
N4668	user-33	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza húmidos e secos	11:57:03	00:00:28	1	0	3	1	5	7	0,7143
N8519	user-69	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies húmidas e secas	17:19:25	00:00:49	0	1	4	1	6	7	0,8571
N8265	user-68	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies secas e úmidas	17:20:21	00:00:02	0	1	4	1	6	7	0,8571
N8181	user-66	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies secas e húmidas	12:21:41	00:00:05	0	1	4	1	6	7	0,8571
N7917	user-60	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies secas e molhadas	17:21:21	00:00:46	0	1	4	1	6	7	0,8571
N7742	user-58	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	toalhetes de limpeza para superfícies húmidas e secas.	17:22:05	00:00:21	0	1	4	1	6	7	0,8571
N7700	user-57	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes De Limpeza de Superfícies Secas e Húmidas	17:21:18	00:00:36	0	1	4	1	6	7	0,8571
N7323	user-54	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies secas e húmidas	12:19:51	00:00:29	0	1	4	1	6	7	0,8571
N7140	user-53	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza de superfícies húmidas e secas	12:18:49	00:00:04	0	1	4	1	6	7	0,8571
N7137	user-53	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza de superfícies húmidos e secos	12:18:38	00:00:23	0	1	4	1	6	7	0,8571
N6957	user-52	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies húmidas e secas	12:19:36	00:00:26	0	1	4	1	6	7	0,8571
N6767	user-51	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza de superfícies molhadas e secas	12:20:33	00:00:02	0	1	4	1	6	7	0,8571
N6317	user-46	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies molhadas e secas	11:58:44	00:00:40	0	1	4	1	6	7	0,8571
N5525	user-37	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza de superfícies secas e molhadas	16:50:37	00:00:49	0	1	4	1	6	7	0,8571
N4840	user-34	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies secas e húmidas	11:56:29	00:00:27	0	1	4	1	6	7	0,8571
N4243	user-31	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza secos e húmidos de superfícies	11:57:16	00:00:55	0	1	4	1	6	7	0,8571
N4027	user-30	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies húmidas e secas	11:56:29	00:00:26	0	1	4	1	6	7	0,8571

N3710	user-27	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza secos e húmidos de superfícies	11:57:52	00:00:54	0	1	4	1	6	7	0,8571
N3553	user-26	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies húmidas e secas	12:01:22	00:01:01	0	1	4	1	6	7	0,8571
N7503	user-55	C-Catalogue	5	AC	confirm-se	C	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes secos e humídos de limpeza de superfícies	17:20:30	00:00:33	0	1	4	2	7	7	1
N5959	user-39	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhitas de limpeza para superfícies húmidas ou secas	16:52:23	00:00:06	0	1	5	1	7	7	1
N5122	user-35	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Superfícies de limpeza para superfícies secas e húmidas	16:55:16	00:00:01	0	1	5	1	7	7	1
N3196	user-24	C-Catalogue	5	AC	confirm-se	A	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza húmidos e a seco para superfícies	11:56:33	00:00:42	0	2	4	1	7	7	1

CATALOGUE - PE Segm. 5 Average TER: 75%

No.LINE	USER	TEXT	SEGMENT	MODE	EVENT	SESSION	SOURCE	REFERENCE	RESULT	TIME	DURATION	Delete	Insert	Replace	Move	EDITS	LENGTH	TERp
N6612	user-50	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza Wet and Dry	12:26:06	00:00:02	1	0	0	1	2	7	0,2857
N2024	user-16	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	toalhetes de limpeza de superfície Wet and Dry	17:05:40	00:00:11	0	1	0	2	3	7	0,4286
N7349	user-54	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza	12:26:12	00:00:18	4	0	0	0	4	7	0,5714
N7170	user-53	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza	12:24:40	00:00:15	4	0	0	0	4	7	0,5714
N7085	user-53	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza	12:07:11	00:00:18	4	0	0	0	4	7	0,5714
N2413	user-18	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	toalhetes de limpeza	17:02:03	00:00:09	4	0	0	0	4	7	0,5714
N1094	user-07	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	toalhetes de limpeza	12:07:21	00:00:24	4	0	0	0	4	7	0,5714
N0968	user-06	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	toalhetes de limpeza	12:12:50	00:00:00	4	0	0	0	4	7	0,5714
N0650	user-04	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza de superfícies Wet and Dry	12:09:03	00:00:03	1	2	0	1	4	7	0,5714
N0221	user-02	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza de superfícies Wet and Dry	12:09:38	00:00:00	1	2	0	1	4	7	0,5714
N8552	user-69	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies húmidas e secas	17:24:23	00:00:02	0	1	4	1	6	7	0,8571
N8296	user-68	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfície seca e úmida	17:23:53	00:00:01	0	1	3	2	6	7	0,8571
N8211	user-66	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies húmidas e secas	12:24:46	00:00:06	0	1	4	1	6	7	0,8571
N8007	user-64	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza de superfícies húmidas e secas	17:27:31	00:00:01	0	1	4	1	6	7	0,8571
N7953	user-60	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies secas e molhadas	17:26:35	00:00:17	0	1	4	1	6	7	0,8571
N7770	user-58	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	toalhetes de limpeza para superfícies húmidas e secas	17:26:15	00:00:00	0	1	4	1	6	7	0,8571
N7732	user-57	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies secas e húmidas	17:27:46	00:00:01	0	1	4	1	6	7	0,8571
N7541	user-55	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza secos e húmidos para superfícies	17:25:29	00:00:01	0	1	4	1	6	7	0,8571
N6990	user-52	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies húmidas e secas	12:24:40	00:00:02	0	1	4	1	6	7	0,8571
N6799	user-51	C-Catalogue	5	PE	confirm-se	D	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	toalhetes de limpeza de superfícies molhadas e secas	12:27:32	00:00:27	0	1	4	1	6	7	0,8571
N3086	user-22	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies húmidas e secas	17:04:21	00:00:05	0	1	4	1	6	7	0,8571
N2845	user-20	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	toalhetes de limpeza para superfícies secas ou molhadas	17:02:58	00:00:01	0	1	4	1	6	7	0,8571
N2641	user-19	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies secas e húmidas	17:01:41	00:00:00	0	1	4	1	6	7	0,8571
N2190	user-17	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	toalhetes de superfície húmida e seca de limpeza	17:00:52	00:00:01	0	1	3	2	6	7	0,8571
N1808	user-15	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza de superfícies	17:06:34	00:00:22	2	0	3	1	6	7	0,8571
N1594	user-14	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza secos e húmidos para superfícies	17:04:20	00:00:04	0	1	4	1	6	7	0,8571
N1446	user-09	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies secas e molhadas	12:11:11	00:00:02	0	1	4	1	6	7	0,8571
N1275	user-08	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	toalhetes de limpeza para superfície molhada ou seca	12:09:28	00:00:01	0	1	3	2	6	7	0,8571
N0793	user-05	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies molhadas e secas	12:10:46	00:00:00	0	1	4	1	6	7	0,8571
N0420	user-03	C-Catalogue	5	PE	confirm-se	B	Wet and Dry Surface Cleaning Wipes	Wet and Dry superfície toalhetes de limpeza	Toalhetes de limpeza para superfícies molhadas e secas	12:06:54	00:00:02	0	1	4	1	6	7	0,8571

CATALOGUE - AC Segm. 6 Average TER: 61%

No.LINE	USER	TEXT	SEGMENT	MODE	EVENT	SESSION	SOURCE	REFERENCE	RESULT	TIME	DURATION	Delete	Insert	Replace	Move	EDITS	LENGTH	TERp
N6584	user-50	C-Catalogue	6	AC	confirm-se	C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies de mesa, laptop e teclado limpos de sujeira e poeira com oss <tag> toalhetes de limpeza Wet and Dry.	12:22:06	00:00:39	1	0	6	1	8	24	0,3333
N6319	user-46	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies de mesa, laptop e teclado limpas de sujeira e poeira com os toalhetes de limpeza para superfície molhada e seca.	12:00:08	00:01:24	0	0	6	2	8	24	0,3333
N3368	user-25	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies de mesa, laptop e teclado limpas de sujeira e poeira com os toalhetes de limpeza <tag>	11:59:40	00:00:12	4	0	3	1	8	24	0,3333
N6763	user-51	C-Catalogue	6	AC	confirm-se	C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies de mesa, laptop e teclado limpas de sujeira e poeira com os toalhetes de limpeza de superfícies molhadas e secas.	12:20:12	00:00:00	1	0	6	2	9	24	0,375
N4245	user-31	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies de mesa, laptop e teclado limpas de sujidade e poeira com os <tag> toalhetes de limpeza secos e húmidos.	11:58:56	00:01:40	2	0	6	2	10	24	0,4167
N5124	user-35	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies de mesa, portátil e teclado limpas de sujidade e poeira com os <tag> toalhetes de limpeza para superfícies secas e molhadas.	16:55:17	00:00:01	0	1	8	2	11	24	0,4583
N4497	user-32	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies de mesa, laptop e teclado limpas de sujidade e pó com os toalhetes de limpeza Wet and Dry.	11:58:15	00:01:18	3	0	8	1	12	24	0,5
N8264	user-68	C-Catalogue	6	AC	confirm-se	C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da mesa, do laptop e do teclado sem sujeira e poeira com os toalhetes de limpeza para superfícies secas e úmidas.	17:20:19	00:00:00	1	1	9	3	14	24	0,5833
N7744	user-58	C-Catalogue	6	AC	confirm-se	C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da mesa, laptop e teclado limpas e livres de pó e sujidade com os toalhetes de limpeza <tag> para superfícies húmidas e secas.	17:23:34	00:01:23	1	3	9	2	15	24	0,625
N7505	user-55	C-Catalogue	6	AC	confirm-se	C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da secretária, computador portátil e teclado livres de sujidade e pó com os toalhetes de limpeza secos e húmidos.	17:21:49	00:01:19	2	0	11	2	15	24	0,625

N6959	user-52	C-Catalogue	6	AC	confirm-se	C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua mesa, computador portátil e teclado limpas de sujidade e poeira com os toalhetes de limpeza para superfícies húmidas e secas.	12:20:30	00:00:54	0	1	11	3	15	24	0,625
N5725	user-38	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha a superfície da sua mesa, laptop e teclado limpa, sem sujeira e pó com os toalhetes de limpeza de superfícies Wet and Dry.	16:52:11	00:02:02	0	0	14	1	15	24	0,625
N8521	user-69	C-Catalogue	6	AC	confirm-se	C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha a superfície da sua secretária, laptop e teclado livre de sujidade e pó com os toalhetes de limpeza para superfícies húmidas e secas.	17:21:07	00:01:42	0	0	14	2	16	24	0,6667
N8183	user-66	C-Catalogue	6	AC	confirm-se	C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies da mesa, do computador portátil e do teclado limpas de sujidade e de poeira com os toalhetes de limpeza para superfícies secas e húmidas.	12:21:48	00:00:04	0	4	9	3	16	24	0,6667
N7919	user-60	C-Catalogue	6	AC	confirm-se	C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha a sua mesa, laptop e teclado sem sujidades e poeiras com os toalhetes de limpeza para superfícies secas e molhadas <tag>.	17:22:49	00:01:28	2	0	10	4	16	24	0,6667
N5961	user-39	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua mesa, computador portátil e teclado isentas de sujidade e pó com as Toalhitas de Limpeza para Superfíces Húmidas ou Secas.	16:53:53	00:01:30	0	1	13	2	16	24	0,6667
N5527	user-37	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua mesa, do laptop e do teclado limpas de sujidade e pó com os toalhetes de limpeza de superfícies secas e molhadas.	16:51:55	00:01:17	0	2	11	3	16	24	0,6667
N7702	user-57	C-Catalogue	6	AC	confirm-se	C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies do seu computador, portátil e teclado sem sujidade e pó com os <tag> Toalhetes de Limpeza para Superfícies Húmidas e Secas.	17:23:14	00:01:56	1	1	12	3	17	24	0,7083
N7325	user-54	C-Catalogue	6	AC	confirm-se	C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua mesa, portátil e teclado limpos e sem pó com os toalhetes de limpeza para superfícies secas e húmidas.	12:21:02	00:01:11	1	0	12	4	17	24	0,7083
N4670	user-33	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha a sua mesa, portátil e teclado limpos com os toalhetes de limpeza húmidos e secos.	11:58:10	00:01:07	8	0	8	1	17	24	0,7083
N3712	user-27	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha limpas as superfícies de mesas, portáteis e teclados sem sujidade nem pó com os toalhetes de limpeza húmidos e secos de superfícies.	11:59:47	00:01:55	1	0	13	3	17	24	0,7083
N7142	user-53	C-Catalogue	6	AC	confirm-se	C	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua secretária, portátil e teclado limpas e isentas de pó com os toalhetes de limpeza de superfícies húmidas e secas.	12:19:43	00:00:54	0	0	13	5	18	24	0,75
N4842	user-34	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies das suas mesas, computadores portáteis e teclados sem sujidade e poeiras com os toalhetes de limpeza <tag> para superfícies húmidas e secas.	11:57:19	00:00:50	0	1	13	4	18	24	0,75
N4029	user-30	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua secretária, computador portátil e teclado sem sujidade e poeira utilizando os toalhetes de limpeza para superfícies húmidas e secas.	11:57:45	00:01:16	1	1	13	3	18	24	0,75
N3555	user-26	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha a sua secretária, o seu portátil e o seu teclado livres de pó e sujidade com os toalhetes de limpeza para superfícies húmidas e secas.	12:01:52	00:00:30	0	2	15	2	19	24	0,7917
N3198	user-24	C-Catalogue	6	AC	confirm-se	A	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha limpa a superfície da secretária, computador e teclado, sem sujidade nem poeira, graças aos toalhetes de limpeza húmidos ou a seco para superfícies.	11:58:59	00:02:26	0	0	20	1	21	24	0,875

CATALOGUE - PE | Segm. 6 | Average TER: 53%

No.LINE	USER	TEXT	SEGMENT	MODE	EVENT	SESSION	SOURCE	REFERENCE	RESULT	TIME	DURATION	Delete	Insert	Replace	Move	EDITS	LENGTH	TERp
N6804	user-51	C-Catalogue	6	PE	confirm-se	D	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies de mesa, laptop e teclado limpas de sujeira e poeira com os molhadas e superfície seca toalhetes de limpeza.	12:28:18	00:00:00	1	0	2	0	3	24	0,125
N0981	user-06	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies de secretária, laptop e teclado limpo de sujeira e poeira com as molhadas e superfície seca limpeza.	12:16:16	00:00:21	4	0	1	0	5	24	0,2083
N0429	user-03	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as toalhetes de limpeza de superfícies molhadas e superfície.	12:08:52	00:00:01	0	0	4	2	6	24	0,25
N6621	user-50	C-Catalogue	6	PE	confirm-se	D	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies de mesa, laptop e teclado limpas de sujeira e poeira com os toalhetes de limpeza	12:27:16	00:00:11	5	0	3	0	8	24	0,3333
N1296	user-08	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies de mesa, laptop e teclado limpas de sujeira e poeira com as toalhetes de limpeza para superfície molhada ou seca .	12:14:43	00:00:09	0	1	5	2	8	24	0,3333
N1286	user-08	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies de mesa, laptop e teclado limpas de sujeira e poeira com as toalhetes de limpeza para superfície molhada ou seca .	12:11:46	00:00:02	0	1	5	2	8	24	0,3333
N1106	user-07	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies de mesa, e teclado limpo de sujeira e poeira com os toalhetes de limpeza wet and dry .	12:10:48	00:00:02	2	0	6	1	9	24	0,375

N0235	user-02	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies de mesa, laptop e teclado limpas de sujidade e poeira com toalhetes de limpeza Wet and Dry.	12:14:36	00:00:10	3	0	6	1	10	24	0,4167
N2201	user-17	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies de mesa, portátil e teclado limpos de sujidade e pó com as toalhitas de limpeza de superfície húmida e seca .	17:03:09	00:00:01	0	1	8	3	12	24	0,5
N2655	user-19	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as suas superfícies da sua mesa, laptop e teclado limpas de sujidade e poeiras com os toalhetes de limpeza para superfícies secas e húmidas.	17:04:42	00:00:02	0	1	10	2	13	24	0,5417
N7101	user-53	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua mesa, portátil e teclado limpos e sem pó e com os toalhetes de limpeza.	12:10:02	00:00:39	5	0	7	2	14	24	0,5833
N2049	user-16	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha superfície da sua mesa , laptop e teclado limpos de sujidade e poeira com os toalhetes de limpeza Wet and Dry .	17:11:26	00:00:04	2	0	10	2	14	24	0,5833
N0809	user-05	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies de secretárias, portáteis e teclados limpas de sujidade e pó com os toalhetes de limpeza para superfícies molhadas e secas.	12:13:24	00:00:09	1	0	11	2	14	24	0,5833
N0665	user-04	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua secretária, laptop e teclado livres de sujidade e poeira com os toalhetes de limpeza de superfícies Wet and Dry.	12:12:54	00:00:05	0	0	12	2	14	24	0,5833
N8565	user-69	C-Catalogue	6	PE	confirm-se	D	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua secretária, laptop e teclado livres de sujidade e pó com os toalhetes de limpeza para superfícies húmidas e secas.	17:27:17	00:00:01	0	0	12	3	15	24	0,625
N8313	user-68	C-Catalogue	6	PE	confirm-se	D	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha a superfície da mesa, do laptop e do teclado sem sujeira e poeira com os toalhetes de limpeza para superfície seca e úmida .	17:25:17	00:00:00	0	1	12	2	15	24	0,625
N7000	user-52	C-Catalogue	6	PE	confirm-se	D	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua mesa, computador portátil e teclado limpas de sujidade e poeira com os toalhetes de limpeza para superfícies húmidas e secas.	12:26:07	00:00:06	0	1	11	3	15	24	0,625
N3099	user-22	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies de secretárias, portáteis e teclados livres de sujidade e pó com os toalhetes de limpeza para superfícies húmidas e secas.	17:07:53	00:00:14	1	0	12	2	15	24	0,625
N1827	user-15	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da secretária, portátil e teclado limpas de sujidade e pó com os toalhetes de limpeza de superfícies úmidas e secas.	17:09:43	00:00:09	1	0	12	2	15	24	0,625
N1615	user-14	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies de mesas, portáteis e teclados sem sujidade e pó com os toalhetes de limpeza secos e húmidos para superfícies.	17:07:24	00:00:01	2	0	12	1	15	24	0,625
N1463	user-09	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha a sua mesa, portátil e teclado limpos de sujidade e de pó com os toalhetes de limpeza para superfícies molhadas e secas.	12:14:35	00:00:02	1	0	10	4	15	24	0,625
N7560	user-55	C-Catalogue	6	PE	confirm-se	D	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da secretária, computador de secretária e teclado livres de sujidade e pó com os toalhetes de limpeza secos e húmidos para superficies.	17:27:41	00:00:01	0	1	12	3	16	24	0,6667
N7189	user-53	C-Catalogue	6	PE	confirm-se	D	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua secretária, portátil e teclado limpas e isentas de pó com os toalhetes de limpeza.	12:26:30	00:00:15	5	0	8	3	16	24	0,6667
N8231	user-66	C-Catalogue	6	PE	confirm-se	D	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da mesa, do computador portátil e do teclado limpas de sujidade e de poeira com os toalhetes de limpeza para superfícies húmidas e secas.	12:26:46	00:00:02	1	4	9	3	17	24	0,7083
N7362	user-54	C-Catalogue	6	PE	confirm-se	D	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua mesa, portátil e teclado limpas e sem pó com de limpeza	12:29:12	00:00:14	8	0	7	2	17	24	0,7083
N2867	user-20	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua secretária, portátil e teclado sem pó e sujidades com os toalhetes de limpeza para superfícies secas ou molhadas.	17:07:15	00:00:02	1	0	14	2	17	24	0,7083
N2435	user-18	C-Catalogue	6	PE	confirm-se	B	Keep your desk, laptop and keyboard surfaces clean of dirt and dust with the <tag> Wet and Dry Surface Cleaning Wipes.	Mantenha as suas superfícies de mesa, laptop e teclado limpo de sujeira e poeira com as <tag> molhadas e superfície seca toalhetes de limpeza.	Mantenha as superfícies da sua mesa, do seu computador e do teclado limpo de sujidade e poeira uilizando os toalhetes de limpeza para superfícies secas e molhadas.	17:04:17	00:00:01	0	3	11	3	17	24	0,7083

CATALOGUE - AC Segm. 17 Average TER: 62%

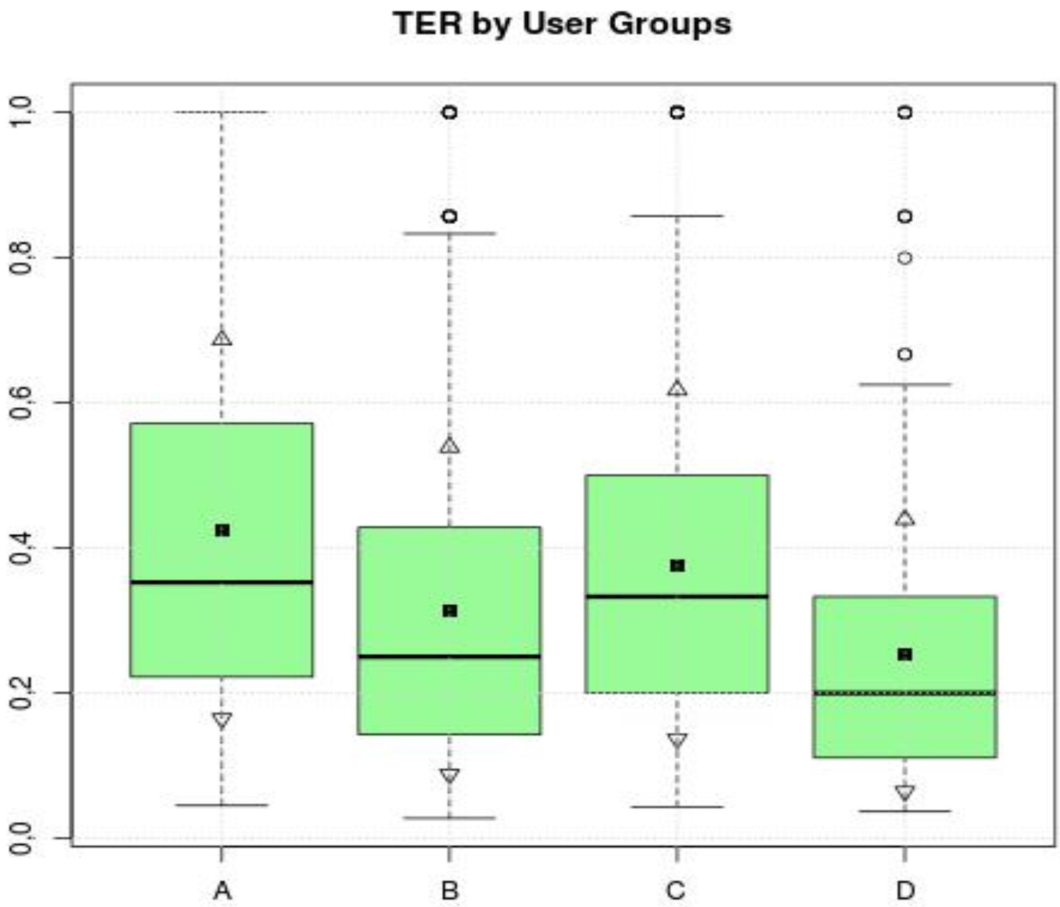
No.LINE	USER	TEXT	SEGMENT	MODE	EVENT	SESSION	SOURCE	REFERENCE	RESULT	TIME	DURATION	Delete	Insert	Replace	Move	EDITS	LENGTH	TERp
N5549	user-37	C-Catalogue	17	AC	confirm-se	A	Thoroughly clean telephone sets, keyboards and other hard surfaces with the <tag> Anti-Static Cleaning Cloths.	conjuntos completamente limpas de telefone, teclados e outras superfícies duras com a <tag> Anti-estático panos de limpeza.	Aparelhos de telefone, teclados e outras superfícies duras completamente limpos com os panos de limpeza anti-estáticos.	16:56:55	00:00:50	2	1	4	2	9	17	0,5294
N4864	user-34	C-Catalogue	17	AC	confirm-se	A	Thoroughly clean telephone sets, keyboards and other hard surfaces with the <tag> Anti-Static Cleaning Cloths.	conjuntos completamente limpas de telefone, teclados e outras superfícies duras com a <tag> Anti-estático panos de limpeza.	Limpe telefones, teclados e outras superfícies rígidas com os toalhetes de limpeza anti-estática <tag>.	12:01:51	00:00:45	3	0	8	1	12	17	0,7059

CATALOGUE - PE Segm. 17 Average TER: 40%

No.LINE	USER	TEXT	SEGMENT	MODE	EVENT	SESSION	SOURCE	REFERENCE	RESULT	TIME	DURATION	Delete	Insert	Replace	Move	EDITS	LENGTH	TERp
N2721	user-19	C-Catalogue	17	PE	confirm-se	B	Thoroughly clean telephone sets, keyboards and other hard surfaces with the <tag> Anti-Static Cleaning Cloths.	conjuntos completamente limpas de telefone, teclados e outras superfícies duras com a <tag> Anti-estático panos de limpeza.	Limpe cuidadosamente conjuntos de telefone, teclados e outras superfícies duras com os panospanos de limpeza anti-estáticos.	17:12:51	00:00:01	1	0	6	2	9	17	0,5294

N2492	user-18	C-Catalogue	17	PE	confirm-se B	Thoroughly clean telephone sets, keyboards and other hard surfaces with the <tag> Anti-Static Cleaning Cloths.	conjuntos completamente limpas de telefone, teclados e outras superfícies duras com a <tag> Anti-estático panos de limpeza.	Limpeza profunda de telefones, teclados e outras superfícies rígidas com os panos de limpeza Anti-estático .	17:11:20	00:00:11	1	0	7	1	9	17	0,5294
N2253	user-17	C-Catalogue	17	PE	confirm-se B	Thoroughly clean telephone sets, keyboards and other hard surfaces with the <tag> Anti-Static Cleaning Cloths.	conjuntos completamente limpas de telefone, teclados e outras superfícies duras com a <tag> Anti-estático panos de limpeza.	conjuntos de telefones completamente limpos , teclados e outras superfícies duras com a Anti-estático panos de limpeza.	17:10:49	00:00:10	1	0	2	2	5	17	0,2941
N0472	user-03	C-Catalogue	17	PE	confirm-se B	Thoroughly clean telephone sets, keyboards and other hard surfaces with the <tag> Anti-Static Cleaning Cloths.	conjuntos completamente limpas de telefone, teclados e outras superfícies duras com a <tag> Anti-estático panos de limpeza.	Conjuntos completamente limpos de telefone, teclados e outras superfícies duras com a panos de limpeza Anti-estático	12:13:36	00:00:34	1	0	2	1	4	17	0,2353

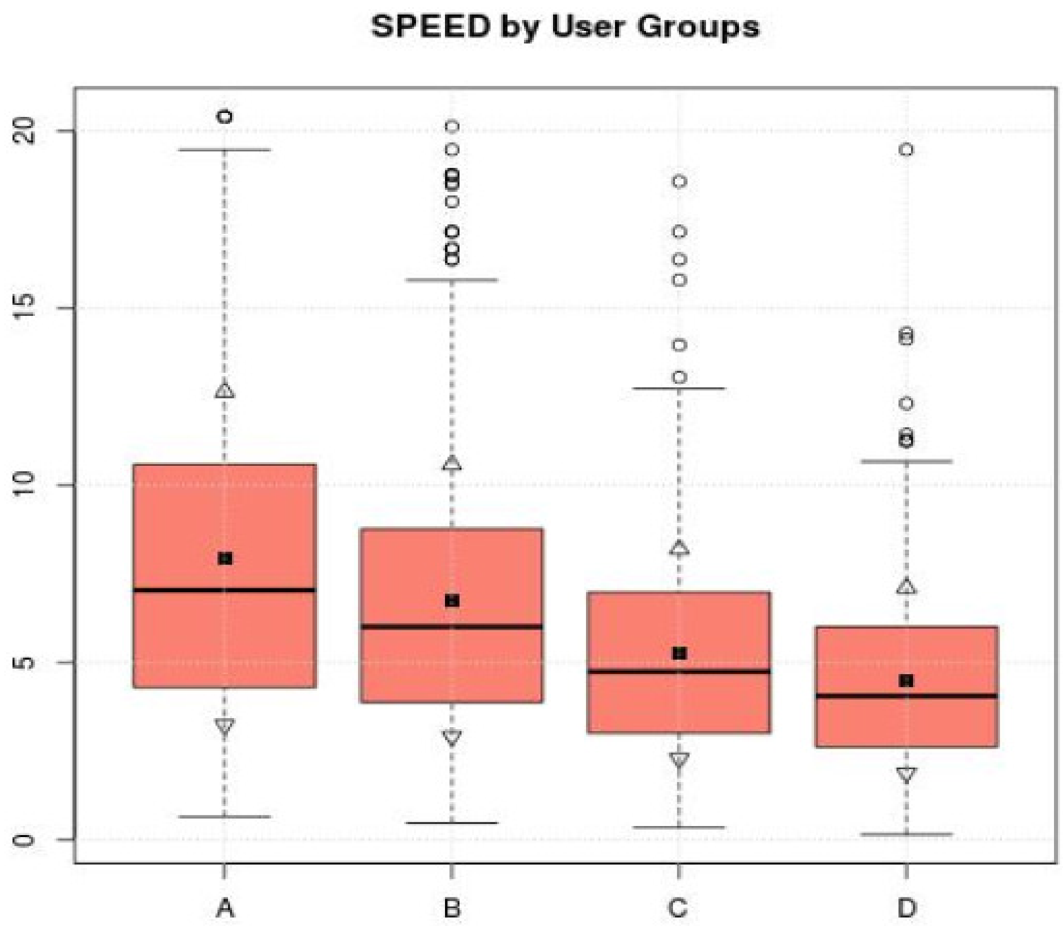
- Result variable: TER
- Grouping variable: 4ranks.avgUser.TERxSPEED



Note:
Green box sizes are defined by the first and third quartiles; black squares inside the green boxes mark averages and lines medians.
Whiskers (horizontal lines outside green boxes) represent the ranges, and circles outside whiskers are outliers.

TER	A	B	C	D
Count	396	452	346	358
Minimum	0,05	0,03	0,04	0,04
1st Quartile	0,22	0,14	0,20	0,11
Median	0,35	0,25	0,33	0,20
Mean	0,43	0,31	0,38	0,25
3rd Quartile	0,57	0,43	0,50	0,33
Maximum	1,00	1,00	1,00	1,00
Standard deviation	0,26	0,23	0,24	0,19

- Result variable: SPEED
- Grouping variable(s): 4ranks.avgUser.TERxSPEED



SPEED	A	B	C	D
Count	396	452	346	358
Minimum	0,64	0,46	0,35	0,14
1st Quartile	4,29	3,87	3,00	2,61
Median	7,04	6,00	4,74	4,05
Mean	7,93	6,74	5,23	4,49
3rd Quartile	10,59	8,76	6,97	6,00
Maximum	20,40	20,12	18,57	19,46
Standard deviation	4,70	3,84	2,96	2,62

TER

	A	B	C	D
average	0,45382	0,31479	0,37963	0,25130
Standard Deviation	0,07167	0,04845	0,04569	0,03947
max	0,60988	0,36959	0,45089	0,31131
min	0,37048	0,22312	0,32839	0,19391
25,00%	0,37048	0,22312	0,32839	0,19391
50,00%	0,37048	0,22312	0,32839	0,19391
75,00%	0,37048	0,22312	0,32839	0,19391
80,00%	0,37048	0,22312	0,32839	0,19391
90,00%	0,37048	0,22312	0,32839	0,19391
95,00%	0,37048	0,22312	0,32839	0,19391
99,00%	0,37048	0,22312	0,32839	0,19391
SKEW	0,93865	-0,58946	0,43954	0,34058
KURTOSIS	0,39122	-0,81324	-1,44663	-1,11296

SPEED

	A	B	C	D
average	9,10873	6,94098	4,92398	4,49946
Standard Deviatio	2,47397	1,01406	1,12001	0,56567
max	14,14376	8,93442	6,08429	5,37847
min	6,73033	5,53977	2,63288	3,40180
25,00%	6,73033	5,53977	2,63288	3,40180
50,00%	6,73033	5,53977	2,63288	3,40180
75,00%	6,73033	5,53977	2,63288	3,40180
80,00%	6,73033	5,53977	2,63288	3,40180
90,00%	6,73033	5,53977	2,63288	3,40180
95,00%	6,73033	5,53977	2,63288	3,40180
99,00%	6,73033	5,53977	2,63288	3,40180
SKEW	1,32230	0,67228	-0,96361	-0,40344
KURTOSIS	0,96321	-0,25021	-0,15418	-0,15239

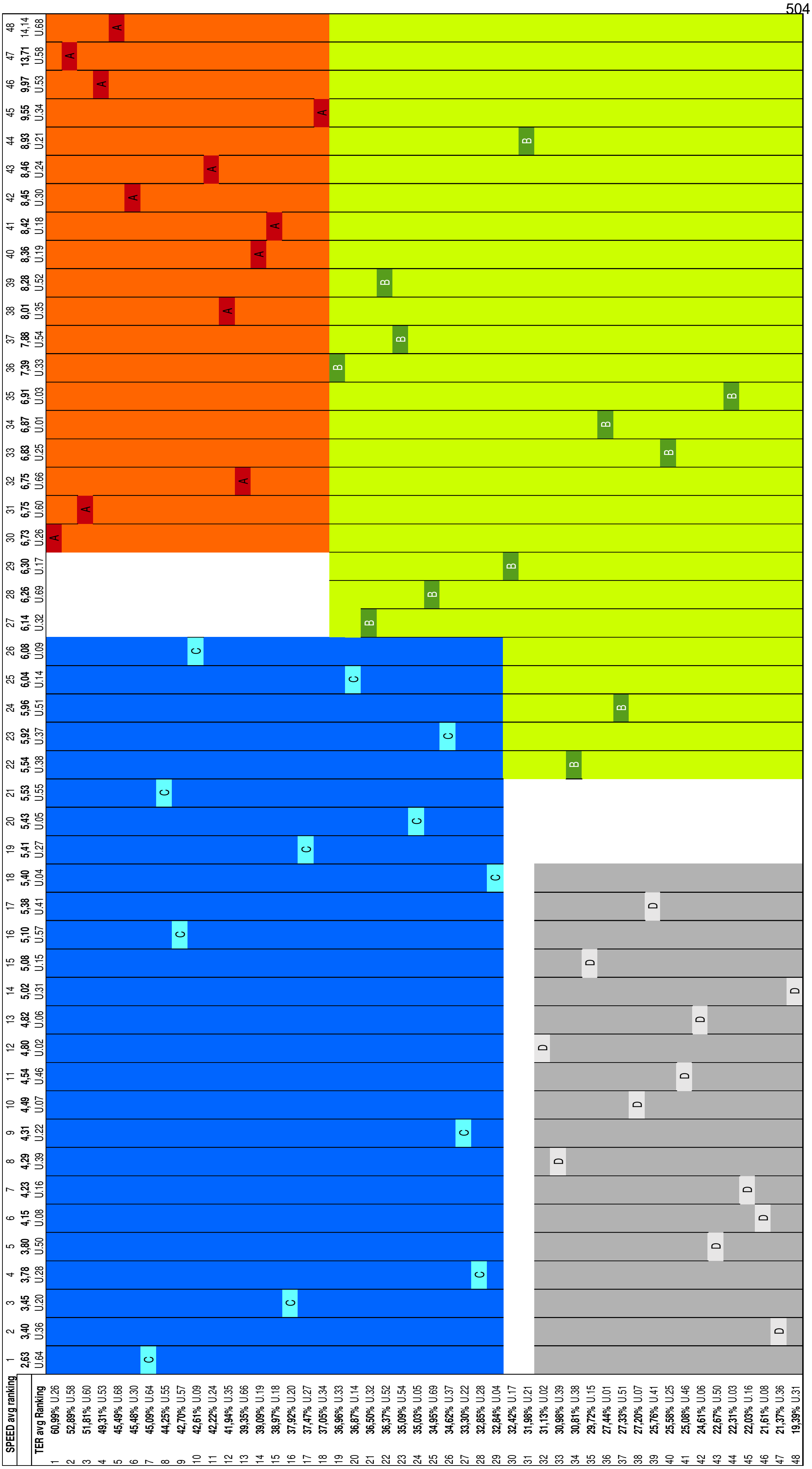
Average - Edited.Tk.avgU.TERxSPEED

USER	A	B	C	D	Total Result
user-01		0,2744			**0,2744**
user-02				0,3113	**0,3113**
user-03		0,2231			**0,2231**
user-04			0,3284		**0,3284**
user-05			0,3503		**0,3503**
user-06				0,2461	**0,2461**
user-07				0,2720	**0,2720**
user-08				0,2161	**0,2161**
user-09			0,4261		**0,4261**
user-14			0,3687		**0,3687**
user-15				0,2972	**0,2972**
user-16				0,2203	**0,2203**
user-17		0,3242			**0,3242**
user-18	0,3897				**0,3897**
user-19	0,3909				**0,3909**
user-20			0,3792		**0,3792**
user-21		0,3198			**0,3198**
user-22			0,3330		**0,3330**
user-24	0,4222				**0,4222**
user-25		0,2558			**0,2558**
user-26	0,6099				**0,6099**
user-27			0,3747		**0,3747**
user-28			0,3285		**0,3285**
user-30	0,4548				**0,4548**
user-31				0,1939	**0,1939**
user-32		0,3650			**0,3650**
user-33		0,3696			**0,3696**
user-34	0,3705				**0,3705**
user-35	0,4194				**0,4194**
user-36				0,2137	**0,2137**
user-37			0,3462		**0,3462**
user-38		0,3081			**0,3081**
user-39				0,3098	**0,3098**
user-41				0,2576	**0,2576**
user-46				0,2508	**0,2508**
user-50				0,2267	**0,2267**
user-51		0,2733			**0,2733**
user-52		0,3637			**0,3637**
user-53	0,4931				**0,4931**
user-54		0,3509			**0,3509**
user-55			0,4425		**0,4425**
user-57			0,4270		**0,4270**
user-58	0,5289				**0,5289**
user-60	0,5181				**0,5181**
user-64			0,4509		**0,4509**
user-66	0,3935				**0,3935**
user-68	0,4549				**0,4549**
user-69		0,3495			**0,3495**
Total Result	**0,4538**	**0,3148**	**0,3796**	**0,2513**	**0,3499**

Average - SPEEDk.avgU.TERxSPEED

USER	A	B	C	D	Total Result
user-01		6,869			**6,869**
user-02				4,803	**4,803**
user-03		6,913			**6,913**
user-04			5,397		**5,397**
user-05			5,429		**5,429**
user-06				4,823	**4,823**
user-07				4,488	**4,488**
user-08				4,145	**4,145**
user-09			6,084		**6,084**
user-14			6,039		**6,039**
user-15				5,079	**5,079**
user-16				4,227	**4,227**
user-17		6,295			**6,295**
user-18	8,419				**8,419**
user-19	8,360				**8,360**
user-20			3,453		**3,453**
user-21		8,934			**8,934**
user-22			4,307		**4,307**
user-24	8,463				**8,463**
user-25		6,830			**6,830**
user-26	6,730				**6,730**
user-27			5,411		**5,411**
user-28			3,782		**3,782**
user-30	8,450				**8,450**
user-31				5,018	**5,018**
user-32		6,144			**6,144**
user-33		7,394			**7,394**
user-34	9,545				**9,545**
user-35	8,009				**8,009**
user-36				3,402	**3,402**
user-37			5,923		**5,923**
user-38		5,540			**5,540**
user-39				4,291	**4,291**
user-41				5,378	**5,378**
user-46				4,543	**4,543**
user-50				3,796	**3,796**
user-51		5,956			**5,956**
user-52		8,278			**8,278**
user-53	9,970				**9,970**
user-54		7,878			**7,878**
user-55			5,531		**5,531**
user-57			5,100		**5,100**
user-58	13,711				**13,711**
user-60	6,750				**6,750**
user-64			2,633		**2,633**
user-66	6,753				**6,753**
user-68	14,144				**14,144**
user-69		6,261			**6,261**
Total Result	**9,109**	**6,941**	**4,924**	**4,499**	**6,368**

6.4.3 - User matrix of TER and Speed averages - table



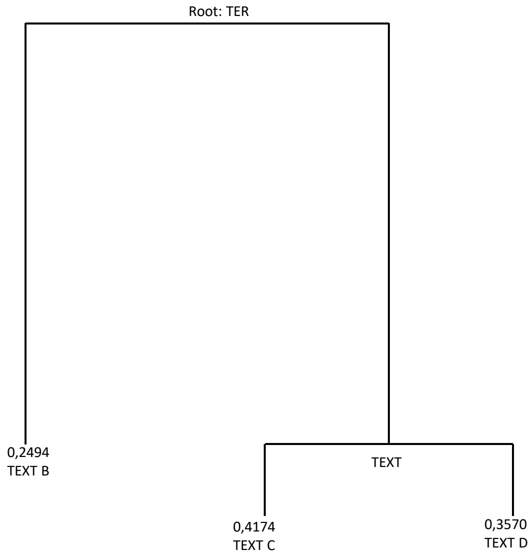
10-07-2017

```
>  summary(tree.TERP)

node), split, n, deviance, yval
      * denotes terminal node

1) root 1552 88,78 0,3422
  2) TEXT: B-Questionnaire 472 13,25 0,2494 *
  3) TEXT: C-Catalogue,D-Manual 1080 69,69 0,3827
    6) TEXT: C-Catalogue 460 27,10 0,4174 *
    7) TEXT: D-Manual 620 41,64 0,3570 *
```

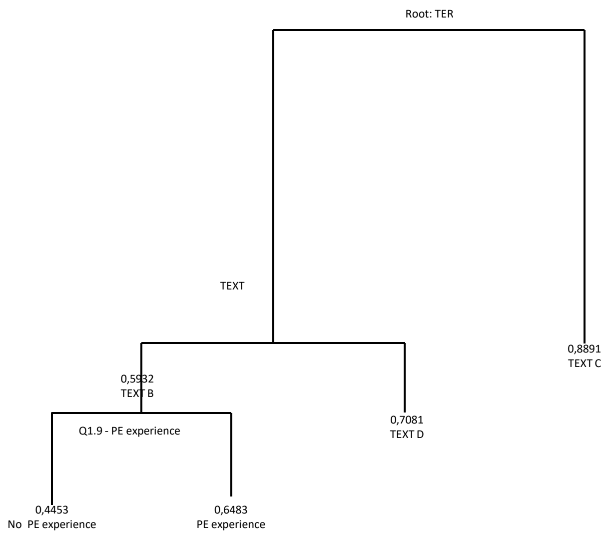
TER



10-07-2017

```
1) root 1552 308,20 0,7268
  2) TEXT: B-Questionnaire,D-Manual 1092 245,60 0,6584
    4) TEXT: B-Questionnaire 472 113,90 0,5932
      8) D.Experience.profile < 0.5 128  31,62 0,4453 *
      9) D.Experience.profile > 0.5 344  78,44 0,6483 *
    5) TEXT: D-Manual 620 128,20 0,7081 *
  3) TEXT: C-Catalogue 460  45,35 0,8891 *
```

TER - detail



10-07-2017

SPEED

```
node), split, n, deviance, yval
      * denotes terminal node

 1) root 1552 23540 6,188
   2) MODE: AC 779 14800 7,543
     4) D.Use.Predictive.writing < 0.5 413  7316 6,884
       8) D.taskPE < 0.5 255  3481 6,208 *
       9) D.taskPE > 0.5 158  3530 7,976 *
     5) D.Use.Predictive.writing > 0.5 366  7105 8,287
      10) D.Use.MT < 0.5 224  3671 7,300 *
      11) D.Use.MT > 0.5 142  2872 9,844 *
   3) MODE: PE 773  5862 4,823 *
```