

Alma Mater Studiorum – Università di Bologna

SCUOLA DI LINGUE E LETTERATURE,
TRADUZIONE E INTERPRETAZIONE

Sede di Forlì

Corso di Laurea Magistrale in Traduzione Specializzata (LM – 94)

TESI DI LAUREA

in Metodi e Tecnologie per la Traduzione

**Perceived difficulty and time effectiveness in translation:
a comparison of machine translation post-editing and translation memory use**

CANDIDATO

Alessandra Rossetti

RELATORE

Prof. Federico Gaspari

CORRELATORE

Prof.ssa Silvia Bernardini

Anno Accademico 2013/2014

Sessione II

TABLE OF CONTENTS

<i>RINGRAZIAMENTI</i>	7
RIASSUNTO.....	9
RESUMEN.....	10
ABSTRACT.....	11
LIST OF FIGURES.....	13
LIST OF TABLES.....	14
LIST OF APPENDICES.....	15
ABBREVIATIONS AND ACRONYMS.....	17
1. Introduction.....	19
1.1 Aim of This Study.....	19
1.2 Rationale behind This Study.....	19
1.3 Overview of the Contents.....	20
2. Related Work.....	23
2.1 Contents and Structure of the Chapter.....	23
2.2 Methods Adopted for Process-oriented Research.....	24
2.3 The Translation Process.....	28
2.4 The PE Process.....	32
3. TAP Experiments.....	39
3.1 Reasons for the Choice of TAPs within This Study.....	39
3.2 Set-up of the Experiment.....	40
3.2.1 Selection of the Sample.....	40
3.2.2 Avoidance of the Learning Effect.....	41
3.2.3 Translation Direction.....	42
3.2.4 Source Text Features.....	42
3.2.5 Information and Instructions.....	43
3.3 Experimental Sessions.....	45
3.3.1 Pilot Study.....	45
3.3.2 Main Experiments.....	45

4. Translation Solutions Provided by the TM Database and the Raw Output.....	49
4.1 Method Adopted for the Identification of Translation Problems and the Corresponding Translation Solutions.....	49
4.2 Data Analysis and Discussion.....	52
4.2.1 The CAT Setting.....	53
4.2.2 The PE Setting.....	56
4.2.3. A Comparison of the CAT and the PE Scenarios.....	59
4.3 Summary.....	62
5. Translation Strategies and the Level of Perceived Difficulty.....	65
5.1 Aim of this Part of the Research.....	65
5.2 Definition of “Translation Strategy”.....	65
5.3 Identification of the Number and Types of Translation Strategies.....	66
5.4 Assessment of the Level of Perceived Difficulty in Relation to Translation Strategies.....	68
5.4.1 Level of Perceived Difficulty within the CAT Setting: Data Analysis and Discussion.....	70
5.4.2 Level of Perceived Difficulty within the PE Setting: Data Analysis and Discussion.....	71
5.5 Summary.....	73
5.6 Analysis of Cognitive Effort and Relation with the Level of Perceived Difficulty.....	74
5.6.1 Aim of the Analysis of Cognitive Effort in the Present Study.....	74
5.6.2 Criteria for the Identification of Pauses to Be Analysed.....	75
5.6.3 Data Analysis and Discussion.....	76
5.6.4 Summary.....	79
6. Time Effectiveness.....	81
6.1 Aim of This Part of the Research and Method Adopted.....	81
6.2 Data Analysis and Discussion.....	82
6.3 Summary.....	84
6.4 Findings of the Present Work.....	86
6.4.1 Connecting Previous Results.....	86
6.4.2 Main Findings.....	87

7. Conclusions.....	89
7.1 Assessment and Limitations of the Research.....	89
7.1.1 Areas for Future Work.....	90
7.2 Lessons Learned.....	91
REFERENCES.....	93
APPENDICES.....	99
Appendix A: Information and Instructions Provided to the Participants Working Within the CAT Setting.....	99
Appendix B: Information and Instructions Provided to the Participants Working Within the PE Setting.....	101
Appendix C: Source Text to Be Translated with the Help of the TM Software by the Participant of the First Pair.....	103
Appendix D: Source Text to Be Translated with the Help of the TM Software by the Participant of the Second Pair.....	103
Appendix E: Source Text to Be Translated with the Help of the TM Software by the Participant of the Third Pair.....	104
Appendix F: Raw Output to Be Post-edited by the Participant of the First Pair.....	104
Appendix G: Raw Output to Be Post-edited by the Participant of the Second Pair..	105
Appendix H: Raw Output to Be Post-edited by the Participant of the Third Pair....	105
Appendix I: Translation Delivered by the Participant of the First Pair.....	106
Appendix J: Translation Delivered by the Participant of the Second Pair.....	106
Appendix K: Translation Delivered by the Participant of the Third Pair.....	107
Appendix L: Post-edited Text Delivered by the Participant of the First Pair.....	107
Appendix M: Post-edited Text Delivered by the Participant of the Second Pair.....	108
Appendix N: Post-edited Text Delivered by the Participant of the Third Pair.....	108
Appendix O: Rankings Obtained by the Participant of the First Pair Working within the CAT Setting.....	109
Appendix P: Rankings Obtained by the Participant of the Second Pair Working within the CAT Setting.....	109
Appendix Q: Rankings Obtained by the Participant of the Third Pair	

Working within the CAT Setting.....	109
Appendix R: Rankings Obtained by the Participant of the First Pair Working within the PE Setting.....	110
Appendix S: Rankings Obtained by the Participant of the Second Pair Working within the PE Setting.....	110
Appendix T: Rankings Obtained by the Participant of the Third Pair Working within the PE Setting.....	110

RINGRAZIAMENTI

I primi e sentitissimi ringraziamenti vanno al mio relatore, il Prof. Federico Gaspari, per la cura estrema con cui ha seguito questo lavoro dall'inizio alla fine, per la pazienza con cui ha sempre chiarito tutti i miei dubbi, per il sostegno costante senza il quale non avrei superato i momenti di incertezza, per l'atteggiamento incoraggiante che mi ha permesso di imparare dai miei innumerevoli errori, per tutti gli approfondimenti e gli spunti di riflessione che mi hanno fatta appassionare a quello che scrivevo, e per il tempo illimitato dedicato a parlarmi e ascoltarmi.

Ringrazio moltissimo anche la mia correlatrice, la Prof.ssa Silvia Bernardini, per le preziose e puntuali indicazioni che mi hanno permesso di migliorare questo lavoro.

Ringrazio di cuore i sette colleghi che hanno accettato di partecipare al mio esperimento a metà del semestre, in un periodo in cui l'attività didattica era in pieno fermento e le scadenze si accumulavano. Grazie per l'entusiasmo straordinario e per l'impegno che ci avete messo. Senza il vostro aiuto, il contenuto di questa tesi sarebbe stato tutt'altro.

Ringrazio i compagni e le compagne di corso che hanno reso unico ogni giorno di questi due anni. In particolare, Cristina, Silvia, Veronica e Michela, per le scampagnate, i viaggi al mare e in montagna, i gruppi di "studio", lo spirito di squadra, le mille e una pausa chiacchiere-caffè e molto altro ancora. Un ringraziamento speciale va poi a Cristina, per essermi accanto nella buona e nella cattiva sorte, per le passeggiate rigeneranti, le serate cinema e le cene a base di gelato, capaci di illuminare le giornate più buie.

Infine, ringrazio Delia, la mia compagna di (dis)avventure da sei anni a questa parte, per avermi convinta a tornare da Granada e provare il test alla SSLMIT. Grazie per non aver mai smesso di credere in me.

RIASSUNTO

Con il presente studio si è inteso analizzare l'impatto dell'utilizzo di una memoria di traduzione (TM) e del post-editing (PE) di un output grezzo sul livello di difficoltà percepita e sul tempo necessario per ottenere un testo finale di alta qualità. L'esperimento ha coinvolto sei studenti, di madrelingua italiana, del corso di Laurea Magistrale in Traduzione Specializzata dell'Università di Bologna (Vicepresidenza di Forlì). I partecipanti sono stati divisi in tre coppie, a ognuna delle quali è stato assegnato un estratto di comunicato stampa in inglese. Per ogni coppia, ad un partecipante è stato chiesto di tradurre il testo in italiano usando la TM all'interno di SDL Trados Studio 2011. All'altro partecipante è stato chiesto di fare il PE completo in italiano dell'output grezzo ottenuto da Google Translate. Nei casi in cui la TM o l'output non contenevano traduzioni (corrette), i partecipanti avrebbero potuto consultare Internet. Ricorrendo ai Think-aloud Protocols (TAPs), è stato chiesto loro di riflettere a voce alta durante lo svolgimento dei compiti. È stato quindi possibile individuare i problemi traduttivi incontrati e i casi in cui la TM e l'output grezzo hanno fornito soluzioni corrette; inoltre, è stato possibile osservare le strategie traduttive impiegate, per poi chiedere ai partecipanti di indicarne la difficoltà attraverso interviste a posteriori. È stato anche misurato il tempo impiegato da ogni partecipante. I dati sulla difficoltà percepita e quelli sul tempo impiegato sono stati messi in relazione con il numero di soluzioni corrette rispettivamente fornito da TM e output grezzo. È stato osservato che usare la TM ha comportato un maggior risparmio di tempo e che, al contrario del PE, ha portato a una riduzione della difficoltà percepita. Il presente studio si propone di aiutare i futuri traduttori professionisti a scegliere strumenti tecnologici che gli permettano di risparmiare tempo e risorse.

RESUMEN

Este estudio trata del impacto del uso de una memoria de traducción y de la posesión de un texto meta generado por un sistema de traducción automática sobre el nivel de dificultad que percibieron los participantes y sobre el tiempo que necesitaron para realizar una traducción de calidad alta. Los participantes fueron seis estudiantes italianos del curso de Traducción Especializada de la Universidad de Bolonia (en Forlì), a los que se les dividió en tres parejas. Cada pareja recibió un comunicado de prensa en inglés: un participante tradujo el texto al italiano usando la memoria de traducción y SDL Trados Studio 2011, mientras que el otro poseyó al italiano el texto generado por Google Translate. Cuando la memoria de traducción o el texto meta de la TA no proporcionaban traducciones correctas, los participantes podían buscar en Internet. A través de los protocolos de pensamiento en voz alta (TAPs), todos los participantes reflejaron en voz alta durante los procesos de traducción y posesión. Este método permitió reconocer los problemas traductivos encontrados y los casos en los que la memoria de traducción y el texto meta de Google Translate proporcionaron soluciones correctas; además, fue posible clasificar las búsquedas en Internet agrupándolas en estrategias traductivas, para que luego los participantes las pusiesen en orden de dificultad a través de entrevistas retrospectivas. Se midió la duración de cada tarea. Los datos sobre la dificultad y el tiempo necesario se relacionaron con el número de traducciones correctas contenidas en las dos herramientas. Se observó que traducir empleando la memoria de traducción llevó a una disminución del nivel de dificultad y permitió ahorrar más tiempo. El objetivo de este estudio es el de ayudar a los futuros profesionales de la traducción a elegir tecnologías que les permitan ahorrar tiempo y recursos.

ABSTRACT

The present work analyses the respective impact of translation memory (TM) use and post-editing (PE) of raw machine translation (MT) output on the level of difficulty perceived by trainee translators and on the time which they needed in order to deliver a high-quality text. The experiment involved six Italian translation students, all enrolled in the MA Programme in Specialised Translation at the University of Bologna at Forlì. The participants were divided into three pairs and each pair was given an excerpt from an English press release. Within each pair, one subject translated the text assigned into Italian by using the TM software within SDL Trados Studio 2011; the other subject performed full PE of the raw MT output obtained by Google Translate. In those cases in which the TM software or the raw output did not contain correct translations, the participants could consult any webpage. By adopting the method of Think-aloud Protocols (TAPs), the subjects were instructed to think aloud during the performance of the tasks. This allowed the identification of translation problems and of those cases in which the TM and the raw output provided correct solutions; moreover, it was possible to identify the translation strategies adopted and ask the participants to rank them in terms of difficulty during retrospective interviews. The time taken by each participant was measured. The data on the difficulty and the duration of the tasks were related to the number of successful translation solutions respectively provided by the TM and the raw output. It was observed that, in contrast to PE, using the TM led to a decrease in the difficulty perceived by the participants, and that it allowed them to save more time. The aim of this work is to help prospective professional translators to use effectively the technological tools at their disposal.

LIST OF FIGURES

Figure 1: Webpage containing the excerpt assigned to the second pair (URL: http://goo.gl/4rjJE4).....	47
Figure 2: Variation in the number of solutions provided by the TM database.....	55
Figure 3: Variation in the number of solutions contained in the raw output.....	59
Figure 4: Number of successful translation solutions respectively provided by the TM database and the raw output out of the overall number of common translation problems identified.....	62
Figure 5: Duration of the translation and the PE tasks.....	82

LIST OF TABLES

Table 1.1: Successful solutions (not) provided by the TM database during the 1 st TAP session.....	53
Table 1.2: Successful solutions (not) provided by the TM database during the 2 nd TAP session.....	54
Table 1.3: Successful solutions (not) provided by the TM database during the 3 rd TAP session.....	54
Table 1.4: Successful solutions (not) contained in the raw output during the 1 st TAP session.....	57
Table 1.5: Successful solutions (not) contained in the raw output during the 2 nd TAP session.....	58
Table 1.6: Successful solutions (not) contained in the raw output during the 3 rd TAP session.....	58
Table 1.7: Comparison between the TM database and the raw output in terms of solutions respectively provided to the translation problems identified by both participants within the 1 st pair.....	60
Table 1.8: Comparison between the TM database and the raw output in terms of solutions respectively provided to the translation problems identified by both participants within the 2 nd pair.....	61
Table 1.9: Comparison between the TM database and the raw output in terms of solutions respectively provided to the translation problems identified by both participants within the 3 rd pair.....	61
Table 2.1: Relation between the number of successful solutions provided by the TM database, the most frequent strategies and their rankings.....	70
Table 2.2: Relation between the number of successful solutions provided by the raw output, the most frequent strategies and their rankings.....	72
Table 3.1: Strategies adopted for the solution of translation problems and related pauses (when present) resulting from cognitive effort.....	77
Table 3.2: Rankings of the four Internet search strategies most frequently employed by the 1 st pair of participants.....	78
Table 4.1: Relation between time gains and successful solutions within the CAT setting.....	83
Table 4.2: Relation between time losses and successful solutions within the PE setting.....	83

LIST OF APPENDICES

A:	Information and Instructions Provided to the Participants Working within the CAT Setting.....	99
B:	Information and Instructions Provided to the Participants Working within the PE Setting.....	101
C:	Source Text to Be Translated with the Help of the TM Software by the Participant of the First Pair.....	103
D:	Source Text to Be Translated with the Help of the TM Software by the Participant of the Second Pair.....	103
E:	Source Text to Be Translated with the Help of the TM Software by the Participant of the Third Pair.....	104
F:	Raw Output to Be Post-edited by the Participant of the First Pair.....	104
G:	Raw Output to Be Post-edited by the Participant of the Second Pair.....	105
H:	Raw Output to Be Post-edited by the Participant of the Third Pair.....	105
I:	Translation Delivered by the Participant of the First Pair.....	106
J:	Translation Delivered by the Participant of the Second Pair.....	106
K:	Translation Delivered by the Participant of the Third Pair.....	107
L:	Post-edited Text Delivered by the Participant of the First Pair.....	107
M:	Post-edited Text Delivered by the Participant of the Second Pair.....	108
N:	Post-edited Text Delivered by the Participant of the Third Pair.....	108
O:	Rankings Obtained by the Participant of the First Pair Working within the CAT Setting.....	109
P:	Rankings Obtained by the Participant of the Second Pair Working within the CAT Setting.....	109
Q:	Rankings Obtained by the Participant of the Third Pair Working within the CAT Setting.....	109
R:	Rankings Obtained by the Participant of the First Pair Working within the PE Setting.....	110

S:	Rankings Obtained by the Participant of the Second Pair Working within the PE Setting.....	110
T:	Rankings Obtained by the Participant of the Third Pair Working within the PE Setting.....	110

ABBREVIATIONS AND ACRONYMS

APR	Average Pause Ratio
CAT	Computer-assisted Translation
CE	Confidence Estimation
CL	Controlled Language
CNA	Choice Network Analysis
CP	Collaborative Protocol
EWR	Event to Word Ratio
L1	First Language
L2	Second Language
LTM	Long-term Memory
MA	Master of Arts
MT	Machine Translation
OP	Orientation Pause
PE	Post-editing
RP	Revision Pause
SMT	Statistical Machine Translation
STM	Short-term Memory
TAP	Think-aloud Protocol
TA	Thinking Aloud
TM	Translation Memory

Chapter 1

INTRODUCTION

1.1 Aim of This Study

This study deals with a comparison between translation memory (TM) use and full post-editing (PE) of raw machine translation (MT) output in order to determine which of these two working scenarios can reduce the difficulty perceived by trainee translators to a larger extent, and which one turns out to be the most time-effective option. This work started with the analysis of the usefulness of the TM software and the raw output in terms of the successful translation solutions respectively provided by each of the two tools. Subsequently, the data obtained were related to the participants' perceptions of difficulty and to the time which they needed in order to complete the translation and the PE tasks on excerpts of three English press releases. Therefore, by combining think-aloud protocols (TAPs) and retrospective interviews, the aim of this study is to clarify the impact that these different translation technologies (i.e. TM software and raw output) can have on the performance of trainee translators. This should lead translation students to make a more effective use of the technological tools at their disposal and, in turn, could help them during their future careers, especially when having to cope with time and resource constraints. As O'Brien and Moorkens (2014) point out, MT is steadily more used for dissemination purposes, with PE being on the increase in professional translation. As a result, both TMs and raw MT output might represent viable options to obtain high-quality final texts. Nonetheless, preferring one over another has implications on the performance of translators which should be taken into account.

1.2 Rationale behind This Study

So far, research in the fields of the translation and the PE process has focused on a variety of aspects, such as the time required by these two tasks, the cognitive load which they involve, the types of strategies which are adopted, etc. Some of these aspects are analysed also within the present study, such as the duration of the translation and the PE process and the types of strategies employed by the participants. However, further aspects are taken into account too, since it was felt that they needed a more in-depth investigation. First of all, as far as strategies are concerned, within the present work, they were classified so that the subjects could indicate the level of difficulty which they perceived when adopting them. To the best of my knowledge, for the language combination English-Italian and looking at the performance of

translation students, perceptions of difficulty associated with translation strategies had not been analysed yet. Nonetheless, this is an aspect worthy of attention, since (together with the time taken by a task and the expected level of quality) it is one of the main factors which can influence the subjects' preference for one working scenario over another in their future career.

Secondly, the relation between, on the one hand, the number of correct translation solutions provided either by the TM database or the raw output and, on the other hand, perceptions of difficulty and duration of the tasks, had not yet, to my knowledge, been explored in detail. Nevertheless, this relation is particularly interesting since it can shed light on the extent to which the performance of trainee translators is influenced by the very features of the translation technologies being used. In turn, this allows the observation of whether these features alone can account for differences in terms of difficulty and duration of the tasks between the computer-assisted translation (CAT) and the PE setting. Therefore, one of the motivations for this study lies in the need to gain a better understanding of the factors which can influence the translation and the PE process of trainee translators.

1.3 Overview of the Contents

After this introduction, which clarifies the motivations behind this work, its aim and its structure, the following six chapters deal with specific parts of the research. More precisely, Chapter 2 provides a literature review in which publications focusing on the translation and the PE process are analysed. The works presented mainly deal with the translation strategies employed by the participants both during the translation and the PE process, the duration of the translation and the PE tasks and the cognitive load involved in them; moreover, further studies dealing with methodological issues are included. Chapter 3 focuses on the rationale behind the adoption of TAPs as the primary method for the gathering of data within the present work. Furthermore, it contains a description of the set-up of the experiments and of the sessions which were held. Chapter 4 reports the data regarding the respective usefulness of the TM software and the raw MT output in terms of successful translation solutions provided to translation problems (which comprise the problems encountered both during the translation and the PE process by the six trainee translators involved in the experiment). Chapter 5 deals with the translation strategies adopted by the subjects when the two tools proved to be useless, as well as with the level of perceived difficulty assigned to those strategies.

In order to do so, results from retrospective interviews during which the participants had to rank translation strategies in terms of perceived difficulty are reported and discussed.

Moreover, the data on the level of perceived difficulty are related to the findings regarding the number of successful translation solutions respectively provided by each of the two tools, in order to determine whether a relation between these two aspects can be identified. Furthermore, a sub-section of Chapter 5 analyses more detailed data for the first pair of participants, by combining the data on the level of perceived difficulty with evidence of cognitive effort emerging at unconscious level through pauses. The aim of this part of the research is to determine whether, by using a mixed approach, it is possible to gain a more complete picture of the overall difficulty experienced by the participants when employing translation strategies.

Chapter 6 focuses on the analysis of the time taken by the subjects when performing the translation and the PE tasks, so as to determine which between the CAT and the PE setting represented the most time-effective working scenario. As in the case of the data on the difficulty of translation strategies, in this Chapter too, the data on the different durations of the tasks are related to the number of successful solutions provided by the TM software and the raw output, in order to determine whether this latter aspect can influence the former. Moreover, this Chapter summarises the main findings of this entire work: it combines the results of the various parts of the research in order to determine whether and how the number of successful translation solutions respectively provided by the TM software and the raw MT output can be related to the level of perceived difficulty and the time required by the translation and the PE tasks. This is followed by a discussion on which of the two working scenarios can reduce the difficulty perceived and the time needed by trainee translators to a larger extent. Finally, Chapter 7 draws some general conclusions on this work, in particular by assessing its limitations and possible areas for future research.

Chapter 2

RELATED WORK

2.1 Contents and Structure of the Chapter

This chapter reviews some of the key publications dealing with, on the one hand, the methodology of process-oriented research and, on the other hand, aspects of the translation and the PE process. Due to the large amount of works addressing these topics, a selection was made with regard to the studies to be presented and, in the vast majority of cases, the material included refers to research which has been carried out in the last two decades or so. Sub-section 2.2 presents an overview of the methodological issues referring to the data gathering techniques adopted within the fields of the translation and the PE process, by mainly focusing on TAP-based research; sub-section 2.3 provides an overview of studies dealing with the translation process, while sub-section 2.4 introduces works focusing on the PE process. It is worth noting that research dealing with both the translation and the PE process has been conducted by using different methods for gathering data, such as TAPs, collaborative protocols (CPs), key-logging programs, etc. Therefore, in order to avoid an excessively fragmented literature review, the works presented within each sub-section are not classified according to the methods adopted, although methods are reviewed and commented. The contributions are presented in chronological order within each sub-section.

2.2 Methods Adopted for Process-oriented Research

Over the years, process-oriented research has constantly evolved both in terms of the methodology adopted and the objects of study. With regard to the former, thinking aloud (TA), namely the verbalisation of mental processes while performing a task, was used as the primary research method in order to shed light on the translator's "black box" through the analysis of TAPs, i.e. the written transcripts of the utterances recorded during the experiments. It is worth noting that the reporting which takes place during these experimental sessions is concurrent (i.e. simultaneous with the task being carried out) and indirect (i.e. the subjects are not instructed to be analytical about their thoughts) (Jääskeläinen, 2011). Since TAPs are in the verbal, and not in the numerical form, TAP-based research can be regarded as qualitative research (Sun, 2011). Moreover, similarly to other types of qualitative research and unlike quantitative research, TAP studies involve relatively small samples, which are selected so as to analyse information-rich cases and derive in-depth understanding rather than

empirical generalisations.

With regard to this latter point, Krings (2001) explains that a further reason for the small number of subjects involved in TAP experiments is that the effort required by the transcriptions of the verbalisations makes it almost impossible for a single researcher to analyse a large sample in an acceptable amount of time. Nonetheless, while one of the most important rules of qualitative research is natural situation, TAP experiments might involve a form of manipulation. For instance, the experimenter may want to determine whether a given tool can influence the subjects' performance and, accordingly, the subjects might be told which tools to use. With regard to this, Jääskeläinen (2011) states that validity (i.e. the extent to which the situation under analysis is manipulated and subjected to experimental control) is an issue with TA. Accordingly, the setting up of a TAP experiment has to aim at creating the least artificial environmental conditions, while at the same time controlling both subjects and task variables as closely as possible.

TAPs were firstly adopted in the field of psychology by Claparède (1932) and Duncker (1935). Subsequently, they were used by the American psychologists Ericsson and Simon (1993 (1984)), whose study provided the theoretical and methodological framework for cognitive explorations of the translation and the PE process. This framework is based on the view of human cognition as information processing and on the assumption that subjects are able to report accurately only on consciously processed information, namely the information which is at the focus of their conscious attention. In particular, according to the two American psychologists, subjects can report static and conscious mental states (i.e. the input and the output of the processing system), but not the dynamic processes themselves. Therefore, the data verbalised refer to steps of information processing, not to the process itself. Moreover, Ericsson and Simon (1993 (1984)) point out that information can be stored in several memories, which are:

- a. ultra-short-term memory, which does not keep information for more than about two seconds;
- b. short-term memory (STM), with limited capacity (about seven unconnected items) and intermediate duration;
- c. long-term memory (LTM), with large capacity and relatively permanent storage, but with longer access times with respect to the other memories.

Only information in STM can be reported, since that is the information being at the focus of the subjects' conscious attention. Therefore, the impossibility to verbalise information present in ultra-short-term memory and LMT is regarded as a limitation of TAPs,

which can lead to the incompleteness of data. Furthermore, several other weaknesses can be identified, such as a slow down effect (Krings, 2001 and Jakobsen, 2003). In addition, it has been observed that, if the task being carried out is very demanding, all available cognitive processes might be employed for the performance, thus leaving the subjects unable to verbalise (cf. sub-section 5.6). Moreover, some scholars have shown concern that TA could interfere with the main task being performed. With regard to this, it is worth noting that Ericsson and Simon (1993 (1984)) distinguish between three types of verbalisations which can be obtained through TAPs, namely Level 1, Level 2 and Level 3 verbalisations.

Level 1 verbalisations are the direct vocalisation of information in the form in which it is needed; Level 2 verbalisations involve description, explication and encoding of the thought content: the subject is required to label the information which is processed. Finally, Level 3 verbalisations require the explanation of thoughts, ideas and motives in relation to earlier thoughts and information, as well as additional interpretative processes. In this latter case, the subject is not expected to merely recode the information present in his/her STM, but he/she also has to link information to unrelated thoughts, for example by answering specific questions which are not directly linked to the task performed. According to Ericsson and Simon (1993 (1984)), unlike Level 3 verbalisations, Level 1 and Level 2 verbalisations do not interfere with the performance of the main task. Nonetheless, Toury (1991) suspects that, even with Level 1 and Level 2 verbalisations, TA could interfere with translating, since the need to verbalise leads the subjects to produce a spoken translation of a linguistic element before the written one, and it might occur that spoken and written translation do not involve the same strategies.

The adoption of TAPs in translation studies began in Europe in the late 1980s with the aim of complementing the normative models which had been applied to the study of the translation process up to that moment. On the contrary, as far as PE is concerned, Krings (2001) points out that attention began to be paid to the mental processes involved in this activity at a later time. This new interest was the result of the widespread use of translation technologies, which called for a comparison between MT PE and CAT tools, in particular TMs. Furthermore, the decision to adopt TAPs within the fields of translation and PE was encouraged by the fact that the two tasks mainly operate in verbally encoded data, which, according to Lörcher (1992), represents a favourable precondition for TA. Moreover, solving translation and PE problems is often carried out as a series of steps, and this step-by-step nature facilitates the gathering of information on the on-going mental processes. However, it should also be noted that, in addition to the analysis of the more general mental processes

involved in the translation and the PE task, TAPs have also been used in order to shed light on more specific aspects within these two processes, such as problem-solving strategies (Krings, 1986 and Lörcher, 1991), criteria for decision-making (Tirkkonen-Condit, 1989) and creativity (Kußmaul, 1991), among others. Furthermore, there has been a line of TAP-based research dealing with aspects of methodology; here, some of these works dealing with methodological issues are presented and reviewed.

Jääskeläinen (2000) deals with some examples of the methodological questions on which research adopting TAPs should focus. In particular, she discusses three issues, i.e. pre-experimental testing of subjects, a comparison between TA and CPs and, finally, the potential effects of TA on the performance of the translation task. With regard to the pre-experimental testing of subjects, she points out that, in addition to mapping the subjects' training and work experience, it might be useful to gather information about some of their personality traits which might influence the TAP experiment (in particular, tolerance of stressful situations and fear of failure). As far as the comparison between TA and CPs is concerned, Jääskeläinen (2000) points out that CPs do not provide access to the actual translation process, since it is mainly solitary. Accordingly, by adopting CPs, it is not possible to observe the thoughts routinely occurring to a single mind at work. Finally, with regard to possible interference from TA, Jääskeläinen (2000) compares eight translations produced by subjects while thinking aloud with eight translations of the same text produced by other subjects without thinking aloud. In particular, she analyses the influence of the syntactic structure of the source text on the target text, as well as the formal correspondence between them at the lexical level.

In his study on the cost and effort involved in PE, as opposed to those involved in human translation, Krings (2001) shows that TAPs slow down the primary task by roughly 30 per cent. He also observes that text production occurs in smaller steps and/or in a less linear way when the subjects are asked to think aloud. In particular, he draws his conclusions from the observation of students post-editing instructional texts which had been translated from English into both German and French by two rule-based MT engines. A further study aiming at analysing the effects of TA on the main task being performed (in particular, on translation speed, revision and segmentation) is the one conducted by Jakobsen (2003) with four Danish Master of Arts (MA) translation students and five Danish expert translators. Each of them translated four texts (two from Danish into English, and two from English into Danish); the two Danish texts were respectively 367 and 522 characters long (including spaces), while the two English texts contained 760 and 1001 characters respectively (including spaces).

In each language direction, one task was performed while thinking aloud and one

without thinking aloud. All verbalisations were recorded on audiotape, while keystrokes were logged with Translog. Subjects took part in warm-up sessions, so as to become familiar with both TA and Translog, and they were allowed to access Internet information; no time limit was set. It should be noted that the level of similarity between the two texts assigned for each language direction is not specified. Nonetheless, this is an important aspect to take into consideration. As a matter of fact, the four tasks were performed by each subject in random order; therefore, the translation speed of a subject might have been influenced by the learning effect (and not only by TA), if he/she translated two (possibly) similar texts one after the other.

Li (2004) summarises important safeguards which should be taken into account when conducting TAP research. Here only some of his methodological points are included, namely those which are more relevant to my work. More precisely, the safeguards presented are: voluntary participation and guarantee of anonymity; purposeful sampling (generalisations should be avoided and findings should be treated as descriptive of a given context; in order for findings to be transferable to other contexts, the researcher should seek subjects who are representative of different groups); triangulation (different data sources, methods and investigators should be involved); prolonged engagement (studies should be longitudinal in order to overcome problems of distortion due to biases and perceptions of either the researcher or the subjects) and (near-)natural situation (the research context should be kept as close to the subjects' routine working environment as possible and any external interference should be kept to a minimum).

Finally, O'Brien (2010) is devoted to some methodological challenges related to both eye tracking and translation process research in general. In particular, she divides these challenges into categories which, in some cases (such as natural situation and importance of anonymity) coincide with the safeguards identified by Li (2004). Here only those categories are reported which are deemed to be important so as to clarify further methodological aspects of my study. First of all, as far as validity is concerned, O'Brien (2010) identifies additional factors which might influence it. As a matter of fact, she writes that, although short texts should be used so as not to make the participants feel tired or bored, in order to investigate a real scenario, entire texts should be provided, since translators do not normally work with chunks of texts. Moreover, O'Brien (2010) points out that, in order to ensure validity, subjects should be free to use the resources which they routinely employ.

It is worth noting that the methodological shortcomings of TA identified within studies dealing with methodological issues led researchers to employ other methods, often in

combination which each other and/or with TA. These further methods include, among others, retrospective interviews (participants are asked questions on their actions after the performance of a task), CPs (subjects are asked to translate jointly, either in pairs or small groups), video recording (with the camera directed either at the translator or at the computer screen), keystroke logging (mainly by means of Translog, a programme which records every keystroke and cursor movement made by the subjects, as well as their eye movements), screen recording (which allows the recording of any screen activity carried out on a computer) and eye tracking (used to record fixations, gaze paths, pupil size and eye movements). As a matter of fact, although TAPs and other forms of verbal reports are far from being exhausted as a source of information, the advent of objective recording methods of gathering information about the translation and the PE process has supported verbal data and allowed the triangulation of research findings, thus improving their reliability. As a result, numerous studies have adopted triangulation, such as the one carried out by Faber and Hjort-Pedersen (2009), who use TAPs, retrospective interviews and Translog, or the study conducted by Angelone (2010), who resorts to both TAPs and screen recording.

2.3 The Translation Process

This sub-section deals with some of the studies carried out within the field of translation process research. The studies which were selected are characterised by a certain degree of variability in terms of the issues addresses; therefore they can provide a broader (although not complete) picture of translation process research. Muñoz Martín (2010) states that three levels can be identified within the more general notion of “translation process”: the first level is characterised by the mental states and operations involved in translation; the second level comprises the tasks resulting from those mental states and operations (e.g. reading, typing, research); the third level includes everything and every agent engaged in the period of time which goes from the contact by the commissioner to the delivery of the final product. With regard to the objects of study, it is important to note that translation process research has focused on a variety of aspects, such as decision criteria (Tirkkonen-Condit, 1989), problem-solving strategies (Krings, 1986 and Lörscher, 1991), subject profiling (Muñoz Martín, 2010), effort in translation (Alves et al., 2012), etc. Here some of these works are reviewed.

First of all, it is worth mentioning the line of translation process research which focused on translation strategies. Numerous classifications of translation strategies (mainly on the basis of TAPs) have been proposed so far. To give some examples, Krings (1986)

suggests that translation strategies can be classified as strategies of comprehension, equivalent retrieval, equivalent monitoring, decision-making and reduction; Gerloff (1986) proposes a further classification in which the strategies identified are: problem identification, linguistic analysis, storage and retrieval, general search and selection, text inferencing and reasoning, text contextualisation and task monitoring; Malone (1988) identifies nine strategies which can be adopted by the translators, namely equation, substitution, divergence, convergence, amplification, reduction, diffusion, condensation and reordering; Jääskeläinen (1993) makes a distinction between global and local translation strategies, the former being related to the text in its entirety, whereas the latter to specific units within it; Chesterman (1997) distinguishes between comprehension strategies and production strategies.

A further work dealing with strategies is the one conducted by Asadi and Séguinot (2005), who carried out a study with nine professional translators working at the linguistics department of a pharmaceutical company. They assigned the subjects a text which belonged to their area of expertise and asked them to translate for 20 minutes as they normally did. More precisely, two participants translated from French into English, while the others from English into French. Combining screen recording, TAPs and retrospective interviews, the researchers aimed at determining whether experience-related or text-specific shortcuts, strategies or general patterns shared by the translators within this homogeneous group could be identified. However, it should be noted that the use of a screen video recorder, although being invisible to the subjects, may have led them to abandon some of the strategies routinely employed (for example, because they were not always successful).

Moreover, in Asadi and Séguinot's (2005) study, after the performance of the task, the subjects were asked questions about their decisions, pauses and editing. This approach compels the subjects to retrieve detailed information from their LTM; therefore, the possibility cannot be ruled out that they are reinterpreting their thoughts or creating them anew. Furthermore, within this experiment, two language directions were analysed, but it is not specified which was the subjects' mother tongue, despite the fact that language competence is an aspect that should be taken into account when adopting TAPs. As a matter of fact, it has been shown that, when subjects translate or post-edit out of their mother tongue, i.e. out of the same language in which they are thinking aloud, interferences from TA are greater (Göpferich, 2009).

Translation process research also focused on pauses, mainly with the aim of assessing the cognitive load involved in translation. In order to analyse this latter aspect, most scholars have resorted to the conclusions drawn by Ericsson and Simon (1993 (1984)) about the

meaning of pauses within TAPs (although the experiments of the two American psychologists were not conducted within the fields of translation or PE). In particular, according to the two scholars, in case of competition between the verbalisations and the task-oriented processes (i.e. with high cognitive load), the task-oriented processes will have priority over the verbalisations. This is mainly proven by the fact that “[s]ubjects tend to stop verbalizing in conditions where they are giving indications of being under a high cognitive load.” (Ericsson and Simon, 1993 (1984): 91). The studies on pauses which have been conducted so far are characterised by remarkable variability with regard to the criteria according to which pauses are analysed (e.g. their meaning, their minimum length, their distribution within a text, etc.), as well as with regard to the method through which they are recorded. For example, Séguinot (1989) defines pauses as interruptions during typing; however she does not mention how pauses are measured, nor their minimum length of time. Jakobsen (1998) analyses pauses during the translation process by using Translog and points out that a time unit of one second can be adequate for the identification of all the delays in the process of text production which would be interesting for his study. According to Alves (2006), pauses are indicators of moments of intense monitoring or problem-solving and decision-making processes.

Pauses are an object of study also in Alves and Liparini Campos’s (2009) study, which investigates the impact of both TMs and time pressure on the types of support used by professional translators. The subjects involved were 12 professional translators with the same years of experience and all familiar with the use of TM software. The language pairs analysed were English-Brazilian Portuguese and German-Brazilian Portuguese. Each of the eight source texts was taken from a different technical manual and contained approximately 500 words. The source texts, characterised by the same level of difficulty, were divided into four groups of two texts each, one for each language direction. The first two texts were translated without the aid of a TM and without time pressure; for other two texts the subjects used a TM and worked without time pressure; two more texts were translated with time pressure but without the help of a TM and, finally, the remaining two texts were translated with both a TM and time pressure. With regard to the methods employed, it should be noted that Alves and Liparini Campos (2009) adopted observation charts, Translog, Camtasia (a screen recorder) and retrospective interviews, which allowed the triangulation of the data gathered.

Alves and Liparini Campos (2009) began their analysis by classifying pauses either as orientation pauses (OP) or revision pauses (RP) (Jakobsen, 2005). However, in those cases in which a pause was longer than a given time criterion, it proved to be difficult to determine whether the focus of a subject’s mental activity was on the translation of a new segment or the

revision of a previously translated one. On these occasions, in order to classify pauses, the two researchers took into account the actions immediately following pauses. Nevertheless, this approach is questionable, since, during a pause, the focus of a subject's attention may easily shift from orientation to revision, and vice versa, as well as to thoughts not related to the task being performed. Therefore, adopting TAPs would have provided more data on what was going on in the translators' minds during moments of silence. Following the classification of pauses in OP and RP, each pause was also classified according to the type of support used during it so as to solve a translation problem. In particular, the categories identified were: simple internal support, when the subject did not perform any kind of search during a pause; simple external support, when a subject consulted a resource which played a vital role in the solution of a problem; dominant internal support, when a solution was not the result of any consultation (even though it is performed); dominant external support, when a suggestion provided by an aid was adopted by a subject.

Finally, Enríquez Raído (2014) carried out research into the use of the Web as an external aid for the translation process by analysing the Web search behaviours of six participants: four postgraduate translation trainees in their first year of studies, a PhD student of translation with three years of professional experience in different domains and a translation teacher with over 15 years of experience in the discipline. With regard to the four students, two of the subjects were native speakers of English, one was native speaker of Russian and one considered English as his/her first language (L1) and Mandarin as his/her second language (L2). The subjects were asked to translate two popular-science texts belonging to different domains from Spanish into English; the first text was translated by all six participants, while the second one only by the four students. In order to gather data, Enríquez Raído (2014) resorted to different methods: screen recording, background questionnaires (used to assess the subjects' knowledge and experience with both translation and web searching), online search reports (in which the subjects described the web search tasks which they performed so as to solve translation problems) and one-to-one interviews. It should be noted that the subjects involved in the study represented a very heterogeneous group from which no general conclusions can be drawn. As a matter of fact, they differed not only in terms of language skills, but also with regard to the levels of expertise in translation and web searching skills.

2.4 The PE Process

This sub-section is concerned with research into the PE process, and it charts its goals and its methodologies. The most extensive analysis of the PE process, conducted by using TAPs, dates back to Krings (2001), who researched the translation strategies employed during the PE process, starting from the assumption that the differences between human translation and PE influence the translation problems which the subjects encounter during the translation and the PE process respectively and, as a result, the number and types of translation strategies which they adopt in order to overcome them. Krings's (2001) study was particularly important also because he identified three levels of PE effort, i.e. temporal, technical and cognitive effort. Temporal effort refers to the time required in order to post-edit a given output; technical effort consists of the keystrokes and cut-and-paste operations needed so as to produce a post-edited version; finally, cognitive effort refers to the mental processes aimed at identifying and correcting the errors within the raw output. However, while it is possible to measure temporal and technical effort directly, cognitive effort can be assessed only in an indirect way.

Most of the subsequent works dealing with the PE process adopted Krings's (2001) classification of PE effort, even when resorting to different data gathering techniques. O'Brien (2005), starting from the drawbacks involved in the use of TAPs (sub-section 2.2), assessed the potential of two other methods to measure the effort involved in PE, i.e. keystroke logging (through Translog) and Choice Network Analysis (CNA), a method aimed at measuring the difficulty experienced by the post-editors by taking into account the number of choices made by the subjects. As a matter of fact, when a source text item is translated in the same way by all the participants, it is assumed that the item requires minimal processing and is easily translatable; on the contrary, when a number of different translations is provided for the same source text item, it is assumed that, due to the range of possible translations, greater cognitive effort is required on the part of the subjects in order to choose among them.

More precisely, O'Brien (2005) presented an experiment in which data regarding cognitive effort (based on the pauses recorded through Translog) were triangulated with data obtained thanks to CNA. However, it should be noted that both methods have drawbacks. First of all, there is no certainty that pauses recorded through Translog are actually indicators of cognitive effort (e.g. they may be due to lack of concentration). Accordingly, using TA in conjunction with Translog might provide more data regarding what is going on in the post-editor's mind during pauses, even when the verbalisations are incomplete. Secondly, with regard to CNA, when post-editors modify the output by adopting different solutions for the

same source text item, this may be due either to the different resources consulted (which are not specified in O'Brien, 2005) or to the different solutions occurring to the subjects' minds, rather than to the difficulty of the very source text item.

Based on the assumption that pauses indicate cognitive effort, O'Brien (2006) conducted a study with the aim of identifying a relation between source text machine translatability and PE effort (measured on the basis of pauses). In particular, the purpose of her work is to determine whether post-editing sentences to which controlled language (CL) rules are applied requires less cognitive effort on the part of the subjects than post-editing sentences to which CL rules are not applied. She used two types of English sentences, one type edited using CL rules, while another type without the application of such rules. Subsequently, the two types of sentences were machine-translated into German and nine subjects were asked to post-edit the raw output. It is worth noting that, in this study too, O'Brien (2006) resorted to Translog and CNA as methods for gathering data. Since she did not find a relation between pauses (and, as a result, between cognitive effort) and the expected difficulty of a segment (determined by the adoption of CL rules), she argues that, one possible explanation for this is the fact that, taken in isolation, pauses are not reliable indicators of cognitive effort in PE.

The second possible explanation indicated by O'Brien (2006) is the fact that CL rules do not affect the cognitive effort required on the part of subjects. Nevertheless, she does not mention two further aspects which might have influenced the results obtained, namely the subjects' mother tongue and their degree of familiarity with the PE activity. As a matter of fact, it might be assumed that, if some subjects post-edited out of their L1 while others into it, this may have required more effort (with, possibly, more pauses) on the part of the subjects post-editing out of their L1 as a result of the translation direction, regardless of the very features of the source text. Furthermore, if the subjects were not familiar with the PE activity, this may have resulted in pauses caused by uncertainty about the very scenario within which they were working.

Temnikova (2010) carried out a study aimed at assessing how much cognitive effort is required on the part of post-editors when correcting the different kinds of MT errors; in particular, Temnikova (2010) tried to determine whether the use of a controlled language for crisis management (CLCM) improved the machine translatability of texts, thus diminishing the cognitive effort on the part of post-editors. In order to overcome the limits of both TAPs and CNA, she proposed a new evaluation approach which starts from and enriches the error classification of the MT output presented in (Vilar et al., 2006). More precisely, Temnikova

(2010) assumed that the less cognitively costly errors are those at word level, while the most cognitively expensive ones are those involving syntactic and semantic processing of the whole sentence. She combined this MT error classification with manual evaluation of the human cognitive effort required. Nevertheless, she does not mention the aspects that human evaluators had to take into account when assessing the difficulty connected to the detection and correction of errors.

Tatsumi and Roturier (2010) focused on the relation between source text characteristics and temporal and technical PE effort; in particular, they aimed at determining whether characteristics such as ambiguity, complexity and style guide compliance influence PE effort. For their experiment, nine subjects with different levels of familiarity with the PE process were selected; more precisely, seven of them had experience in post-editing IT-related documentation, one in post-editing non-IT-related documentation, and one had no experience in post-editing. Therefore, it remains to be seen whether, by solely involving more specific groups of subjects (e.g. professionals with the same years of experience in post-editing IT-related documentation), the results obtained in terms of the time required or the changes made to the output would be different. Furthermore, it is worth noting that, since Tatsumi and Roturier (2010) performed their study with one language pair (i.e. from English into Japanese), it would be interesting to extend their work to additional language pairs and, in turn, to different source text characteristics.

Specia (2011) used three different annotation types (i.e. PE time, PE distance and PE effort scores) in order to experiment with confidence estimation (CE) models, used to filter low-quality segments which would require more effort on the part of post-editors than translating from scratch. She compared CE models respectively based on PE time, PE distance and PE effort scores so as to determine which one(s) can facilitate the work of the subjects to a larger extent in a practical working scenario. However, it is worth noting that the CE model based on PE effort scores was obtained by asking translators to indicate the amount of PE required by a given segment (e.g. complete or little PE). Therefore, there is no guarantee that the amount of PE which was needed correlated well with cognitive effort. To give just one example, it can be assumed that, even when little PE was needed, the correction of the few errors identified may have required more cognitive effort on the part of the subjects than the complete retranslation of a sentence for which an appropriate translation easily occurred to their minds.

A similar problem related to the assessment of PE cognitive effort can be observed in the study conducted by Koponen (2012), namely an experiment aimed at studying the relation

between technical PE effort and perceived cognitive PE effort, by analysing cases in which the edit distance and a manual score reflecting perceived cognitive effort differed. More precisely, the manual score indicating the perceived cognitive effort was obtained using a 5-point scale based on the percentage of text to be post-edited: “(1) indicates the segment is incomprehensible and needs to be translated from scratch; (2) significant editing is required (50-70%) of the output; (3) about 25-50% of the output needs to be edited; (4) about 10-25% needs to be edited, and (5) little to no editing is required” (Koponen, 2012: 183). Therefore, in this case too, a problem can be identified with regard to the relation between the percentage of text to be post-edited and the cognitive effort required, since no data are gathered on the types of errors identified.

Koponen et al. (2012) suggest that PE time might be used so as to assess the cognitive effort involved in PE. In order to do so, effort indicators resulting from the performance of the post-editors involved were logged using PET, a freely available PE tool (Aziz et al., 2012). In order to relate the error types recorded by means of PET with the level of cognitive effort, the classification proposed by Temnikova (2010) was adopted, although with some modification. Within the same study, a second experiment was conducted which aimed at analysing individual differences among post-editors. The results suggest that: 1) even with the same instructions to minimally change the output, subjects differ with regard to what constitutes minimal; 2) some post-editors maximise cut-paste operations while others prefer writing out the whole corrected segment; 3) some post-editors proceed sequentially, while others move around in the sentence.

Lacruz et al. (2012) is a study which starts from O'Brien's (2006) work. The researchers introduced a new metric for pause activity, namely the average pause ratio (APR), which is computed for each segment as the average time per pause in the segment divided by the average time per word in the segment. Moreover, Lacruz et al. (2012) assumed that cognitive effort could be measured by counting the number of complete editing events (i.e. collections of individual editing actions which can be considered to naturally form part of the same overall action). Within their study, the only participant involved was told to post-edit the text to his/her satisfaction; therefore, the number of complete editing events recorded (and the pauses associated with them) might have been the result of the preferences of the subject, rather than the very features of the raw output.

Lacruz and Shreve (2014) related pause activity in keystroke log reports with the event to word ratio (EWR), which is calculated by dividing the number of complete editing events by the number of words. More precisely, Lacruz and Shreve (2014) investigated the

correlation between cognitive effort (as indexed by EWR) and various pause metrics, by using Translog so as to record keystrokes and pauses during the PE process. EWR was assumed to be an indirect index of cognitive effort; as a matter of fact, although it succeeded in capturing density of effort, it could not distinguish between the different levels of cognitive effort required. Therefore, it is not always a reliable indicator of cognitive effort, since some complete editing events might require more (or less) cognitive effort than others, depending on the types of errors identified.

Popović et al. (2014) investigated five types of PE operations and their relation with both cognitive PE effort (indicated as “quality level”) and PE time. Classification of edit operations was performed automatically, while human quality level scores were used as a measure of cognitive effort. PE time was measured on the sentence level in a controlled way, so as to isolate factors such as pauses between sentences. Unlike previous studies, within this work, data regarding the cognitive effort required were obtained by asking annotators to assess the quality level of segments. However, it is not specified whether annotators were familiar with the PE process, despite the fact that this factor may have influenced the expectations of the subjects in terms of the quality of the MT output and, accordingly, their subsequent evaluation. Furthermore, before post-editing, the participants were asked to perform the minimum number of edits necessary to make the translation acceptable; however, since acceptability is a relative concept, this guideline is likely to have influenced the number and type of changes which were made.

Finally, it is worth noting that, both within the field of the translation and the PE process, there is a line of research dealing with the impact of translation technologies on translation and PE speed. To give just a few examples, Guerberof (2009) researched the difference in terms of productivity (measured by means of processing speed) between the PE of MT output and the processing of fuzzy matches from a TM database, finding that professional translators show higher productivity when post-editing MT output. On the other hand, Koehn (2009) investigated the impact of the use of a CAT tool based on SMT methods on the translation speed of non-professional translators, and he found that users work faster than when unassisted by this aid. A further study dealing with the influence of a MT enhanced CAT tool on translation speed is the one conducted by Federico et al. (2012), whose results demonstrated that, in the majority of cases, professional translators save time when passing from the TM to the TM+MT suggestion mode. Gaspari et al. (2014) focused on two bidirectional language pairs (i.e. English-German and English-Dutch) in order to research the difference between, on the one hand, the subjects’ perceptions of the time required by the PE

task as opposed to the time required by manual translation and, on the other hand, the actual PE time. The results showed that, although the participants perceive MT PE to be slower, it proves to be the faster option for two translation directions (i.e. from Dutch into English and from English into German).

Chapter 3

TAP EXPERIMENTS

3.1 Reasons for the Choice of TAPs within This Study

As far as this study is concerned, TAPs were used as the primary method in order to gather data on the translation problems identified by the subjects (cf. sub-section 4.1), on the correctness of the translation solutions respectively provided by the TM database and the raw output (cf. sub-section 4.1), on the number and type of translation strategies employed by the participants (cf. sub-section 5.3) and on the pauses indicating cognitive effort (cf. sub-section 5.6). The primary rationale behind the choice to use TAPs was the need to create the least invasive environmental conditions and make the subjects work within a routine situation. This need led to the exclusion of keystroke logging, screen recording and eye tracking as methods of data gathering. As far as keystroke logging is concerned, although keystroke loggers are not intrusive, they have fewer functions compared with Microsoft Word or CAT tools. Accordingly, their use would have forced the subjects to work with a piece of software different from the one which they routinely use when translating or post-editing, thus possibly causing a slow down effect or uncertainty on the part of the subjects due to the use of the tool employed, rather than to the very features of the task being performed.

Screen recording and eye tracking were ruled out too, although for different reasons: Hansen (2008) points out that, when resorting to eye tracking or screen recording, there is a factor of stress that should be taken into account, since the subjects might feel intimidated by the idea of being observed during every stage of their performance. As a result, adopting eye tracking or screen recording might have led the participants to alter their behaviour (e.g. by avoiding consulting a low-quality online dictionary or spending a long amount of time considering a single problem). It should also be noted that the method of CAN, adopted by O'Brien (2005), was ruled out too, since the subjects were told that they could consult any webpage; therefore, a large amount of different solutions could not have been regarded as an indicator of difficulty, but rather as the result of different problem-solving behaviours. Furthermore, it is worth pointing out that TAPs were deemed to be more suitable than other types of verbal reports. In particular, CPs were excluded because, when observing two minds at work, thoughts may be recorded which would have never occurred to a single subject (Jääskeläinen, 2000).

Nonetheless, for one part of the present research (namely the one aiming at assessing

the level of difficulty perceived by the participants in relation to translation strategies), TAPs were combined with retrospective interviews (cf. sub-section 5.6). Excluding objective recording methods (such as eye tracking, keystroke logging, etc.) and relying on TAPs as the primary means of gathering data had a main drawback, namely the impossibility to know whether the participants actually verbalised all the information needed. Therefore, the possibility could not be ruled out that the data which were gathered were incomplete. Nonetheless, it should be noted that resorting to TAPs had various merits. More precisely, this method allowed the participants to work within routine conditions and without feeling under pressure. Therefore, this ensured that the data which were gathered (although possibly incomplete) mirrored the actual translation and PE process of the trainee translators involved in the experiment.

3.2 Set-up of the Experiment

3.2.1 Selection of the Sample

The present study was conducted with six trainee translators all enrolled in the final year of the MA Programme in Specialised Translation at the University of Bologna at Forlì (Italy). Apart from the greater availability of the subjects, this choice was mainly due to two reasons. First of all, the students who accepted to participate in the study could be matched by translation and language skills in English; as a matter of fact, since they had received the same training in translation between English and Italian, it could be assumed that, by the time this experiment was conducted, they had reached a similar level of proficiency. Moreover, in the previous year, all the subjects had attended the same lessons on the use of CAT tools (in particular SDL Trados Studio 2011) and on MT PE, thus becoming familiar with both working scenarios. Therefore, unlike several previous studies (e.g. Tatsumi and Roturier 2010), the subjects selected represented a homogeneous group. However, it is worth noting that involving participants who belonged to the same context had a drawback, namely the fact that it was not possible to employ the safeguard described by Li (2004) as “purposeful sampling”, whose aim is to apply the results obtained to a context as wider as possible.

The second motivation to involve trainee translators was the fact that, if the participants had been professional translators, it would have been more problematic to match them by translation and language skills, since it would have been necessary to take into account a variety of aspects, such as their training, their years of work experience, their specialisations, etc. (Jääskeläinen, 2000). Secondly, it may have been more difficult for

professionals to verbalise their thoughts during the performance of the tasks assigned. As a matter of fact, subjects seem to stop verbalising not only when they are deep in thought (i.e. when coping with a heavy cognitive load¹), but also when they have to do little thinking, namely when problem solving becomes routine (as in the case of professionals). This process, known as “automation”, is characterised by a reduction in the amount of processing carried out in STM, and is explained as follows:

Before overlearning has occurred, processes have to be interpreted, with substantial feedback from intermediate processing stages in STM. Overlearning amounts to compiling these processes, so that fewer tests are performed when they are being executed, hence less information is stored at intermediate stages in STM.

(Ericsson and Simon, 1993 (1984): 126-127)

Therefore, as subjects can produce verbalisations only on thoughts which are to some extent conscious (Jääskeläinen, 2000), conducting the study with trainee translators (i.e. non-professionals) allowed the recording of a large number of verbalisations about intermediate processes of which the subjects were aware, since they had not automated them yet.

3.2.2 Avoidance of the Learning Effect

It is necessary not to underestimate the process of automation which may take place during the course of a task, even though it is performed by a non-professional. With regards to this, Ericsson and Simon (1993 (1984): 127) point out that

We may distinguish between automatic processes that subjects already possessed prior to an experiment, as part of their cognitive skills, and processes whose intermediate stages became more automatic, and hence less reportable, during the course of the experiment. In the case of the latter, reports from the automated processes at the end of the experiment will omit information about intermediate states of which the subjects were aware at the beginning of the experiment.

For this reason, each of the participants was administered solely one experimental task (i.e. either translation or PE) on just one text, so as to have as little a learning effect (and, as a

¹ It is worth noting that, within the present study, a distinction is made between “cognitive load” and “cognitive effort”, on the basis of the distinction between “cognitive demand” and “cognitive effort” made by Lacruz and Shreve (2014) and by Lacruz et al. (2014). In particular, “cognitive load” is assumed to indicate the degree of challenge required by a given task, while “cognitive effort” corresponds to the overall amount of mental resources which a subject has to deploy while carrying out a task.

result, as little automation) as possible. As a matter of fact, it is important to keep in mind that the three texts used within this experiment (available in appendices C-E) were very similar in terms of content and formal features, with some variability being present only in the last paragraph of each of them. Therefore, the learning effect might have resulted not only from carrying out the two different tasks on the same text, but also from performing one task on more than one text. Furthermore, it should be noted that the paragraphs constituting each of the texts were characterised by a difference in complexity, in particular in terms of lexicon and syntactic structures; therefore, this led the subjects to employ a variety of problem-solving strategies during the course of the experiment.

3.2.3 Translation Direction

Each of the subjects (who were all native speakers of Italian) was asked to either translate or post-edit the text assigned from English into Italian. The rationale behind the choice of this translation direction was the need to ensure that the translation problems encountered (cf. subsection 4.1), the difficulty experienced (cf. Chapter 5) and the time taken (cf. Chapter 6) would not be the result of the specific difficulties related to translating or post-editing into a language other than their native language. Furthermore, with regard to the possible interferences of concurrent verbalisations with the performance of a task, the participants were asked to translate into their native language also because results from previous studies (Göpferich, 2009) demonstrated that, when subjects translate or post-edit into their mother tongue, i.e. the same language in which they are thinking aloud, interferences due to concurrent verbalisations are smaller than in studies where the language in which the subjects think aloud differs from the language of the target text.

3.2.4 Source Text Features

The English source texts which were selected for this study (available in appendices C-E) were three passages slightly longer than 100 words, and each of them had been taken from a press release containing data on the economic performance of a package delivery company. As far as the length of the three passages is concerned, it is worth noting that, although O'Brien (2010) points out that translators and post-editors rarely work with chunks of texts, brief passages from longer press releases were selected, so as to avoid a drop in motivation caused by the performance of a long task. Nonetheless, in order to provide the subjects with a source text as coherent and cohesive as possible, the chunks selected from each press release

were the first two or three paragraphs of the text and they appeared one after the other. These excerpts were selected for two main reasons. First of all, the press release is a text type which the subjects had already translated as part of their coursework, although not so frequently so as to become familiar with it; therefore, this choice was made to avoid routine tasks which would have caused a reduction in the amount of verbalisations (cf. sub-section 3.2.2).

Secondly, the texts selected represented realistic assignments for trainee translators and were potential candidates for translation with either TM software or MT PE. Although the three source texts were different, since each of them reported the data of a different quarter, they were very similar in terms of terminology, syntactic structures and stylistic features. These similarities were fundamental in retaining the same level of difficulty and thus in controlling this variable. However, the choice of three different texts was meant to have a larger amount of possible problems (which were all assumed to be characterised by the same level of complexity with respect to the subjects' skills) so as to broaden the range of problem-solving strategies that could be observed. As far as the level of difficulty of the texts is concerned, it is also important to underline that very demanding tasks might lead the subjects to reduce their verbalisations, as a result of excessive cognitive effort. To give just one example, when describing Matrat's TAP experiment, Jääskeläinen (2000) points out that her choice to use difficult texts resulted in a small amount of verbalisations. For this reason, the translation and the PE tasks assigned within the present study were set up by taking into account the subjects' translation and language skills, so as not to be very complex.

3.2.5 Information and Instructions

At the beginning of each experimental session, before the subjects started translating or post-editing, they signed a consent form in which it was specified that their participation in the experiment was voluntary, that their anonymity would be guaranteed and that they could ask any question to the experimenter. Furthermore, it was explained that the subjects could withdraw from the study at any stage without penalty. Therefore, the safeguards of voluntary participation and guarantee of anonymity (Li, 2004) were ensured so as to avoid distortion of data due, on the one hand, to lack of motivation and, on the other hand, to possible fear on the part of the participants of having their translations evaluated. The subjects were also given written instructions and pieces of information (available in appendices A and B). In the information section, the task to perform was explained in detail, together with the aim of the investigation, the resources which the subject could use and the main features of the text type that he/she would translate or post-edit. Furthermore, the participants were told that no time

limit was set and that their translations would not be evaluated. Although Jääskeläinen (2000) stresses the importance of collecting data on personality traits (and, in particular, fear of stressful situations) before conducting TAP experiments, within the present study, due to the information provided, it was assumed that there would be no need to gather this type of data, since no pressure was put on the participants.

As far as instructions are concerned, they were assumed to be very important since it has been demonstrated that the subjects' TAPs are influenced by their exact wording (Ericsson and Simon, 1993 (1984)). In particular, requiring Level 3 verbalisations (cf. sub-section 2.2) might have led the participants to alter the structure of their thought processes. Jääskeläinen (2000) too agrees that if the subjects feel that their ability to be analytical about their thoughts is under investigation, they might be intimidated. For this reason, within this study, the types of information which the subjects were asked to verbalise can be ascribed to either Level 1 or Level 2 verbalisations (cf. sub-section 2.2). As a matter of fact, the subjects were told that they could verbalise any thought occurring to their minds but, at the same time, they were asked to mainly focus on their actions, to express their considerations on the results found, etc. This decision was also meant to reduce the possible effects of TA on the translation (Jakobsen, 2003) and the PE process (Krings, 2001). As a matter of fact, it was assumed that making the subjects focus on more "external" actions (such as the webpages that they consulted) would leave them free to consider problems and possible solutions as much as they needed. Finally, within the instruction section, it was also explained that the participants had to deliver high-quality texts. This was meant to reduce biases due to individual differences.

More precisely, with regard to the translation task, the subjects were provided with information about two possible scenarios which could have occurred during the translation process and the types of thoughts on which they were expected to focus. In particular, it was explained that 100% or context matches may have been found by the TM that the subjects would keep, and in this case they were not required to verbalise any thought, although they were not prevented from doing it. In the scenario in which the TM did not provide any match or provided a fuzzy match whose translations had to be checked, the subjects were expected to search for possible translations by firstly using the Concordance Search, after verbalising the portion of text on which they were performing searches. If the Concordance Search facility proved to be useless, they were allowed to consult any website; in this case, they were asked to verbalise some specific types of information, such as the words which they were searching, the websites which they were consulting, the solutions contained in the webpages,

the quality of such solutions, the changes which they would make to them, etc. Furthermore, the participants using the TM database were instructed to update it as they translated.

As far as PE is concerned, the subjects were asked to turn on Word's track changes feature before starting post-editing. They were subsequently presented with two possible scenarios which may have occurred and were instructed on the types of information on which they should have focused in relation to them: in those cases in which the subjects considered the raw output to be correct, they were not required to verbalise, although they were not prevented from doing it. Instead, if they had some doubts about a translation or deemed it to be incorrect, they could consult any website and, in this case, they were expected to verbalise some specific types of information, such as the words which they were searching, the websites which they were consulting, the solutions contained in the webpages, the quality of such solutions, the changes which they would make to them, etc.

3.3 Experimental Sessions

3.3.1 Pilot Study

Prior to the main experiment, a pilot study was carried out. Like the experimental participants, the subject who participated in the pilot study was a trainee translator enrolled in the final year of the MA Programme in Specialised Translation at the University of Bologna at Forlì (Italy). Since the aim of the pilot study was to determine whether the instructions (their final version is available in appendices A and B) provided to the participants were clear in asking them to verbalise the types of information needed, and since the verbalisations required were very similar (regardless of whether the subjects were translating with the help of TM software or post-editing MT output) (cf. sub-section 3.2.5), it was assumed that the pilot study could have been conducted only within one of the two working scenarios considered. Accordingly, the subject performing the pilot study was asked to solely post-edit one of the three texts which would be used for the main experiment, while thinking aloud. On the basis of her suggestions, the instructions were slightly modified so as to make them more clear, in particular with regard to the types of information which would be more useful for the purpose of this study.

3.3.2 Main Experiments

Before conducting the main experiments, it would have been advisable to give each of the participants a warm-up task in order to help them familiarise with TA (O'Brien, 2009).

Although this was not possible due to the time constraints under which this research was conducted, it should be noted that the participants were not asked to verbalise every thought occurring to their minds (although they were not prevented from doing so), but rather focus on some specific actions and considerations (cf. sub-section 3.2.5); therefore, this was assumed not to require previous training. The students were divided into three pairs, and each pair was given one of the passages to translate from English into their native Italian, with the instruction of aiming for publishable quality of their target texts (available in appendices I-N). The members of each pair worked on the very same extract while thinking aloud, but under different conditions: one student used the TM software in SDL Trados Studio 2011, while the other post-edited the raw MT output provided by Google Translate. Although O'Brien (2010) points out that subjects should be free to use the resources which they routinely employ, within my experiment, the participants were told that they could consult any webpage, but they were not allowed to resort to tools different from those provided to them (i.e. TM software or raw output). This decision was meant to gather data on the respective impact of using a TM and post-editing raw output on the performance of trainee translators, also by focusing on the difficulty of the Internet searches which the participants had to perform when these two tools did not provide correct solutions (cf. Chapter 5).

Each of the subjects was provided with both a paper version of the text and an electronic one on his/her computer. The participants were made aware up-front that their verbalisations would be recorded (O'Brien, 2009) and a visible voice recorder was indeed used during the experimental sessions. This allowed the transcribing and the analysis of the verbalisations at a later stage. No objective methods for the gathering of data (such as keystroke logging or screen recording) were used, since it was assumed that they would lead the subjects to alter their normal behaviour as a result of the awareness of being constantly observed (cf. sub-section 3.1). Nonetheless, the safeguard of triangulation (Li, 2004) was employed by resorting to retrospective interviews after the performance of the tasks, when the participants were asked to rank their four most frequent translation strategies in terms of perceived difficulty (cf. sub-section 5.4). Each of the subjects could work at his/her normal pace and performed the task within his/her routine working environment; therefore, a (near-) natural situation was created (Li, 2004). Furthermore, no technical changes to the subjects' computers had to be made and subjects were not asked to work by using software unfamiliar to them. It is also important to note that the interaction between the researcher and the subjects was reduced to a minimum. As a matter of fact, if a subject had felt that he/she was taking part in social interaction, this could have led him/her to alter his/her verbalisations (e.g.

by using explanations or justifications for his/her actions or considerations) (Bernardini 2001). Accordingly, interaction consisted solely of reminders to verbalise when the subjects were silent for more than one minute. Data from each subject were collected in just one session of about an hour. As a result, although it would have been advisable, prolonged engagement (Li, 2004) was not be used.

Here, one of the three excerpts assigned to the participants is provided:

The screenshot shows a webpage with a dark teal header containing the text 'Investor Relations'. Below the header, the page title is 'Financial News Release'. The main headline is 'UPS Board Declares Dividend'. There is a link with a printer icon that says 'View printer-friendly version'. On the right side, there is a '<< Back' link. The main text of the release reads: 'ATLANTA, May 05, 2011 (BUSINESS WIRE) -- The UPS (NYSE:UPS) Board of Directors today declared a regular quarterly dividend of \$0.52 per share on all outstanding Class A and Class B shares. The dividend is payable June 1, 2011, to shareholders of record on May 16, 2011. Earlier this year, the UPS Board boosted the regular quarterly dividend by 11% to the current level of \$0.52 per share. UPS's dividend has more than tripled since 2000, when it stood at \$0.17 per share. The company has either increased or maintained its dividend every year for more than four decades.'

Figure 1: Webpage containing the excerpt assigned to the second pair (URL: <http://goo.gl/4rjJE4>)

Chapter 4

TRANSLATION SOLUTIONS PROVIDED BY THE TM DATABASE AND THE RAW OUTPUT

4.1 Method Adopted for the Identification of Translation Problems and the Corresponding Translation Solutions

In this chapter, the focus of interest is the respective ability of the TM software and the raw output to provide successful translation solutions to the translation problems encountered by the trainee translators involved in this experiment. Therefore, before analysing the respective usefulness of these two tools in terms of translation solutions, it was necessary to set some criteria in order to identify those cases in which the participants encountered translation problems, both within the texts and the translation technologies provided. In the instructions provided to them (available in appendices A and B), the participants were asked to verbalise all the portions of text on which they were performing searches, either because they had to decide between more than one way of rendering linguistic elements (Pym, 2010) or because they did not know what a correct translation would be. Therefore, thanks to the verbalisations, it was possible to identify the translation problems which the participants encountered within the texts assigned to them. Nonetheless, once the translation problems had been identified, for the purpose of this study (cf. sub-section 1.1), it was also necessary to assess whether the translation solutions respectively provided by the TM database and the raw output were deemed to be correct. Therefore, through the instructions (available in appendices A and B), the participants were also asked to verbalise their considerations on the correctness of the results contained in these two tools.

Nevertheless, in order to gather more complete data on the usefulness of the TM database and the raw output, other aspects emerging from TAPs were taken into account, namely further indicators of problems referring to the solutions contained in these two tools and suggesting that the participants either did not find any solution in the tool being used or were not satisfied with the results provided. As far as problem indicators are concerned, Krings (1986) distinguishes between primary and secondary problem indicators; Lörcher (1991), on the other hand, explains that indicators can be characterised by different degrees of vagueness; therefore, he makes a distinction between explicit indicators (such as the verbalisation of a problem) and implicit ones (e.g. a pause during the performance of a task). As far as implicit indicators are concerned, he points out that they are more vague and more

difficult to interpret than explicit ones; therefore, in order to be considered as reliable data, they should be analysed in relation to further pieces of additional information. For example, with regard to pauses, Lörscher (1991: 61-62) stresses that

[a] pause in target-language text production may indicate a translational problem, but also a momentary distraction, lack of concentration, etc. If, however, the pause occurs at a point in the text which the analyst interprets to be difficult to translate, and if further signs such as [...] phatic utterances such as “hm” or “oh” with a fall-rise intonation contour, drawls, repetitions, and/or self-corrections can be found in the environment of that point, all the signals in their entirety are most likely to indicate a translational problem of a subject.

Within the present study, since numerous phenomena emerged from the TAP experiments which were conducted, a selection had to be made with regard to which of them should have been treated as problem indicators referring to the solutions contained in the TM database and the raw output. This selection was influenced by classifications reported in previous studies (e.g. Krings, 1986 and Jääskeläinen, 1987). More precisely, it was observed whether the indicators proposed in those works could be identified also within the verbalisations of the participants involved in the present study. The list below contains those phenomena emerging from the TAP experiments which were deemed to be (either implicit or explicit) problem indicators referring to the solutions provided by the two tools:

- Explicit statement of a problem (e.g. “Non mi convince.”, “Ok, qua c’è qualche problema più importante.”, “C’è un problema di articoli.”)²
- Uncertainty markers³ (e.g. “Non so...”, “Credo che...”, “Vediamo se...”, etc.)⁴
- Use of aids (e.g. “Proviamo su IATE.”, “Per sicurezza mi conviene dare un’occhiata su Internet.”)⁵
- Fillers (e.g. “Ehm...”)

² Since the experiment was carried out in Italian, the transcripts of the sessions are in Italian. The English translation of the examples reported is provided: “I’m not sure.”, “Ok, there are some bigger problems here.”, “There is a problem with articles.” (my translation)

³ To the best of my knowledge, there are no previous studies about Italian expressions indicating uncertainty during the performance of TAP experiments. Therefore, the conclusions reported by Tirkkonen-Condit (2000: 127) were used as guidelines in order to identify Italian expressions indicating uncertainty: “uncertainty phenomena [...] contain expressions of epistemic and deontic modality, hedges on quality and quantity, questions, hypothetical statements, references to ignorance, uncertainty, etc.”

⁴ “I don’t know...”, “I think that...”, “Let’s see...” (my translation)

⁵ “Let’s try with IATE.”, “I’ll google it, just to be sure.” (my translation)

It is worth noting that “explicit statement of a problem” is a problem indicator corresponding to the subjects’ verbalisations on the correctness of the results respectively contained in the TM database and the raw output (as required through the instructions provided). The further elements in the list were used in order to identify those cases in which, despite not being satisfied with the solutions contained in the TM database or the raw output, the participants did not clearly state their considerations. Since no objective recording method (such as screen recording, keystroke logging, etc.) was used, further problem indicators were indeed paramount to gather more reliable data on those cases in which the two tools employed were (not) useful and, accordingly, the participants had to resort to Internet search strategies (cf. sub-section 5.3).

Here some excerpts of the transcriptions of the participants’ verbalisations are provided in order to clarify the extent to which the analysis of different problem indicators allowed to identify all those cases in which the subjects either did not find or were not satisfied with the results contained in the TM database or the raw output:

1: “Allora, nel primo segmento non ho risultati dalla memoria. Quindi sul Concordance cerco ‘boosts’. Non lo trovo neanche sul Concordance. Quindi apro il dizionario Wordreference per vedere se ci sono delle alternative che mi piacciono. ‘Boost’ mi dà ‘sostenere’, ‘incoraggiare’, ma qui credo sia un titolo, quindi ci vuole qualcosa che sia ad effetto...”⁶

2: “Cerco ‘regular quarterly dividend’. Lo cerco su Eur-Lex. Però non trovo niente, quindi vado su Linguee. Non trovo granché neanche qui, quindi cerco su Google se quello che ha messo Google è corretto. Sì, penso che sia corretto perché ci sono tantissime soluzioni.”⁷

3: “‘Projection’ sarà ‘proiezione’? ‘Previsione’, mmm.”⁸

Each of these three excerpts contains a different problem indicator suggesting that the participant was not satisfied with the tools being used and, accordingly, had to resort to a translation strategy. In particular, in the first example, there is the explicit statement of a problem, since the subject clearly stated that the TM database was useless in that case. In the second example, it can be noted that the participant did not say that he/she was not satisfied with the translation contained in the raw output. Nonetheless, the very fact that he/she resorted to an aid indicates that he/she did not regard the solution provided as convincing and,

⁶ “Well, for the first segment, I have no matches from the TM. So I am using the Concordance to look up ‘boosts’. I can’t find it in the Concordance neither. So I am consulting Wordreference, to see whether I can find translations which I like. I’ve found ‘sostenere’, ‘incoraggiare’, but I think this is a heading, so I need something more impactful”. (my translation)

⁷ “I’m looking up ‘regular quarterly dividend’ on Eurlex. I can’t find anything. So I am consulting Linguee, but I can’t find anything here neither. So I am searching on Google to see whether the translation from Google is correct. Yes, I think it’s correct, because there are many solutions.” (my translation)

⁸ “Does ‘projection’ mean ‘proiezione’? ‘Previsione’ uhm.” (my translation)

accordingly, felt the need to perform some searches. Finally, in the third example, the participant's uncertainty about the translation found is indicated by the presence of a filler within his/her verbalisations.

4.2 Data Analysis and Discussion

As far as the analysis of data is concerned, it is worth pointing out that the differences between the two settings were taken into account. With regard to the CAT scenario, those cases in which resorting to the TM database either did not provide any result or provided an unsuccessful one were listed together, since in both cases the TM software proved to be useless (see Tables 1.1-1.3). Accordingly, only those translations found in the TM and employed by the subjects (even when Internet searches were performed in order to check them) were considered as successful translation solutions. With regard to the PE setting, the translations contained in the raw output were treated as successful when they either were left as such or the changes made to them did not alter the root of the translated word(s) contained in the output (even when Internet searches were performed in order to check them), but rather other elements, such as the position of a word within a sentence, its grammatical categories, etc. On the contrary, when the translation provided by the output was replaced, the result was regarded as being rejected by the subject (see Tables 1.4-1.6).

Finally, in order to adequately interpret these data, it is important to remember that the work of the second and the third subject using the TM software was influenced by the updates which were stored in the TM after the previous session(s). On the contrary, with regard to the PE scenario, although the output was uploaded into Google Translator Toolkit each time a session was concluded (so as to allow data storage), no change could be observed in the texts which were subsequently obtained from Google Translate. As a result, post-editors did not benefit from the work of the post-editors performing the same tasks before them. On the basis of these differences, the analysis was divided into three parts. In the first two parts (cf. sub-sections 4.2.1 and 4.2.2), the focus is on the number of (un)successful solutions respectively provided by each of the two tools for the translation problems identified by the subjects within each of the two settings. On the contrary, the third part of the analysis (cf. sub-section 4.2.3) deals with a direct comparison between the two scenarios in terms of translation solutions provided to the participants for the common translation problems.

4.2.1 The CAT Setting

Since the participants working within the CAT setting had been instructed to firstly resort to the Concordance Search when they either did not know what a correct translation would be or were uncertain about it (cf. sub-section 3.2.5), the usefulness of the TM software was primarily analysed by looking at the number of successful solutions which this tool was able to provide via the Concordance Search. Subsequently, the impact of matches was also discussed. Tables 1.1-1.3 report the data regarding the three TAP sessions. Each of them refers to the performance of one of the three subjects working within the CAT scenario. In the first column, all the translation problems identified by the subjects are reported. The second and the third columns respectively indicate those cases in which the solution contained in the TM was deemed to be correct and those when it was either not found or not used once it had been found.

TRANSLATION PROBLEMS	RESULTS FROM THE CONCORDANCE SEARCH	
	Found and accepted	Not found/ not accepted
Boosts		X
Board		X
Earnings Outlook		X
Strong Cash flow		X
NYSE	X	
Board of Directors		X
Regular		X
\$		X
Outstanding		X
Class A		X
Class B		X
Payable		X
Of record		X
Operating environment	X	

Table 1.1: Successful solutions (not) provided by the TM database during the 1st TAP session

TRANSLATION PROBLEMS	RESULTS FROM THE CONCORDANCE SEARCH	
	Found and accepted	Not found/ not accepted
Declares		X
Board of Directors	X	
Regular quarterly dividend	X	
Payable		X
Shareholders of record	X	
Boosted	X	

Table 1.2: Successful solutions (not) provided by the TM database during the 2nd TAP session

TRANSLATION PROBLEMS	RESULTS FROM THE CONCORDANCE SEARCH	
	Found and accepted	Not found/ not accepted
Board	X	
Boosts	X	
Per Share	X	
Directors	X	
Earnings Outlook	X	
We believe	X	
distributions	X	
Chairman	X	

Table 1.3: Successful solutions (not) provided by the TM database during the 3rd TAP session

The data which were gathered from all three TAP sessions within the CAT scenario show that:

- the number of translation problems for which the solutions provided by the TM database were accepted by the subjects steadily increased. In the first TAP session, only two out of fourteen results (approximately 14%) were deemed to be correct; in the second TAP session, four out of six results (approximately 66%) were accepted; finally, in the third TAP session, eight out of eight solutions (100%) were inserted by the subject in his/her final product;
- the number of translation problems for which the TM database did not provide any (successful) solution steadily decreased. In the first TAP session, the subject was not helped by the TM find a solution in twelve out of fourteen cases (approximately 85%); in the second TAP session, this happened for two out of six translation problems (approximately 33%) and finally, in the third TAP session, the subject was able to find a solution to all eight translation problems that he/she had encountered.

These results are visible in Figure 2.

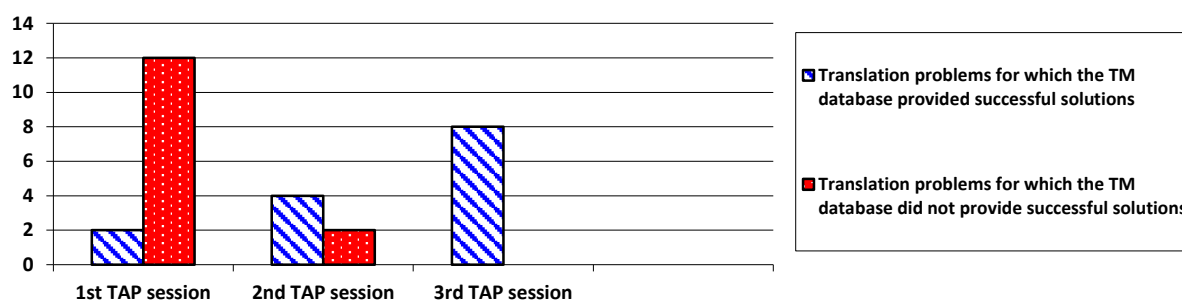


Figure 2: Variation in the number of solutions provided by the TM database

In addition to the growing number of correct solutions provided by the Concordance Search, a further way in which the TM software facilitated the subjects' performance is represented by matches. With regard to their occurrence, a difference can be identified among the three sessions: in the first one, no matches were found; in the second one, an 83% fuzzy match was inserted by the TM database and the subject performed searches about two of the translations proposed in the match (i.e. those for "payable" and "shareholders of record"). Next, when searching the solution for a further translation problem by using the Concordance Search, the subject found another fuzzy match which had not been automatically inserted, since it was a 70% fuzzy match;⁹; this match proved to be very useful since the participant inserted it in

⁹ The fuzzy match threshold which was set was 75%, in order to increase the usefulness of the matches which were automatically inserted in the target text. As a matter of fact, O'Brien and Moorkens (2014) state that there

his/her final text in a straightforward way, without performing additional searches. As far as the third TAP session is concerned, four matches (a 100%, an 84%, a 96% and a 92% match) were automatically inserted by the TM in the target text. In this case too, the subject did not perform any search for any of the translation solutions contained in the matches: he/she just changed the percentages and the dates within them, therefore it could be assumed that these matches facilitated his/her work to a large extent.

4.2.2 The PE Setting

The participants working within the PE scenario had been instructed to verbalise the translation problems which they encountered within the raw output while post-editing (cf. sub-section 3.2.5). Tables 1.4-1.6 report the translation problems in the first column; while the second and the third column respectively indicate those cases in which the solution contained in the raw output was deemed to be correct by the subject and those cases in which it was deemed to be correct, although partly modified. Finally, the fourth column reports those translations contained in the output which were rejected by the subjects, since they were considered as incorrect.

is wide agreement that fuzzy matches below a specific value (e.g. 75% similarity) are useless and, accordingly, need full human translation.

TRANSLATION PROBLEMS	RESULTS FROM THE OUTPUT		
	Accepted	Accepted and modified	Not accepted
UPS Boosts Dividend		X	
Board			X
Cites			X
Earnings Outlook			X
Strong Cash Flow			X
(NYSE: UPS)		X	
Board of Directors	X		
Regular quarterly dividend	X		
Outstanding	X		
Class A shares		X	
Dividend is payable		X	
Shareholders of record			X
Turned in a great performance			X
Global operating environment	X		
Volatile	X		
UPS Chairman and CEO		X	
That projection			X

Table 1.4: Successful solutions (not) contained in the raw output during the 1st TAP session

TRANSLATION PROBLEMS	RESULTS FROM THE OUTPUT		
	Accepted	Accepted and modified	Not accepted
Board	X		
Declares	X		
Regular quarterly dividend	X		
Outstanding			X
Dividend is payable	X		
Shareholders of record	X		

Table 1.5: Successful solutions (not) contained in the raw output during the 2nd TAP session

TRANSLATION PROBLEMS	RESULTS FROM THE OUTPUT		
	Accepted	Accepted and modified	Not accepted
Earnings Outlook			X
Quarterly	X		
Outstanding shares		X	
Class A			X

Table 1.6: Successful solutions (not) contained in the raw output during the 3rd TAP session

The data which were gathered from each of the three TAP sessions within the PE scenario show that:

- in the first TAP session considered, ten out of seventeen results (approximately 58%) were deemed to be correct; in the second TAP session, five out of six results (approximately 83%) were accepted by the subject; finally, in the third TAP session, two out of four solutions to the translation problems identified by the subject (namely 50%) were deemed to be successful solutions;

- with regard to the number of unsuccessful translations contained in the raw output, it can be observed that, in the first TAP session, the subject was not helped by the output find a solution in seven out of seventeen cases (approximately 41%). In the second TAP session, this happened for one out of six translation problems (approximately 16%) and finally, in the third TAP session, the subject rejected two out of four translation solutions (50%) to problems, thus assuming that they were not correct.

These data are summarised in Figure 3.

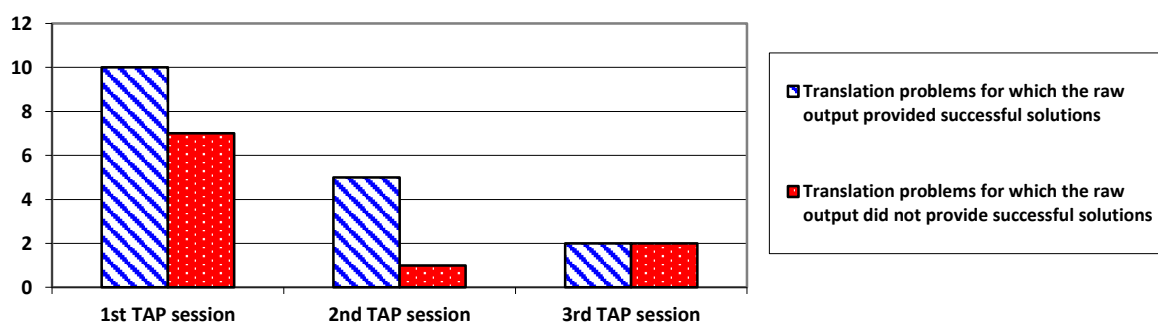


Figure 3: Variation in the number of solutions contained in the raw output

4.2.3 A Comparison of the CAT and the PE Scenarios

Tables 1.7-1.9 report the translation problems encountered by both participants within each of the three pairs, and show which of the two tools between TM software and raw output proved to be more useful in solving problems. This was possible because the members of each pair worked on the same text (cf. sub-section 3.3.2); therefore, each Table corresponds to the results obtained from the performance of two subjects working with different tools on the same text. The first column reports the translation problems encountered, while the second and the third column respectively show those cases in which the results in the TM were regarded as correct (and, accordingly, employed) and those in which they were either incorrect or not found. The remaining three columns respectively indicate those cases in which the results in the raw output were accepted as such, accepted although partly modified

or rejected.

TRANSLATION PROBLEMS	RESULTS FROM THE CONCORDANCE SEARCH		RESULTS FROM THE OUTPUT		
	Found and accepted	Not found/ not accepted	Accepted	Accepted and modified	Not accepted
Boosts		X		X	
Board		X			X
Earnings Outlook		X			X
Strong Cash flow		X			X
NYSE	X			X	
Board of Directors		X	X		
Regular		X	X		
Outstanding		X	X		
Class A		X		X	
Payable		X	X		
Of record		X			X
Operating environment	X		X		

Table 1.7: Comparison between the TM database and the raw output in terms of solutions respectively provided to the translation problems identified by both participants within the 1st pair

TRANSLATION PROBLEMS	RESULTS FROM THE CONCORDANCE SEARCH		RESULTS FROM THE OUTPUT		
	Found and accepted	Not found/ not accepted	Accepted	Accepted and modified	Not accepted
Declares		X	X		
Regular quarterly dividend	X		X		
Payable		X	X		
Shareholders of record	X		X		

Table 1.8: Comparison between the TM database and the raw output in terms of solutions respectively provided to the translation problems identified by both participants within the 2nd pair

TRANSLATION PROBLEMS	RESULTS FROM THE CONCORDANCE SEARCH		RESULTS FROM THE OUTPUT		
	Found and accepted	Not found/ not accepted	Accepted	Accepted and modified	Not accepted
Strong earnings Outlook	X				X

Table 1.9: Comparison between the TM database and the raw output in terms of solutions respectively provided to the translation problems identified by both participants within the 3rd pair

The data which were gathered show that:

- during the first TAP session, two out of twelve results (approximately 16%) were accepted by the subject using the TM software, while eight out of twelve solutions (approximately 66%) were accepted by the subject working within the PE scenario. During the second TAP session, the subject working within the CAT setting accepted two out of four solutions (50%) from the TM database, while the subject working within the PE scenario regarded as successful the solutions contained in the output in four out of four cases (100%). During the third TAP session, one out of one (100%) solutions were accepted by the subject using the TM, while zero out of one results (0%) were accepted by the post-editor;
- during the first TAP session, ten out of twelve results (approximately 83%) provided by the TM were not accepted or not found by the subject, while four out of twelve solutions contained in the output (approximately 33%) were not accepted by the subject working within the PE scenario. During the second TAP session, the subject working within the CAT setting either did not accept or did not find two out of four solutions (50%), while the subject working within the PE scenario deemed as incorrect zero out of four results (0%) in the raw output. During the third TAP session, zero out of one (0%) solutions were either not accepted or not found by the subject using the TM, while one out of one results (100%) from the output were not accepted by the post-editor.

Figure 4 summarises these results, by showing the difference between the two tools in terms of the respective successful solutions provided to the same translation problems.

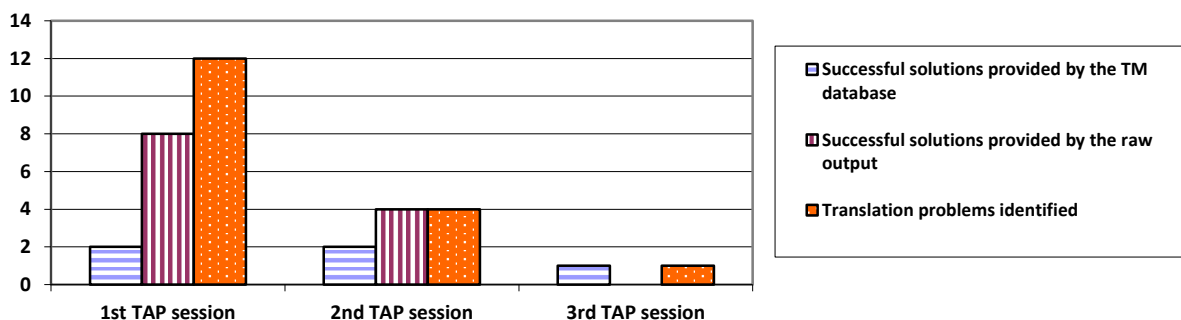


Figure 4: Number of successful translation solutions respectively provided by the TM database and the raw output out of the overall number of common translation problems identified

4.3 Summary

By looking at the data presented within this Chapter, some conclusions can be drawn. First of all, it should be noted that, although the number of successful translation solutions provided by the TM database steadily increased as a result of the updates, in two out of three cases the

percentage of successful solutions contained in the TM was lower than the corresponding percentage contained in the raw output. This result is confirmed by the data regarding the percentages of correct solutions respectively provided by the two tools to the common problems identified by both participants within each pair. As a matter of fact, in this case too, for two out of three pairs of participants, it was observed that the raw output contained a larger number of successful solutions than the TM database.

Secondly, although this study focuses on the impact of the number of translation solutions on the perceived difficulty and the duration of the tasks, here a further aspect is worth briefly mentioning, namely the number of translation problems respectively identified within the CAT and the PE scenarios. In particular, when looking at the number of translation problems, what can be observed is the fact that, for the first pair of participants, it is larger within the PE setting; for the second pair, it is the same within both working scenarios; while, for the third pair of participants, it is larger within the CAT setting. This latter data is particularly surprising when considering the increased number of matches provided by the TM database to the third subject working within the CAT setting.

To sum up, it can be argued that, in the majority of cases, post-editing raw MT output is a more effective option when compared to TM use in terms of the correct translation solutions which the subjects are provided by the tool being used. Nonetheless, it should be noted that a larger number of correct translation solutions does not imply neither a lower level of difficulty nor a lower amount of time. To give just one example, for those problems to which the raw output did not provide successful solutions, the difficulty and the amount of time required in order to solve them might be greater than the difficulty and the amount of time required in order to solve the translation problems identified within the CAT scenario, even when they outnumbered those within the PE scenario. For this reason, the data on the usefulness of the two tools in terms of solutions were related to the difficulty perceived by the participants and the time needed to complete the tasks so as to assess the respective impact of TM use and PE of raw output on the performance of trainee translators.

Chapter 5

TRANSLATION STRATEGIES AND THE LEVEL OF PERCEIVED DIFFICULTY

5.1 Aim of This Part of the Research

The aim of this part of the research is to assess the level of difficulty that the participants perceived when working within the two working scenarios considered, namely the CAT and the PE settings. More precisely, the focus of interest is respective impact of TM use and PE of raw output on the difficulty of the Internet searches performed in order to solve the translation problems for which the two tools either did not provide successful solutions or presented translations which the subjects had to check (cf. sub-section 4.1). After identifying the number of Internet searches carried out (via TAPs) and assigning each of them to a type of translation strategy (cf. sub-section 5.3), the data on the difficulty perceived when employing translation strategies were obtained by means of retrospective interviews (cf. sub-section 5.4). Furthermore, following the assessment of the level of perceived difficulty in relation to the translation strategies employed within each working scenario (cf. sub-sections 5.4.1 and 5.4.2), for the first pair of participants, the data obtained by means of the rankings (available in appendices O and R) were related to evidence of cognitive effort emerging at unconscious level through the pauses within TAPs (cf. sub-section 5.6). These two methods were combined in order to determine whether, by relating the data provided by each of them, it would be possible to gain a better understanding of the overall difficulty experienced by the subjects.

5.2 Definition of “Translation Strategy”

Any categorisation of translation strategies proposed so far is the result of a different perspective on either the translation or the PE process; as a matter of fact, the very notion of “translation strategy” is characterised by high variability and it is rarely defined precisely. Therefore, in the interests of clarity, a definition was selected within the present work, namely the one proposed by Lörscher (1991: 76) as an adaptation of a previous definition by Færch and Kasper (1983): “a translation strategy is a potentially conscious procedure for the solution of a problem which an individual is faced with when translating a text segment from one language into another”. This definition is particularly suitable for the purpose of this study for at least four reasons: firstly, it simplifies the identification and the subsequent analysis of

strategies by including problem-orientedness and potential consciousness among the requirements mental processes have to fulfil in order to be considered as strategies. Secondly, it is descriptive in nature, therefore it justifies the investigation of what strategies are in fact used by the subjects (regardless of which strategies they should have used).

Thirdly, by limiting the concept of strategy to problem-solving and decision-making processes, it facilitates its documentability; finally, the definition provided is characterised by a high degree of generality and therefore can be applied to the strategies adopted during both the translation and the PE task. Therefore, despite differences in the final classifications obtained, it is worth noting that any translation strategy identified originates from the attempt to solve a translation problem (cf. sub-section 4.1). With regard to this, Kiraly (1995) argues that those translation units which pose problems for the translator lead him/her to adopt translation strategies; furthermore, Krings (2001: 169) states that “[i]nextricably bound to the occurrence of problems in the production, reception or translation of any kind of text are the strategies implemented for solving these problems”.

5.3 Identification of the Number and Types of Translation Strategies

Accordingly, within this study, in order to identify the occurrences of translation strategies employed by the subjects, the analysis started from the translation problems which they verbalised and for which either the TM software or the raw output did not provide successful solutions or contained translations which needed to be checked by means of Internet search strategies (cf. sub-section 4.1). As a matter of fact, it is worth noting that, within the present study, translation strategies corresponded to the Internet searches performed by the participants in order to solve translation problems. Therefore, unlike several previous studies (e.g. Asadi and Séguinot, 2005), the focus was not on the general translation strategies underlying the majority of the subjects’ decisions and subsequent actions, but rather on more “local” strategies. The reason why strategies corresponded to Internet searches was that both the subjects using the TM software and those post-editing could rely solely on the Net so as to solve the translation problems which they encountered, when the tools at their disposal (i.e. the TM database or raw output) proved to be useless; as a matter of fact, no corpora nor other resources could be employed.

Since the purpose of this part of the research is to assess the level of difficulty perceived by the subjects in relation to their four most frequent translation strategies, it was necessary to assign each Internet search to a specific category of translation strategies. With

regard to this, it is important to bear in mind that the experimental instructions (available in appendices A and B) asked the participants to think aloud about which webpages they were consulting. This was indeed fundamental because it allowed the observation of the translation strategies to which the participants were resorting at a given moment, although they were not explicitly stated by them. Therefore, the identification of the strategies adopted was possible thanks to assumptions about the purposes of the Internet searches which the participants had performed. To give just one example, when a subject stated that he/she was consulting an online bilingual dictionary, it was assumed that he/she was engaged in a strategy of equivalent retrieval. Accordingly, strategies were identified by looking at the purposes of the Internet searches performed.

Here, an extract taken from the TAPs of a subject is provided so as to show the extent to which the participants' verbalisations were useful in shedding light on the strategies used:

*“NYSE è una sigla che nella Concordance rimane uguale. Solo per capire che cos'è lo cerco. Guardo anche su IATE. Ecco, infatti, è la Borsa di New York. Ok, quindi lo lascio così.”*¹⁰

In this case, it could be concluded that the subject was using a strategy aimed at the comprehension of the meaning of the source term. It is worth noting that, since problem-orientedness was indeed an important aspect when it came to identify translation strategies, those cases in which the subjects translated or post-edited straightforwardly (since they already knew the successful translation solution and did not need to resort to external aids) were excluded from the analysis. As a matter of fact, Lörcher (1991) too argues that the problems leading to the adoption of strategies occur when a subject realises that he/she is unable to (successfully) transfer a source-language text segment into a target-language one. The classification proposed here mainly adopted as starting points the categorisations proposed by Krings (1986) and Gerloff (1986); nonetheless, they were partly modified on the basis of the specific phenomena which were observed during the TAP experiments conducted as a part of this work. More precisely, the translation strategies which were identified are:

¹⁰ An English translation of the extract is provided: “The acronym 'NYSE' is left as it is in the TM. I'll look it up, just to know what it is. I'll use IATE too. Here it is, it's the New York Stock Exchange. Then I'll leave it as it is.”

- Equivalent retrieval
- Equivalent monitoring
- Comprehension of the source-language term
- Comprehension of the target-language term
- Contextualisation
- Reduction
- Reformulation

5.4 Assessment of the Level of Perceived Difficulty in Relation to Translation Strategies

After performing the task assigned to him/her and by means of retrospective interviews, each of the subjects was provided with his/her source text or raw output (depending on whether he had translated or post-edited), with the target text he/she had delivered (available in appendices I-N) and with a list of his/her four most frequent strategies. Next, each participant was asked to put them in order of the level of difficulty which he/she had perceived when using them, from less to more difficult. The decision to solely focus on the four strategies more frequently adopted by the participants was based on two motivations. On the one hand, it was meant to gather more reliable data, since it would have been easier for the participants to retrieve this type of information from their LTM; therefore, the possibility that they were reinterpreting their thoughts or creating them anew was reduced. On the other hand, it was assumed that the conclusions in terms of perceived difficulty drawn from strategies adopted few times could not have been generalised.

It is worth noting that the retrospective interviews were set up so as to eliminate (or, at least, reduce) the factors which might have led to a distortion of the data regarding the subjects' evaluations of the difficulty involved in the adoption of strategies. First of all, since difficulty is a relative concept, before evaluating the difficulty of strategies, the subjects were provided with a notion of "difficulty", namely with an explanation of the aspects to take into account. As a matter of fact, they were asked to think about all those cases in which Internet searches having a specific purpose (i.e. corresponding a strategy) had to be abandoned because they did not give the expected results, thus not providing any successful solution to translation problems. Secondly, had the subjects been asked to rate the level of difficulty perceived by using a, say, 10-point scale, the data obtained with regard to the values assigned might have been influenced by individual differences in the way subjects evaluate difficulty.

Therefore, in order to avoid subjective evaluations, each subject was asked to rank the four strategies most frequently used by him/her, by putting them in the relative order of the level of difficulty perceived when adopting them, from less to more difficult, and he/she was not allowed to give an equal ranking to more than one strategy. It was assumed that, in doing so, i.e. by putting strategies in order of their difficulty rather than merely assigning a value to each of them, the data obtained would have been more reliable, since the subjects would have had to evaluate a strategy by relating its difficulty to that of the others. Therefore, it would have been easier to compare the rankings of the participants. Thirdly, the difficulty perceived when consulting a given webpage may have depended on the very features of the webpage considered (such as in the case of a low-quality online dictionary). Therefore, in order to avoid evaluations of the level of difficulty which might have been influenced by the characteristics of the webpages consulted, the subjects were asked to take into consideration the different webpages which they had to consult for the same purpose (e.g. find the equivalent of a term). This decision was meant to ensure that the level of perceived difficulty would be associated with the aim of a given strategy rather than the means adopted in order to achieve such aim.

Having controlled the possible factors which might have influenced the subjects' evaluations of the difficulty involved in the adoption of a strategy, it was possible to assume that the way in which each subject ranked a given strategy was influenced by the problem from which that strategy originated. For example, it was assumed that, if the subjects perceived a high level of difficulty when adopting a strategy (e.g. in the case of the third- or the fourth- ranking strategy), this was due to the presence of particularly demanding problems. This association was paramount, since it allowed to relate the difficulty perceived by the participants while performing Internet searches to the translation problems for which the two tools either did not provide successful solutions or presented translations which the participants had to check (cf. sub-section 4.1). Before looking deeper into the analysis and discussion of the results, it is important to point out that, although the subjects were asked to rank four strategies, the analysis of this part of the research focused on just one strategy, namely the one which each subject adopted more often and that, as a result, corresponded to the majority of his/her Internet searches. It was assumed that, by focusing on the strategy which each subject employed more often during the performance of his/her task, as well as on the level of difficulty that he/she perceived when adopting it, it would have been possible to gather information about the difficulty perceived by the participant in the majority of phases during the translation or the PE process.

5.4.1 Level of Perceived Difficulty within the CAT Setting: Data Analysis and Discussion

As far as the CAT setting is concerned, the analysis focused on the relation existing between the steady increase in the number of successful translation solutions provided by the TM database and the level of difficulty perceived by each of the subjects when performing his/her most frequent strategy, i.e. during the majority of his/her Internet searches. More precisely, it was observed whether the steady increase in the number of successful solutions led to a change in the type of problem most frequently encountered and, in turn, in the type of translation strategy used more frequently in order to solve it. Subsequently, it was observed whether a change in the type of strategy adopted more often led the subjects to perceive a greater or lower level of difficulty.

Table 2.1 shows the data regarding the percentage of successful solutions provided by the TM database in each of the three sessions, the strategy most frequently adopted by each subject and the corresponding ranking assigned to this strategy on the basis of the level of difficulty perceived when employing it.

Sessions within the CAT setting	Percentage of solutions provided by the TM and accepted by the subjects (out of the overall translation problems identified)	Most frequently adopted strategy	Ranking (out of 4) of the strategy most frequently adopted
1st session	14%	Equivalent retrieval	4
2nd session	66%	Equivalent monitoring	3
3rd session	100%	Equivalent monitoring	3

Table 2.1: Relation between the number of successful solutions provided by the TM database, the most frequent strategies and their rankings

The data which were gathered show that:

- as sessions take place, there is a change in the types of strategies most frequently adopted by the subjects throughout the translation process;
- the ranking assigned to the strategy of equivalent retrieval is higher than the ranking assigned to the strategy of equivalent monitoring.

Starting from these data, some conclusions can be drawn. First of all, it could be easily assumed that the change in the types of strategies which were most frequently adopted by the subjects was the result of the steady increase in the number of solutions provided by the TM

database. As a matter of fact, the shift from the strategy of equivalent retrieval (the most used during the first session) to that of equivalent monitoring (the most employed during the second and third session) suggests that the updating of the TM database resulted in a difference in the types of translation problems which the participants had to solve more often by means of Internet searches. More precisely, the subjects were increasingly led to assess the correctness of the results provided, rather than search them from scratch, and this shift led to a decrease in the level of difficulty perceived during the majority of the Internet searches performed, thus reducing the overall difficulty involved in the translation process itself.

5.4.2 Level of Perceived Difficulty within the PE Setting: Data Analysis and Discussion

With regard to the PE scenario, each session was considered separately. Accordingly, the number of successful solutions contained in the raw output within each of the three sessions was related to the ranking indicating the level of difficulty perceived by each of the subjects when adopting his/her most frequent strategy. Subsequently, the results from the three sessions were compared so as to determine whether the differences in terms of the number of successful translation solutions contained in the raw output corresponded to a difference in the type of translation problem identified more often, in the type of translation strategy adopted in order to solve it and, as a result, in the level of difficulty perceived during the majority of Internet searches.

Table 2.2 shows the data regarding the percentage of successful translation solutions provided by the output within each session, the strategy most frequently adopted by each subject and the corresponding ranking assigned to this strategy on the basis of the level of difficulty perceived when employing it.

Sessions within the PE setting	Percentage of solutions contained in the output and accepted by the subjects (out of the overall translation problems identified)	Most frequently adopted strategy	Ranking (out of 4) of the strategy most frequently adopted
1st session	58%	Contextualisation	4
2nd session	83%	Contextualisation	4
3rd session	50%	Contextualisation	4

Table 2.2: Relation between the number of successful solutions provided by the raw output, the most frequent strategies and their rankings

The data which were gathered show that:

- regardless of the different number of solutions accepted by each subject within each of the three sessions, there is no variation in the type of strategy which each of them adopted more often (i.e. contextualisation);
- all three subjects gave an equal ranking to the strategy of contextualisation.

Starting from these data, some conclusions can be drawn. First of all, the fact that the strategy most frequently adopted by all three subjects was that of contextualisation suggests that the type of translation problem which all three subjects encountered more often was the same. In particular, it should be noted that the strategy of contextualisation refers to those cases in which the subjects performed Internet searches aimed at solving stylistic problems within the output. In those cases, they stated that they were searching comparable texts (i.e. comparable press releases) so as to determine whether the output respected the stylistic features of this text type. Accordingly, it was assumed that, even in those cases in which the output provided successful solutions to more than half the translation problems encountered (such as in the first and second sessions), the solutions provided did not include those to the stylistic problems within the text. Secondly, it should be noted that all three subjects gave the higher ranking to the strategy of contextualisation. This may be regarded as a consequence of the fact that, not only the solutions provided were not solutions to the stylistic problems encountered, but also that, in order to identify the stylistic features of the text type, the participants were led to perform more in-depth (and, accordingly, more difficult searches). Therefore, it could be argued that the different number of successful solutions identified in the output did not influence the level of difficulty perceived by the subjects when performing the

majority of their Internet searches throughout the PE process.

5.5 Summary

The data presented in sections 5.4.1 and 5.4.2 shed light on the extent to which one of the features of the working scenarios considered (i.e. the number of successful translation solutions respectively provided by the tools employed) could affect the types of strategies adopted by the subjects during the process of translation or PE, and, in turn, the level of difficulty involved in them. In particular, it was observed that the tool employed within the CAT scenario (i.e. the TM software) reduced the difficulty perceived by the participants to a larger extent as far as Internet searches are concerned, by providing the subjects with an increasing number of translations which they could use as starting points during their searches for target-language equivalents. On the contrary, with regard to the PE setting, the three texts were mostly characterised by stylistic problems which were regarded by the subjects as the most difficult to solve, regardless of the number of successful solutions which were available within the texts. Therefore, it could be argued that the successful translation solutions contained in the raw output did not reduce the difficulty involved in the PE process when it came to resorting to the most frequent strategy, namely the one aiming at solving stylistic problems. Finally, it should also be noted that, in one case out of three, the post-editors were led to adopt with more frequency a strategy more difficult than the one adopted within the CAT setting even if the raw output contained a larger number of successful translation solutions (i.e. for the second pair of participants).

5.6 Analysis of Cognitive Effort¹¹ and Relation with the Level of Perceived Difficulty

5.6.1 Aim of the Analysis of Cognitive Effort in the Present Study

Within this sub-section, pauses meeting some criteria (cf. sub-section 5.6.2) were deemed to be indicators of cognitive effort on the part of the subjects, and they were used in order to conduct a case study on the overall level of difficulty experienced by the first pair of participants when employing translation strategies. More precisely, evidence of difficulty emerging at unconscious level by means of pauses within TAPs was related to the rankings (available in appendices O and R) of the level of perceived difficulty obtained via retrospective interviews. Therefore, since the participants ranked the difficulty perceived in relation to Internet search strategies (cf. sub-section 5.4), in order to allow a comparison between these two aspects of difficulty, the pauses which were analysed were those occurring during Internet searches (i.e. during the performance of translation strategies). With regard to this latter aspect, it is worth noting several studies have dealt with the relation between pauses and use of aids. To give just one example, Alves and Liparini Campos (2009) analysed the pauses which were observed while the participants of their experiment resorted to both internal or external aids.

Nonetheless, unlike their study, the present work exclusively focused on the pauses recorded while the subjects resorted to external aids. This decision was due to the fact that, in the vast majority of cases, the subjects were not able to find a solution to a translation problem on their own, possibly as a consequence of the fact that they were not familiar with the text type assigned to them. On the contrary, the professional translators involved in Alves and Liparini Campos's (2009) experiment often found solutions without resorting to any external aid (possibly as a result of being more self-confident or having more expertise than trainee translators). It is important to note that this part of the research had a double aim: firstly, to observe the possible differences between two manifestations of difficulty; secondly, to determine whether, by combining these two data gathering techniques (namely TAPs and retrospective interviews), a more comprehensive picture of the overall difficulty experienced by the subjects while employing translation strategies could be obtained.

As a matter of fact, during retrospective interviews, each of the participants ranked solely the four strategies which he/she had employed more often. The decision to select only few strategies to be presented was due to the fact that asking the subjects to rank all those

¹¹ Within the present study, cognitive effort is treated as a part of the overall difficulty experienced by the participants.

which they had used (including, for example, strategies employed once or twice) would have compelled them to retrieve very specific pieces of information already stored in their LTM. Accordingly, this might have resulted in a distortion of the data gathered. Therefore, through rankings, it was possible to gather data on the perceived difficulty only for a limited number of strategies. Furthermore, since these data corresponded to subjective perceptions, it was felt that more objective evidence would help to get a more detailed picture of the overall difficulty experienced during the translation and the PE task.

5.6.2 Criteria for the Identification of Pauses to Be Analysed

Since pauses can depend on factors different from cognitive effort (e.g. lack of concentration, boredom, etc.), some criteria were set in order to identify pauses worth of analysing. The first criterion which was set referred to the length of pauses both within the CAT and the PE setting: only pauses longer than eight seconds were taken into consideration. Although numerous shorter pauses related to Internet searches could be identified, only pauses longer than eight seconds were analysed so as to exclude from the analysis the time which the subjects supposedly spent typing the search, finding the desired webpage(s) and reading the results obtained for the first time. Therefore, the focus of the analysis was on the actual cognitive effort which the participants experienced when considering and evaluating the possible solution(s) provided by the strategies to which they resorted (i.e. by the webpages which they consulted).

The time criterion of eight seconds also allowed me to filter out from the analysis all those cases in which a problem was identified by a subject (as demonstrated by the presence of a problem indicator) (cf. sub-section 4.1), but its solution did not demand cognitive effort on his/her part (such as in the case of a dictionary look-up rapidly providing a solution that a subject deemed to be successful). With regard to the measurement of the length of pauses, only those time intervals were measured which occurred between a problem indicator revealing that a subject was going to resort to an Internet search strategy (cf. sub-section 4.1) and the beginning of the sentence immediately following the pause (which generally contained the subject's reflections on the solution found or on the impossibility to find a solution via a given strategy). Those cases in which the subject began a search and verbalised his/her actions at a later stage (or even when the search had already concluded) were removed from the analysis, since no clear starting point of the process could be identified and, as a consequence, the length of the pause could not be measured.

The second criterion which was set refers to the presence of specific indicators within

pauses which were assumed to signal that a subject was concentrating on the task being performed and that, as a result, the pause was due to cognitive effort on his/her part. The first type of indicators is represented by fillers¹², such as “ehm”, “hm”, etc. In his study, Lörcher (1991) assumes that their presence within a pause might indicate that the subject has identified a translation problem and is concentrating on its solution. The second type of elements which were assumed to indicate that a pause was due to cognitive effort is represented by the subjects’ interruptions of moments of silence during which they lowered their voice and either repeated some words to themselves (Lauffer, 2002) or read the results of their online searches, without saying what they thought about their quality (although they had been instructed to do so) (cf. sub-section 3.2.5).

When one or both these two types of elements were observed within pauses, it was assumed that the moments of silence preceding and following them were moments characterised by cognitive effort during which the participants were engaged in considering the results of their online searches. However, it should be noted that, since both elements are vocalisations, their duration was removed from the measurement of the length of the pauses during which they occurred. Finally, in order to allow a more direct comparison between the CAT and the PE setting, only those pauses were considered which could be observed during the performance of Internet search strategies aimed at solving the translation problems identified by both participants within the first group (see Table 1.7).

5.6.3 Data Analysis and Discussion

On the basis of the criteria which were set so as to identify those moments of silence which would be worth taking into account (cf. sub-section 5.6.2), ten pauses were analysed. They emerged as a result of the cognitive load involved in the Internet searches carried out so as to solve three translation problems represented by the translation of an acronym (“NYSE”), an adjective (“outstanding”) and a collocation (“cash flow”). For each of these three translation problems, all the translation strategies adopted were presented (regardless of whether they required cognitive effort or not).

Table 3.1 summarises the data which were gathered.

¹² It is worth noting that, within this study, fillers are also treated as problem indicators signalling that a participant was not satisfied with the solutions provided by either the TM database or the raw output (cf. sub-section 4.1).

TRANSLATION PROBLEMS	Strategies and (length of pauses) in the CAT scenario	Strategies and (length of pauses) in the PE scenario
NYSE	Comprehension of the source-language term¹³ (43 sec.) + ¹⁴ Equivalent retrieval	Contextualisation (30 sec.) + Equivalent monitoring
OUTSTANDING	Equivalent retrieval (33 sec.) + Equivalent retrieval (30 sec.) + Equivalent retrieval + Equivalent monitoring + Equivalent monitoring	Comprehension of the source-language term (41 sec.) + Comprehension of the source-language term + Equivalent retrieval + Equivalent monitoring
STRONG CASH FLOW	Equivalent retrieval (35 sec.) + Equivalent monitoring (8 sec.) + Equivalent monitoring (11 sec.) + Equivalent monitoring	Equivalent retrieval (30 sec.) + Equivalent monitoring (30 sec.) + Equivalent monitoring + Contextualisation + Contextualisation + Equivalent monitoring

Table 3.1: Strategies adopted for the solution of translation problems and related pauses (when present) resulting from cognitive effort

The data which were gathered show that:

- when being adopted in both settings in order to solve the second common translation problem (i.e. the translation of “outstanding”), the strategy of equivalent retrieval is associated with pauses indicating cognitive effort only within the CAT scenario. Furthermore, when looking at the third translation problem (i.e. the translation of “strong cash flow”), it can be observed that the strategy of equivalent retrieval is associated with pauses indicating cognitive effort in both scenarios, but the duration of these pauses is greater within the CAT setting. Finally, it is worth noting that this strategy is associated with pauses indicating cognitive effort in the majority of cases in which it is employed (i.e. three out of five).

¹³ The translation strategies in bold are those which required cognitive effort, as indicated by the duration of the pauses associated with them. The remaining strategies are those for which no evidence of cognitive effort was identified.

¹⁴ The “plus” sign (+) after a translation strategy indicates that it was followed by a further strategy.

- when requiring cognitive effort in both scenarios (namely for the translation of “strong cash flow”), the strategy of equivalent monitoring is associated with pauses having an overall longer duration within the PE scenario than within the CAT scenario;
- within the CAT setting, the strategy of comprehension of the source-language term is associated with a longer pause indicating cognitive effort.

Starting from these data, some conclusions can be drawn. First of all, as could be expected, the strategies of equivalent retrieval and comprehension of source-language term required greater cognitive effort when being employed within the CAT setting; this was probably caused by the fact that, in the majority of cases, the participant did not have a starting point for his/her searches (unlike the subject who post-edited). On the other hand, the strategy of equivalent monitoring required more cognitive effort within the PE setting; this may be due to the fact that, when having to monitor a translation contained in the output, the participant was aware of the fact that it did not directly come from a human translator, but rather from the recombining carried out by the statistical machine translation (SMT) system used. Accordingly, this might have led him/her to be warier of the translation proposed.

Table 3.2 reports the results obtained via the rankings, in order to allow the combining of the data.

	Frequent strategies employed within the CAT scenario	Frequent strategies employed within the PE scenario
Level of difficulty perceived in relation to strategies (4=most difficult; 1=least difficult)	4- Equivalent retrieval 3- Contextualisation 2- Equivalent monitoring 1- Comprehension of the target-language term	4- Contextualisation 3- Equivalent retrieval 2- Equivalent monitoring 1- Comprehension of the source-language term

Table 3.2: Rankings of the four Internet search strategies most frequently employed by the 1st pair of participants

Before relating the results from the two data gathering techniques, a selection was made with regard to the data to be analysed. Accordingly, the following cases were not object of study:

- cases in which a strategy perceived as being very difficult is not present among those adopted for the solution of a translation problem (e.g. the strategy of contextualisation within the CAT setting). As a matter of fact, due to the limited number of translation problems considered for the analysis of cognitive effort (namely the three problems identified by both subjects within the first pair), it is not possible to draw conclusions regarding a difference in the evidence provided by the two methods. In other words, the possibility cannot be ruled out

that, if a strategy is not used for the solution of the three translation problems, this is due to the very features of those specific translation problems (which, in turn, require specific strategies), rather than on a difference in the types of data respectively provided by the two methods.

Once irrelevant data were filtered out, it was possible to focus on evidence pertinent to the purpose of this part of the research. In particular, it was observed that:

- the strategy of equivalent retrieval requires cognitive effort in the majority of cases in which it is adopted within the CAT setting (i.e. three out of five), and this result is confirmed by the rankings, since the participant using the TM software considers this strategy as the most difficult;
- the strategy of equivalent monitoring requires greater cognitive effort within the PE setting; nonetheless, when looking at the rankings, it can be noted that the two participants assign the same level of perceived difficulty to it;
- within the CAT setting, the strategy of comprehension of the source-language term requires cognitive effort in the only case in which it is adopted for the solution of the three translation problems considered; nonetheless, it is not ranked since it is not among the four most frequent strategies adopted by the subject.

5.6.4 Summary

On the basis of the small sample of data presented within this section, some conclusions can be drawn. First of all, it can be noted that, in some cases (such as for the strategy of equivalent retrieval within the CAT setting), the results obtained via the rankings and via the analysis of pauses coincided, thus showing that the level of difficulty which the participant perceived was mirrored by evidence of cognitive effort emerging in an unconscious way. In other cases (such as for the strategy of equivalent monitoring), the findings of retrospective interviews were not in line with those from TAPs. Since the sample of data analysed in the present section was very limited, it is not possible to draw general conclusions about the extent to which combining these two methods can shed light on the differences and similarities between the various manifestations of the difficulty involved in the translation and the PE task. Nonetheless, the results obtained show that these aspects are worth of further analysis, since they might provide evidence of how the CAT and the PE setting differ in terms of the relation between perceived and unconscious difficulty. Finally, it should be noted that, in some cases (such as for the strategy of comprehension of the source-language term within the CAT setting), analysing pauses within TAPs can provide evidence of difficulty which could

not have been gathered by means of retrospective interviews, due to the limited number of strategies to be ranked. Therefore, by combining these two methods, it would be possible to gain a more comprehensive picture of the overall difficulty experienced by the participants during the various phases of the translation and the PE process.

Chapter 6

TIME EFFECTIVENESS

6.1 Aim of This Part of the Research and Method Adopted

This part of the research aims at determining which of the two working scenarios between using TM software and post-editing raw MT output is the most time-effective scenario, mainly by focusing on the time difference between them. It is worth noting that the time taken either by the translation or the PE process may be influenced by external factors, such as the use of TAPs as a data gathering technique (cf. sub-section 2.2). As a matter of fact, several studies showed that TA slows down both the translation (Jakobsen, 2003) and the PE process (Krings, 2001). Since the time taken by the subjects is one of the aspects being analysed within this study, the decision to resort to TAPs might have represented an issue. Nonetheless, it is worth noting that these studies differ from the present work with regard to numerous aspects, such as sample of participants, text type, language direction, etc. Therefore, the percentages indicating time losses due to TA within these studies can hardly be applied to the present work. Moreover, in order to reduce the possible influence of TA on the time required to complete a task, the participants involved in this study were not asked to verbalise all their thoughts, but rather some specific types of information (cf. sub-section 3.2.5). This was assumed to have less impact on their usual behaviour and, accordingly, on the time which they would normally take. To give just one example, they were instructed to indicate the webpages which they consulted, but they did not have to verbalise the thoughts underlying their choice of a given webpage.

Therefore, the time taken was measured without taking into account the possible influence of TA on either the translation or the PE task, since it was assumed to be negligible. In order to gather evidence of the time difference between the two working scenarios considered, the overall duration of each of the tasks carried out by the subjects was measured. Subsequently, for each pair of participants, the time needed by the subject using the TM software was compared to the time taken by the subject post-editing, so as to measure the difference in duration between the two tasks. Then, for each pair, this difference was related both to the duration of the shorter task and to that of the longer one, with the aim of assessing the percentage represented by time difference out of the overall duration of each of the two tasks. As a result, it was possible to measure the time gains and the time losses respectively resulting from the shorter and the longer task. Finally, these results were related to the data on

the percentage of successful translation solutions respectively contained in the TM database and the raw output (cf. sub-sections 4.2.1 and 4.2.2), so as to determine whether a relation could be identified between the time taken by the participants and the number of successful solutions provided to them.

6.2 Data Analysis and Discussion

Figure 5 reports the respective duration of the translation and the PE tasks.

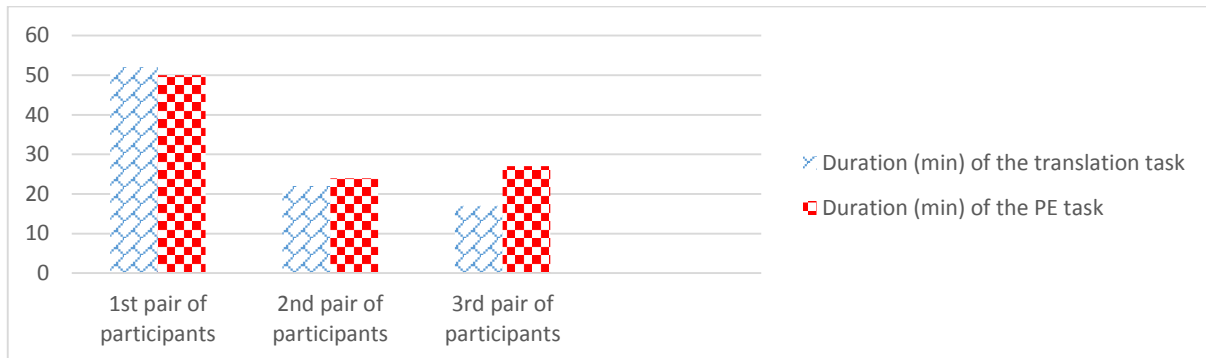


Figure 5: Duration of the translation and the PE tasks

The data gathered show that:

- the duration of the translation tasks performed by using the TM software steadily decreases as experimental sessions are held;
- within the first pair of participants, the performance of the subject using the TM software is two minutes longer than the performance of the subject post-editing the raw MT output. This time difference corresponds to 4% of the duration of the shorter task (i.e. the PE task) and to about 3.8% of the duration of the longer task (namely, the translation task);
- within the second pair of participants, the subject post-editing takes two minutes more than the subject using the TM software. This time difference represents about 9% of the shorter task (i.e. the translation task) and about 8.3% of the longer task (namely, the PE task);
- within the third pair of participants, the performance of the subject post-editing is ten minutes longer than that of the subject using the TM software. This time difference corresponds to about 58.8% of the duration of the shorter task (namely, the translation task) and to about 37% of the duration of the longer task (i.e. the PE task).

Tables 4.1 and 4.2 summarise the data regarding the relation between, on the one hand, the variation in the time gains and time losses of the CAT and the PE setting respectively (which could be observed by comparing the second and third pair of participants) and, on the other hand, the variation in the percentages of successful translation solutions respectively contained in the TM database and the raw output.

	Time gains	Percentage of successful solutions provided (out of the overall translation problems verbalised)
Participant within the 2nd pair	~9%	66%
Participant within the 3rd pair	~58.8%	100%

Table 4.1: Relation between time gains and successful solutions within the CAT setting

	Time losses	Percentage of successful solutions provided (out of the overall translation problems verbalised)
Participant within the 2nd pair	~8.3%	83%
Participant within the 3rd pair	37%	50%

Table 4.2: Relation between time losses and successful solutions within the PE setting

The data gathered show that:

- for the second and third pair of participants, namely in those cases in which translating with the help of the TM software proves to be the most time-effective option, the steady increase in the time gains resulting from this working scenario (i.e. from about 9% to about 58.8%) corresponds to a steady increase in the number of successful solutions provided by the TM database (i.e. from 66% to 100%). On the other hand, with regard to the PE setting, for these two pairs of participants, the higher time loss (namely, of about 37%) corresponds to the lower percentage of successful solutions (i.e. 50%), while the lower time loss (namely, of about 8.3%) corresponds to the higher percentage of successful translation solutions provided (i.e. 83%).

As can be seen in Figure 5, translating with the help of the TM software proved to be more time-effective than post-editing raw MT output in two out of three cases. On the other hand, post-editing raw MT output proved to be more time-effective than using the TM software only in one out of three cases. In particular, the difference in duration between the two tasks became more evident when comparing the performances of the participants within the third pair. These data can be explained by taking into account the updating of the TM

software which was performed by the subjects who used it. As a matter of fact, it is important to remember that the three texts assigned were very similar in terms of form and content; therefore, the number of matches and solutions provided by the TM database steadily increased as sessions were held, thus steadily reducing the amount of time taken by the subjects (see Table 4.1). However, it should be noted that, also within the PE scenario, the variation in the number of successful translation solutions which were available in the various sessions had an impact on the time taken by the post-editors (see Table 4.2). A further aspect which is worth taking into account is the fact that, for the second pair of participants, the subject post-editing took more time than the subject using the TM software despite the fact that the number of successful translation solutions contained in the raw output was greater (see Tables 4.1 and 4.2).

With regard to the data referring to the difference in duration between the two tasks within each pair, it is also worth noting that, both for the first and second pair of participants, this difference amounted to two minutes (see Figure 5). Although this low value may appear not to be noteworthy, when related to the overall duration of the tasks considered, it can contribute to a better understanding of the difference between the time gains and the time losses resulting from each of the two working scenarios considered. As a matter of fact, when the PE task represented the most time-effective option (i.e. for the first pair of participants), the percentage of time saved by the subject post-editing was lower than the percentage of time saved by the subject using the TM software within the second pair of participants, namely when the translation task was the most time-effective option. These data on the percentages of time gains are further reinforced by those indicating the percentages of time losses. As a matter of fact, for the first pair of participants (namely, in the only case in which translating with the help of the TM software required more time than post-editing), the percentage of time losses characterising the translation task was lower than the percentage of time losses resulting from the PE task within the second and the third pair of participants, i.e. when PE proved to be the most time-consuming scenario.

6.3 Summary

On the basis of the results provided by these time measurements, some conclusions can be drawn. First of all, it would seem that, on the whole, translating using a TM database updated with relevant segments allowed trainee translators to save more time than post-editing raw MT output. Secondly, even when translating with the help of TM software required more time than post-editing raw MT output, the resulting time loss affected the overall duration of the

translation task less than the time loss which characterised the PE task (when the latter was the most time-consuming activity). Furthermore, it should be noted that, for both working scenarios, the number of successful translation solutions contained in the TM database and the raw output proved to have an impact on the time needed in order to perform the translation and the PE tasks.

6.4 Findings of the Present Work

6.4.1 Connecting Previous Results

The aim of this work was to determine which of the two working scenarios between using TM software and post-editing raw MT output proved to reduce the perceived difficulty and the time needed by trainee translators to a larger extent. More precisely, in order to determine the respective impact of TM software and raw output on the translation and the PE process, in the present sub-section, the number of successful translation solutions provided by these two tools was related, on the one hand, to the difficulty perceived by the participants when adopting translation strategies and, on the other, to the time required. When looking solely at the number of successful translation solutions provided, what emerged was the fact that, although the CAT scenario was characterised by a steady increase in the number of successful solutions contained in the TM database, this tool was able to provide more correct solutions than the raw output only after its second updating, namely to the third participant. As a matter of fact, when analysing the performances of the participants within the first and the second pair, it was observed that the raw output provided a larger number of successful translation solutions. In order to determine whether and how this aspect influenced the level of difficulty perceived by the trainee translators, the number of successful translation solutions was related to the data obtained via the rankings (available in appendices O-T).

It was observed that the steady increase in the number of correct solutions contained in the TM database led the participants to adopt with more frequency a strategy which they perceived as being less difficult, thus reducing the overall level of difficulty which they experienced during most phases of the translation process. On the contrary, within the PE setting, the variations in the number of successful translation solutions contained in the raw output had no impact on the translation strategy adopted more often by the subjects nor on the level of difficulty assigned to it. As a result, the strategy employed with more frequency by all three participants was also the one which they perceived as being the most difficult. Therefore (unlike within the CAT setting), there was not a decrease in the overall level of difficulty associated with the PE process. With regard to the time respectively taken by the translation and the PE tasks, the findings showed that, in most cases, translating with the help of the TM software was a more time-effective option than post-editing raw output. Furthermore, it was observed that, unlike with perceptions of difficulty, for both working scenarios, the variation in the number of successful translation solutions provided by the two tools had an impact on the time taken by the participants.

6.4.2 Main Findings

By relating the findings regarding the various aspects of the translation and the PE process which were analysed, it was possible to draw some conclusions. Firstly, it was observed that, as sessions were held, unlike post-editing raw MT output, translating by using the TM software allowed trainee translators to steadily reduce the level of difficulty perceived when performing the majority of their Internet searches. Furthermore, on the whole, working within the CAT setting proved to be a more time-effective option than working within the PE setting. Both these results might seem surprising when thinking of the fact that, in most cases, the raw output provided a larger number of translation solutions which were deemed to be successful by the participants. Therefore, it could be argued that, in addition to the very features of the two tools being employed (i.e. their usefulness in terms of translation solutions provided), further aspects (such as the subjects' attitudes) should be taken into account in order to explain the differences in terms of perceived difficulty and time required between the two working scenarios. As a matter of fact, although in some cases these findings could be explained on the basis of the differences in the number of successful translation solutions available to the participants, in other cases they could not.

Therefore, it was argued that also the subjects' attitudes towards the tasks being performed influenced the aspects of the translation and the PE process being analysed. In particular, as far as the level of perceived difficulty is concerned, the fact that the post-editors ranked as the most difficult their most frequent strategy (namely the one aiming at identifying the stylistic features of the texts to be post-edited) might have been caused by their awareness of the fact that the raw output had resulted from a recombining of previous human translations carried out by the SMT system used. This might have led them to think that the style characterising the original human translations had been one of the aspects being particularly altered; accordingly, they might have preferred to perform more in-depth (and, as a result, more difficult) searches. As far as the time needed is concerned, in this case too, the possibility should not be ruled out that the participants took more time when post-editing because they were warier of the solutions provided and, as a result, were led to perform a larger number of Internet searches (thus needing more time) in order to monitor the translations contained in the raw MT output. Accordingly, even though the CAT scenario proved to be the best option in terms of difficulty perceived and time needed, it should be noted that this might not depend solely on the very features of the tool being employed (i.e. the TM software), but also on the participants' stronger sense of trust in the translations

provided by the TM software, as opposed to those contained in the raw output.

This result is in line with the findings of Teixeira's (2014) study, which investigated the correlations between the translators' perceptions and their actual performances when translating without translation suggestions, when translating with translation suggestions from both TM and MT and metadata about the suggestions, and, finally, when translating by using pre-translated texts from TM and MT but without metadata. In the interviews conducted after the task, the participants of Teixeira's (2014) study explained that they did not feel at ease when working without metadata on the provenance of translations because, in their routine tasks, they use different strategies for exact matches, fuzzy matches and MT suggestions. Therefore, he found that different types of translation tasks activate different translation strategies which, in turn, can be assumed to involve different levels of difficulty. Although Teixeira's (2014) study was not conducted with trainee translators, but rather with professional translators, as in the case of my work, it could be argued that being aware of the provenance of a translation suggestion influences the participants' performances.

Chapter 7

CONCLUSIONS

7.1 Assessment and Limitations of the Research

The aim of this work was to assess the respective impact of TM use and PE of raw MT output on the performance of trainee translators in terms of perceived difficulty and time required. In particular, it was observed that the CAT setting led translation students to perceive a lower level of difficulty, while at the same time being the most time-effective option. It is important to note that, within the present study, several limitations can be identified which lay emphasis on the need to further test the results obtained and extend this line of research. First of all, one limitation is connected to the sample of subjects who were selected, since it was very limited and representative of a very specific context. Accordingly, it is not possible to know whether the results obtained can be applied to, say, a larger number of translation students who received different training. Secondly, due to time constraints, it was not possible to test the participants in terms of, on the one hand, translation competence and, on the other hand, technical skills. This would have been advisable in order to determine whether, despite receiving the same training, some individual differences which would affect the performance could be identified. Moreover, the passages used for this experiment belonged to just one text type with specific features; therefore, the possibility cannot be ruled out that, had the participants been asked to translate or post-edit different texts, they might have adopted different translation strategies involving a different level of perceived difficulty and a different amount of time.

In addition, it should be noted that this study concentrated solely on the language combination English-Italian, and that the participants translated and post-edited from English into Italian (i.e. their L1). As a result, in this case too, it is not possible to know whether translating from L1 into L2 would have led the subjects to take a different amount of time and adopt different strategies within each of the two working scenarios; nor is it possible to determine whether, by having the subjects translate and post-edit from another language (i.e. Spanish or French) into Italian, the results would have been the same as those obtained. A further limitation is connected to the very features of the CAT setting; more precisely, to the decision to provide the participants with a TM database which was steadily updated as sessions were held. In several experiments (e.g. Alves and Liparini Campos, 2009), the participants who worked using a TM database had access to the same identical TM content.

Nonetheless, within the present work, the decision to ask the participants to update the TM database with their translations was motivated by the need to compare the TM software and the raw MT output on the basis of the very features which characterise each of these two tools. In other words, translating with the help of a TM database without taking advantage of its distinctive feature (i.e. the storing of previously translated segments) would have created an unrealistic working scenario, which would compromise the validity of the experiment. Despite this, it is worth noting that this decision involved the drawback of having the three participants work within an ever-changing scenario, thus not allowing the complete control of this variable.

Moreover, a further crucial limitation is connected to the primary method being employed, namely TAPs. As a matter of fact, although the adoption of this method had the merit of reducing the pressure put on the participants, the lack of an objective recording tool (e.g. keystroke loggers or screen recorders) did not allow me to know whether the subjects actually verbalised all their actions while performing the tasks. Finally, it is worth noting that, within the present study, although the participants were asked to deliver publishable texts, their final products (available in appendices I-N) were not evaluated. As a result, although the impact of two working scenarios was analysed and compared, it was not possible to determine which of them led the trainee translators to deliver texts of higher quality. This can be regarded as a limitation too, since in the translation market, the need to save time and resources is as important as the quality of the final products. Furthermore, it is worth noting that the analysis of the quality of the target texts (conducted by either professional translators or translator trainers) would have allowed me to rule out the possibility that a lower level of perceived difficulty or a lower amount of time were caused by a performance resulting in a text of lower quality.

7.1.1 Areas for Future Work

Starting from these limitations, it is possible to identify some areas for future research. First of all, the results obtained from the present analysis are worth testing on a larger sample of participants (so as to limit the possible effects of individual differences), as well as on subjects with different backgrounds (e.g. translation students with no experience of PE, or professional translators familiar with this activity, etc.). This would allow the observation of whether and how the type of training and the years of work experience might influence the subjects' perceptions of difficulty and the time which they take when using a given translation technology. Secondly, it would also be interesting to conduct a similar experiment by

assigning the participants other text types, by analysing further language pairs and by changing the translation direction, with the tasks performed into the L2 of the users. Thirdly, it remains to be seen whether, by changing the features of the working scenarios (e.g. by providing the participants with the same TM content), the comparison between the CAT and the PE setting would confirm the findings of the present work. Moreover, it would be interesting to repeat the same experiment by combining the TA methodology with an objective recording tool (e.g. keystroke loggers or screen recorders) which might improve the reliability of the data gathered. Nonetheless, in order not to make the participants work within unfamiliar settings and feel under pressure, it would be advisable to conduct several warm-up exercises during which they might become familiar both with the recording tools being used and with the idea of being constantly observed during the performance of a task.

Finally, it would be advisable to extend the present work also by combining the findings of this process-oriented research with an analysis of the quality of the final products. With regard to this, Alves et al. (2010) favour product-oriented research being combined with process-oriented one, since both approaches present challenges which can best be overcome when adopted in combination with each other. In particular, when it comes to the quality assessment of the final product, relating this aspect to process findings allows a better understanding of the translation process (Saldanha and O'Brien, 2014). An example of a study using a combined approach is the CORPRAT project, which collected five data types: on the one hand, keystroke logs, eye-tracking metrics, audio recordings and screen recordings, and, on the other hand, annotated target texts. In particular, when looking at the aspects analysed within the present work, it can be noted that, in order to expand this line of process-oriented research so as to include quality assessment, special attention should be paid to the relation between quality and time. As a matter of fact, aiming at different levels of quality can deeply influence the time which the participants spend, say, pondering or revising their translations, considering the results of their searches, etc. With regard to this, Guerberof (2009) points out that the analysis of productivity (in terms of processing speed) should be done in relation to an equal level of final quality. Therefore, if the aim of a work is to measure translation or PE speed, an analysis of the quality of the final product would shed light on differences in terms of the time required which might depend on a different level of quality of the target texts (if keeping all other variables controlled).

7.2 Lessons Learned

From my point of view, carrying out this study was a fruitful experience for numerous

reasons. First of all, the very process of selection of a research topic was a way to reflect upon the more general areas of the translation and the PE process in order to identify a narrower area worth of investigating and adding to the body of knowledge. The final choice which I made proved to be interesting in the sense that, in addition to the analysis of the time taken, it also allowed me to investigate the translation and the PE process from an original point of view, i.e. by looking at perceptions of difficulty associated with Internet search strategies. Secondly, when conducting this study, I became aware of the fact that, although translation and PE process experiments are artificial in nature, a deep knowledge of reality is required on the part of the researcher in order to identify all the various aspects which might influence a given phenomenon. Therefore, setting up an experiment led me to carefully observe the reality of the context which I wanted to analyse. Finally, I became aware of the merits and the drawbacks which characterise the most employed methods within these fields. This was particularly useful, since I realised that any research, regardless of the method adopted, presents limitations which call for future work in order to make further progress. As a result, it became clear to me that research is an ongoing and collaborative process.

REFERENCES

- Alves, F. ed. (2003). *Triangulating Translation: Perspectives in Process Oriented Research*. Amsterdam: John Benjamins.
- Alves, F. (2006). "A Relevance-theoretic Approach to Effort and Effect in Translation: Discussing the Cognitive Interface between Inferential Processing, Problem-solving and Decision-making". In *Proceedings of the International Symposium on New Horizons in Theoretical Translation Studies*. Hong Kong, China, 19-20 January, pp. 1-12.
- Alves, F. and Liparini Campos, T. (2009). "Translation Technology in Time: Investigating the Impact of Translation Memory Systems and Time Pressure on Types of Internal and External Support". In A. L. Jakobsen, I. M. Mees and S. Göpferich, eds. *Behind the Mind: Methods, Models and Results in Translation Process Research*. Copenhagen: Samfundslitteratur, pp. 191-210.
- Alves, F., Pagano, A. and da Silva, I. (2010). "A New Window on Translators' Cognitive Activity: Methodological Issues in the Combined Use of Eye Tracking, Key Logging and Retrospective Protocols". In I. M. Mees, F. Alves and S. Göpferich, eds. *Methodology, Technology and Innovation in Translation Process Research. A Tribute to Arnt Lykke Jakobsen*. Copenhagen: Samfundslitteratur, pp. 267-291.
- Alves, F., Pagano, A. and Neumann, S. (2010). "Translation Units and Grammatical Shifts: Towards an Integration of Product- and Process-based Translation Research". In G. M. Shreve and E. Angelone, eds. *Translation and Cognition*. Amsterdam: John Benjamins, pp. 109-142.
- Alves, F. Gonçalves, J. L., Szpak, K. (2012). "Identifying Instances of Processing Effort in Translation through Heat Maps: An Eye-tracking Study Using Multiple Input Sources". In M. Carl, P. Bhattacharya and K. K. Choudhary, eds. *Proceedings of the First Workshop on Eye-tracking and Natural Language Processing*. Mumbai, India, 15 December 2012, pp. 5-20.
- Angelone, E. (2010). "Uncertainty, Uncertainty Management, and Metacognitive Problem Solving in the Translation Task". In G. M. Shreve and E. Angelone, eds. *Translation and Cognition*. Amsterdam: John Benjamins, pp. 17-40.
- Asadi, P. and Séguinot, C. (2005). "Shortcuts, Strategies and General Patterns in a Process Study of Nine Professionals". *Meta*. 50(2): 522-547.
- Aziz, W., Sousa, S. C. M. and Specia, L. (2012). "PET: A Tool for Post-editing and Assessing Machine Translation". In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, J. Odiijk and S. Piperidis, eds. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey, 21-27 May 2012, pp. 3982-3987.
- Bernardini, S. (2001). "Think-aloud Protocols in Translation Research: Achievements, Limits, Future Prospects". *Target*. 13(2): 241-263.
- Chesterman, A. (1997). *Memes of Translation: The Spread of Ideas in Translation Theory*. Amsterdam/Philadelphia: John Benjamins.
- Claparède, E. (1932). "Die Entdeckung der Hypothese". In C. F. Graumann, ed. *Denken*.

Köln/Berlin: Kiepenheuer & Witsch, pp. 109-115.

Dam, H. V., Engberg, J. and Gerzymisch-Arbogast, H. eds. (2005). *Knowledge Systems and Translation*. Berlin: Mouton de Gruyter.

Duncker, K. (1935). *Zur Psychologie des produktiven Denkens*. Berlin: Springer.

Enríquez Raído, V. (2014). *Translation and Web Searching*. New York/London: Routledge.

Ericsson, K. A. and Simon, H. A. (1984). *Protocol Analysis*. Cambridge: MIT Press.

Ericsson, K. A. and Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data* (Rev. ed.). Cambridge: MIT Press.

Færch, C. and Kasper, G. (1983). "On Identifying Communication Strategies in Interlanguage Production". In C. Færch and G. Kasper, eds. *Strategies in Interlanguage Communication*. London: Longman, pp. 210-238.

Færch, C. and Kasper, G. eds. (1983). *Strategies in Interlanguage Communication*. London: Longman.

Federico, M., Cattelan, M. and Trombetti, M. (2012). "Measuring User Productivity in Machine Translation Enhanced Computer Assisted Translation". In *Proceedings of the Tenth Conference of the Association for Machine Translation in the Americas (AMTA 2012)*. San Diego, California, 28 October-1 November.

Gaspari, F., Toral, A., Kumar Naskar, S., Groves, D. and Way, A. (2014). "Perception vs Reality: Measuring Machine Translation Post-editing Productivity". In S. O'Brien, M. Simard and L. Specia, eds. *Proceedings of the Third Workshop on Post-editing Technology and Practice (WPTP-3)*. Vancouver, Canada, 26 October 2014, pp. 60-72.

Gerloff, P. (1986). "Second Language Learners' Reports on the Interpretive Process: Talk-aloud Protocols of Translation". In J. House and S. Blum-Kulka, eds. *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr, pp. 243-262.

Göpferich, S. (2009). "Towards a Model of Translation Competence and its Acquisition: the Longitudinal Study TransComp". In A. L. Jakobsen, I. M. Mees and S. Göpferich, eds. *Behind the Mind: Methods, Models and Results in Translation Process Research*. Copenhagen: Samfundslitteratur, pp. 11-37.

Guerberof, A. (2009). "Productivity and Quality in the Post-editing of Outputs from Translation Memories and Machine Translation". *Localisation Focus*. (7)1: 11-21.

Hansen, G. ed. (1998). *LSP Texts and the Process of Translation* [Copenhagen Working Papers 1]. Copenhagen: Copenhagen Business School.

Hansen, G. (2008). "The Dialogue in Translation Process Research". In Translators Associations of China, ed. *Translation and Cultural Diversity. XVIII FIT World Congress*. Shanghai, China, 2-7 August 2008. Beijing: Foreign Languages Press, pp. 386-397.

Hijort-Pedersen, M. and Faber, D. (2009). "Uncertainty in the Cognitive Processing of a Legal Scenario: A Process Study of Student Translators". *Hermes – Journal of Language and Communication Studies*. 42: 189-209.

House, J. and Blum-Kulka, S. eds. (1986) *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr.

Jääskeläinen, R. (1987). "What Happens in a Translation Process: Think-Aloud Protocols of Translation." Unpublished MA Thesis, School of Translation Studies, University of Joensuu, Savonlinna.

Jääskeläinen, R. (1993). "Investigating Translation Strategies". In S. Tirkkonen-Condit and J. Lafling, eds. *Recent Trends in Empirical Translation Research*. Joensuu: University of Joensuu Faculty of Arts, pp. 99-120.

Jääskeläinen, R. (2000). "Focus on Methodology in Think-aloud Studies in Translating". In S. Tirkkonen-Condit and R. Jääskeläinen, eds. *Tapping and Mapping the Processes of Translation and Interpreting: Outlooks on Empirical Research*. Amsterdam: John Benjamins, pp. 71-82.

Jääskeläinen, R. (2010). "Looking for a Working Definition of 'Translation Strategies'". In I. M. Mees, F. Alves and S. Göpferich, eds. *Methodology, Technology and Innovation in Translation Process Research. A Tribute to Arnt Lykke Jakobsen*. Copenhagen: Samfundslitteratur, pp. 375-386.

Jääskeläinen, R. (2011). "Back to Basics. Designing a Study to Determine the Validity and the Reliability of Verbal Report Data on Translation Processes". In S. O'Brien, ed. *Cognitive Explorations of Translation*. London/New York: Continuum, pp. 15-29.

Jakobsen, A. L. (1998). "Logging Time Delay in Translation". In G. Hansen, ed. *LSP Texts and the Process of Translation* [Copenhagen Working Papers 1]. Copenhagen: Copenhagen Business School, pp. 71-101.

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*. Amsterdam: John Benjamins, pp. 69-95.

Jakobsen, A. L. (2005). "Investigating Expert Translators' Processing Knowledge". In H. V. Dam, J. Engberg and H. Gerzymisch-Arbogast, eds. *Knowledge Systems and Translation*. Berlin: Mouton de Gruyter, pp. 173-189.

Jakobsen, A. L., Mees, I. M. and Göpferich, S. eds. (2009). *Behind the Mind: Methods, Models and Results in Translation Process Research*. Copenhagen: Samfundslitteratur.

Kiraly, D. C. (1995). *Pathways to Translation: Pedagogy and Process*. Kent: Kent State University Press.

Koehn, P. (2009). "A Process Study of Computer-aided Translation". *Machine Translation Journal*. 23(4): 241-263.

Koponen, M. (2012). "Comparing Human Perceptions of Post-editing Effort with Post-editing Operations". In *Proceedings of the 7th Workshop on Statistical Machine Translation (WMT 2012)*. Montréal, Canada, 7-8 June 2012, pp. 181-190.

Koponen, M., Ramos, L., Aziz, W., Specia, L. (2012). "Post-editing Time as a Measure of Cognitive Effort". In S. O'Brien, M. Simard and L. Specia, eds. *Proceedings of the AMTA 2012 Workshop on Post-editing Technology and Practice (WPTP 2012)*. San Diego,

California, USA, 28 October 2012, pp. 11-20.

Krings, H. P. (1986). "Translation Problems and Translation Strategies of Advanced German Learners of French (L2)". In J. House and S. Blum-Kulka, eds. *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr, pp. 263-276.

Krings, H. P. (2001). *Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes*. Kent: Kent State University Press.

Kußmaul, P. (1991). "Creativity in the Translation Process. Empirical Approaches". In K. Van Leuven-Zwart and J. Naaijken, eds. *Translation Studies: The State of the Art. Proceedings of the First James S. Holmes Symposium on Translation Studies*. Amsterdam: Rodopi, pp. 91-101.

Lacruz, I., Shreve, G. M. and Angelone, E. (2012). "Average Pause Ratio as an Indicator of Cognitive Effort in Post-editing: A Case Study". In S. O'Brien, M. Simard and L. Specia, eds. *Proceedings of the AMTA 2012 Workshop on Post-editing Technology and Practice (WPTP 2012)*. San Diego, California, USA, 28 October 2012, pp. 21-30.

Lacruz, I. and Shreve, G. M. (2014). "Pauses and Cognitive Effort in Post-editing". In S. O'Brien, L. W. Balling, M. Carl, M. Simard and L. Specia, eds. *Post-editing of Machine Translation: Processes and Applications*. Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 246-272.

Lacruz, I., Denkowski, M. and Lavie, A. (2014). "Cognitive Demand and Cognitive Effort in Post-editing". In S. O'Brien, M. Simard and L. Specia, eds. *Proceedings of the Third Workshop on Post-editing Technology and Practice (WPTP-3)*. Vancouver, Canada, 26 October 2014, pp. 73-84.

Lauffer, S. (2002). "The Translation Process: An Analysis of Observational Methodology". *O processo de tradução*. 2(10): 59-74.

Li, D. (2004). "Trustworthiness of Think-aloud Protocols in the Study of Translation Processes". *International Journal of Applied Linguistics*. 14(3): 301-313.

Lörscher, W. (1991). *Translation Performance, Translation Process, and Translation Strategies: A Psycholinguistic Investigation*. Tübingen: Gunter Narr.

Lörscher, W. (1992). "Process-oriented Research into Translation and Implications for Translation Teaching". *TTR*. V(1): 145-161.

Malone, J. L. (1988). *The Science of Linguistics in the Art of Translation: Some Tools for Linguistics for the Analysis and Practice of Translation*. Albany: State University of New York press.

Matrat, C. M. (1992). *Investigating the Translation Process: Thinking-aloud versus Joint Activity*. PhD Thesis. Ann Arbor: University Microfilms International.

Mees I. M., Alves F. and Göpferich S. eds. (2010). *Methodology, Technology and Innovation in Translation Process Research. A Tribute to Arnt Lykke Jakobsen*. Copenhagen: Samfundslitteratur.

Muñoz Martín, R. (2010). "On Paradigms and Cognitive Translatology". In G. M. Shreve and

- E. Angelone, eds. *Translation and Cognition*. Amsterdam: John Benjamins, pp. 169-187.
- Muñoz Martín, R. (2010). "The Way They Were: Subject Profiling in Translation Process Research". In I. M. Mees, F. Alves and S. Göpferich, eds. *Methodology, Technology and Innovation in Translation Process Research. A Tribute to Arnt Lykke Jakobsen*. Copenhagen: Samfundslitteratur, pp. 87-108.
- O'Brien, S. (2005). "Methodologies for Measuring the Correlations between Post-editing Effort and Machine Translatability". *Machine Translation*. 19(1): 37-58.
- O'Brien, S. (2006). "Pauses as Indicators of Cognitive Effort in Post-editing Machine Translation Output". *Across Languages and Cultures*. 7(1): 1-21.
- O'Brien, S. (2010). "Eye Tracking in Translation Process Research: Methodological Challenges and Solutions". In I. M. Mees, F. Alves and S. Göpferich, eds. *Methodology, Technology and Innovation in Translation Process Research. A Tribute to Arnt Lykke Jakobsen*. Copenhagen: Samfundslitteratur, pp. 251-266.
- O'Brien, S. (2011). *Cognitive Explorations of Translation*. London/New York: Continuum.
- O'Brien, S. and Moorkens, J. (2014). "Towards Intelligent Post-editing Interfaces". In *FIT XXth World Congress 2014*. Berlin, Germany, 4-6 August 2014.
- O'Brien, S., Balling, L. W., Carl, M., Simard, M., Specia, L. eds. (2014). *Post-editing of Machine Translation: Processes and Applications*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Pym, A. (2010). "Text and Risk in Translation". In M. Sidiropoulou and A. Papaconstantinou, eds. *Choice and Difference in Translation. The Specifics of Transfer*. Athens: University of Athens, pp. 27-42.
- Popović, M., Lommel, A., Burchardt, A., Avramidis, E. and Uszkoreit, H. (2014). "Relations between Different Types of Post-editing Operations, Cognitive Effort and Temporal Effort". In M. Tadić, Koehn, P., J. Roturier and A. Way eds. *Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT 14)*. Dubrovnik, Croatia, 16-18 June 2014, pp. 191-198.
- Saldanha, G. and O'Brien, S. (2014). *Research Methodologies in Translation Studies*. New York: Routledge.
- Séguinot, C. ed. (1989). *The Translation Process*. Toronto: H.G. Publications.
- Shreve, G. M. and Angelone, E. eds. (2010). *Translation and Cognition*. Amsterdam: John Benjamins.
- Specia, L. (2011). "Exploiting Objective Annotations for Measuring Translation Post-editing Effort". In M. L. Forcada, H. Depraetere and V. Vandeghinste, eds. *Proceedings of the 15th Conference of the European Association for Machine Translation (EAMT 2011)*. Leuven, Belgium, 30-31 May 2011, pp. 73-80.
- Sun, S. (2011). "Think-aloud-Based Translation Process Research: Some Methodological Considerations". *Meta*. V. 56(4): 928-951.
- Tatsumi, M. and Roturier, J. (2010). "Source Text Characteristics and Technical and

Temporal Post-editing Effort: What is Their Relationship?”. In V. Zhechev, ed. *Proceedings of the Second Joint EM+/CNGL Workshop “Bringing MT to the User: Research on Integrating MT in the Translation Industry” (JEC’10)*. Denver, Colorado, USA, 4 November, pp. 43-51.

Teixeira, C. (2014). “Perceived vs. Measured Performance in the Post-editing of Suggestions from Machine Translation and Translation Memories”. In S. O’Brien, M. Simard and L. Specia, eds. *Proceedings of the Third Workshop on Post-editing Technology and Practice (WPTP-3)*. Vancouver, Canada, 26 October 2014, pp. 45-59.

Temnikova, I. (2010). “Cognitive Evaluation Approach for a Controlled Language Post-editing Experiment”. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner and D. Tapias, eds. *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC’10)*. Valletta, Malta, 17-23 May 2010, pp. 3485-3490.

Tirkkonen-Condit, S. (1989). “Professional vs. Non-professional Translation: a Think-aloud Protocol Study”. In C. Séguinot, ed. *The Translation Process*. Toronto: H.G. Publications, pp. 73-85.

Tirkkonen-Condit, S. ed. (1991). *Empirical Research in Translation and Intercultural Studies: Selected Papers of the TRANSIF Seminar, Savonlinna 1988*. Tübingen: Gunten Narr.

Tirkkonen-Condit, S. and Laffling, J. eds. (1993). *Recent Trends in Empirical Translation Research*. Joensuu: University of Joensuu Faculty of Arts.

Tirkkonen-Condit, S. (2000). “Uncertainty in Translation Processes”. In S. Tirkkonen-Condit and R. Jääskeläinen, eds. *Tapping and Mapping the Processes of Translation and Interpreting: Outlooks on Empirical Research*. Amsterdam: John Benjamins, pp. 123-142.

Tirkkonen-Condit, S. and Jääskeläinen, R. eds. (2000). *Tapping and Mapping the Processes of Translation and Interpreting: Outlooks on Empirical Research*. Amsterdam: John Benjamins.

Toury, G. (1991). “Experimentation in Translation Studies: Achievements, Prospects and some Pitfalls”. In S. Tirkkonen-Condit, ed. *Empirical Research in Translation and Intercultural Studies: Selected Papers of the TRANSIF Seminar, Savonlinna 1988*. Tübingen: Gunten Narr, pp. 45-66.

Translators Associations of China, ed. (2008). *Translation and Cultural Diversity*. Beijing: Foreign Languages Press.

Van Leuven-Zwart, K. and Naaijken, J. eds. (1991). *Translation Studies: The State of the Art. Proceedings of the First James S. Holmes Symposium on Translation Studies*. Amsterdam, 17-19 December 1990. Amsterdam: Rodopi.

Vilar, D., Xu, J., D’Haro, L., Ney, H. (2006). “Error Analysis of Statistical Machine Translation Output”. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC’06)*. Genoa, Italy, 22-28 May 2006, pp. 697-702.

APPENDIX A

Information and Instructions Provided to the Participants Working within the CAT Setting

TRADUZIONE DI COMUNICATI STAMPA IN AMBIENTE CAT

INFORMAZIONI UTILI (*da leggere prima di iniziare l'esperimento*)

- 1) L'esperimento prevede la traduzione in ambiente CAT dall'inglese all'italiano di un estratto di comunicato stampa di UPS.
- 2) Il testo da tradurre contiene circa 100 parole e riguarda alcuni dati economici dell'azienda.
- 3) In alcuni casi che saranno precisati più sotto, ti sarà chiesto di pensare a voce alta, indicando le fonti consultate, la correttezza o meno dei risultati ottenuti da tali fonti e i gradi/i tipi di interventi necessari per correggere eventuali errori o imprecisioni.
- 4) Il comunicato stampa è un testo utilizzato da molte aziende per far conoscere le proprie iniziative, le novità riguardanti i prodotti e alcuni significativi dati economici. Si tratta di un importante strumento di marketing che contribuisce alla creazione dell'immagine che l'azienda vuole dare di sé.
- 5) In questo esperimento, le tue riflessioni ad alta voce verranno registrate con un registratore. L'obiettivo è capire l'utilità di determinate risorse messe a disposizione del traduttore.
- 6) Gli unici strumenti a cui potrai ricorrere nella fase di traduzione sono la TM e Internet.
- 7) La TM fornita contiene XXX translation units. La percentuale minima di corrispondenza impostata per l'individuazione di un fuzzy match è del 75%.
- 8) Il tempo necessario per tradurre il testo sarà cronometrato.
- 9) Non c'è nessun limite di tempo.
- 10) La qualità del tuo lavoro non sarà valutata.
- 11) In fase di creazione del progetto di traduzione, ti verrà chiesto di selezionare "Prepare" nella finestra "Project Preparation" perché, in una situazione in cui un'azienda chieda a più di un traduttore di tradurre i comunicati stampa e di popolare di volta in volta una TM unica, ogni traduttore potrebbe voler creare e collegare anche una TM di progetto per mantenere separate le proprie proposte di traduzione (della cui validità è sicuro) dalle altre che gli vengono fornite con la TM unica.

ISTRUZIONI (*da leggere prima di iniziare l'esperimento*)

- 1) Crea un progetto di traduzione con SDL Trados Studio 2011, collegando il file source e la translation memory che ti saranno forniti. Non è previsto l'uso né di termbases né di un dizionario di AutoSuggest. Nella finestra "Project Preparation", seleziona "Prepare" in modo che venga creata e aggiornata anche una memoria di progetto.

- 2) Il testo finale deve essere di alta qualità.
- 3) Dopo aver verificato che entrambe le TM siano collegate e che la casella “Update” sia spuntata, inizia a tradurre il file nella Editor view.
- 4) Durante la traduzione, può accadere che:
 - la TM collegata individui un 100% match o un context match della cui correttezza sei sicura, oppure un fuzzy match per cui riconosci subito le modifiche da fare senza consultare altre risorse;
 - la TM individui un fuzzy match con traduzioni che vuoi verificare, o non fornisca nessuna soluzione. In questo caso, utilizza la Concordance Search per le tue ricerche. Puoi utilizzarla per i termini, le collocazioni, la fraseologia e ogni ulteriore elemento del segmento che stai traducendo e, in questi casi, specifica sempre di quale/i elemento/i del testo si tratta. Rifletti ad alta voce anche sull'utilità di questo strumento.
- 5) Se non ottieni risultati o non sei soddisfatto di quelli forniti, ricorri a Internet per le ricerche. Inizia sempre a riflettere ad alta voce, ripetendo per quale elemento del segmento source stai facendo le ricerche in rete (es: “earnings per share”) e, successivamente, ragiona ad alta voce sull'utilità e la correttezza dei risultati eventualmente ottenuti dai siti Internet consultati.
- 6) Puoi tornare sullo stesso segmento più volte e fare tutte le modifiche che ritieni necessarie; devi però specificarlo (dicendo, per esempio: “earnings per share: ulteriore modifica”).
- 7) Non è necessario che tu segua l'ordine dei segmenti così come appare nella Editor view: puoi saltare da un segmento a un altro.
- 8) Se sei soddisfatto della traduzione di un segmento, ricordati di validarlo, in modo che le memorie vengano aggiornate e ti possano eventualmente fornire dei match per gli altri segmenti del testo.
- 9) Prima di concludere la traduzione, valida tutti i segmenti in modo che vengano inseriti nella TM collegata.
- 10) Traduci per intero il testo che ti è stato consegnato.
- 11) Quando ritieni concluso la tua traduzione, avverti la sperimentatrice.
- 12) Consegna il file bilingue sdlxliff, il file target, la TM di progetto (che hai creato e aggiornato) e la TM che ti è stata fornita (e che hai aggiornato).

APPENDIX B

Information and Instructions Provided to the Participants **Working within the PE Setting**

POST-EDITING DI OUTPUT GREZZI DI GOOGLE TRANSLATE

INFORMAZIONI UTILI (da leggere prima di iniziare l'esperimento)

- 1) L'esperimento prevede il post-editing dell'output grezzo ottenuto da Google Translate dopo aver fatto tradurre a questo sistema di traduzione automatica un comunicato stampa di UPS dall'inglese all'italiano. Il post-editing dovrà quindi essere fatto in italiano.
- 2) Il testo su cui fare il post-editing contiene 100 parole circa e riguarda alcuni dati economici dell'azienda.
- 3) In alcuni casi che saranno precisati più sotto, ti sarà chiesto di riflettere a voce alta, spiegando le fonti consultate per le tue ricerche, la correttezza o meno dei risultati ottenuti da tali fonti e i gradi/i tipi di interventi necessari per correggere eventuali errori o imprecisioni dell'output.
- 4) Il comunicato stampa è un testo utilizzato da molte aziende per far conoscere le proprie iniziative, le novità riguardanti i prodotti e alcuni significativi dati economici. Si tratta di un importante strumento di marketing che contribuisce alla creazione dell'immagine che l'azienda vuole dare di sé.
- 5) In questo esperimento, le riflessioni a voce alta verranno registrate con un registratore. L'obiettivo è capire l'utilità di determinate risorse messe a disposizione del traduttore.
- 6) Oltre all'output grezzo fornito come punto di partenza della traduzione, potrai servirti solo di Internet per le tue ricerche.
- 7) Il tempo necessario per fare il post-editing del testo sarà cronometrato.
- 8) Non c'è nessun limite di tempo.
- 9) La qualità del tuo lavoro non sarà valutata.

ISTRUZIONI (da leggere prima di iniziare l'esperimento)

- 1) Considerata la funzione che svolgono i comunicati stampa, il post-editing dell'output che ti viene fornito deve essere completo: il testo finale deve essere di alta qualità.
- 2) Per fare il post-editing utilizza la funzione "Revisioni" di Word.
- 3) Se, durante il post-editing:
 - sei certo della correttezza della proposta ottenuta da Google Translate e decidi di mantenerla nel testo, non è necessario che tu rifletta ad alta voce;
 - non sei certo della correttezza della proposta di Google Translate e decidi di fare delle

ricerche in rete per verificare la sua validità, inizia a riflettere ad alta voce, specificando qual è la porzione di testo interessata, quali siti consulti per la tua verifica e il grado di utilità dei risultati che ottieni;

- consideri errata o imprecisa la proposta di Google Translate e decidi di eliminarla o modificarla in modo sostanziale, rifletti ad alta voce sull'errore che hai individuato, sulle fonti in rete di cui ti servi e sui risultati che ti forniscono.

4) Inizia sempre la tua riflessione ad alta voce dicendo per quale elemento del testo source farai delle ricerche (es: “earnings per share”).

5) Puoi tornare sulla stessa porzione di testo più volte e fare tutte le modifiche che ritieni necessarie; devi però specificarlo (dicendo, per esempio: “earnings per share: ulteriore modifica”).

6) Non è necessario che tu segua l'ordine dei paragrafi nel testo: puoi saltare da un paragrafo a un altro.

7) Il post-editing deve essere fatto su tutto il testo.

8) Quando ritieni concluso il tuo lavoro di post-editing, avverti la sperimentatrice.

9) Dovrai consegnare il file word contenente l'output con le revisioni.

APPENDIX C

Source Text to Be Translated with the Help of the TM Software by the Participant of the First Pair

UPS Boosts Dividend by 10 Percent

Board Cites Earnings Outlook, Strong Cash Flow

The UPS (NYSE: UPS) Board of Directors today increased the regular quarterly dividend by 9.6 percent to \$0.57 per share from \$0.52 on all outstanding Class A and Class B shares. The dividend is payable March 7, 2012, to shareholders of record on Feb. 21, 2012.

"UPS turned in a great performance in 2011 despite a volatile global operating environment," said UPS Chairman and CEO Scott Davis. "Cash flow in 2012 is expected to be strong and clearly today's decision by the Board reflects that projection."

APPENDIX D

Source Text to Be Translated with the Help of the TM Software by the Participant of the Second Pair

UPS Board Declares Dividend

The UPS (NYSE: UPS) Board of Directors today declared a regular quarterly dividend of \$0.52 per share on all outstanding Class A and Class B shares.

The dividend is payable June 1, 2011, to shareholders of record on May 16, 2011.

Earlier this year, the UPS Board boosted the regular quarterly dividend by 11% to the current level of \$0.52 per share. UPS's dividend has more than tripled since 2000, when it stood at \$0.17 per share. The company has either increased or maintained its dividend every year for more than four decades.

APPENDIX E

Source Text to Be Translated with the Help of the TM Software by the Participant of the Third Pair

UPS Board Boosts Dividend by 11 Percent to \$0.52 Per Share

Directors Cite Strong Earnings Outlook

The UPS (NYSE: UPS) Board of Directors today increased the regular quarterly dividend by 11% to \$0.52 per share from \$0.47 on all outstanding Class A and Class B shares. The dividend is payable March 2, 2011, to shareholders of record on Feb. 14, 2011.

"We believe that 2011 is going to be a great year for UPS and we're committed to significantly increasing distributions to shareowners," said UPS Chairman and CEO Scott Davis. "Cash flow is expected to be strong and clearly today's decision by the Board reflects that projection."

APPENDIX F

Raw Output to Be Post-edited by the Participant of the First Pair

UPS Aumenta dividendo del 10 per cento

Consiglio Cites guadagni Outlook, Forte Cash Flow

L'UPS (NYSE: UPS) Consiglio di Amministrazione ha aumentato oggi il dividendo trimestrale regolare del 9,6 per cento a 0,57 dollari per azione da 0,52 dollari su tutte le classi in circolazione azioni di classe A e B. Il dividendo è pagabile 7 marzo 2012, agli azionisti registrati il 21 febbraio 2012.

"UPS ha disputato una grande prestazione nel 2011, nonostante un contesto globale volatile," ha dichiarato UPS Chairman e CEO Scott Davis. "Il flusso di cassa nel 2012 dovrebbe essere forte e chiaramente la decisione odierna del Consiglio che riflette la proiezione."

APPENDIX G

Raw Output to Be Post-edited by the Participant of the Second Pair

Consiglio UPS dichiara dividendo

L'UPS (NYSE: UPS) Consiglio di Amministrazione ha dichiarato oggi un dividendo trimestrale regolare di \$ 0,52 per azione su tutte le classi in circolazione azioni di classe A e B.

Il dividendo è pagabile 1 Giugno 2011, agli azionisti registrati il 16 maggio 2011.

All'inizio di quest'anno, il Consiglio UPS incrementato il dividendo trimestrale regolare del 11% al livello attuale di \$ 0,52 per azione. Dividendo di UPS è più che triplicato dal 2000, quando era pari a 0,17 dollari per azione. La società ha aumentato o mantenuto il suo dividendo ogni anno per più di quattro decenni.

APPENDIX H

Raw Output to Be Post-edited by the Participant of the Third Pair

Consiglio UPS Aumenta dividendo del 11 per cento a \$ 0,52 per azione

Amministrazione Cite Guadagni Forte Outlook

L'UPS (NYSE: UPS) Consiglio di Amministrazione ha aumentato oggi il dividendo trimestrale regolare del 11% a \$ 0,52 per azione da 0,47 dollari su tutte le classi in circolazione azioni di classe A e B. Il dividendo è pagabile 2 marzo 2011, agli azionisti registrati il 14 febbraio 2011.

"Crediamo che il 2011 sta per essere un grande anno per i gruppi di continuità e ci siamo impegnati ad aumentare in modo significativo le distribuzioni di azionisti", ha detto l'UPS presidente e CEO Scott Davis. "Il flusso di cassa dovrebbe essere forte e chiaramente la decisione odierna del Consiglio che riflette la proiezione."

APPENDIX I

Translation Delivered by the Participant of the First Pair

UPS aumenta i dividendi del 10 per cento

Il Consiglio di Amministrazione annuncia le previsioni degli utili e il forte flusso di cassa

Oggi il Consiglio di Amministrazione di UPS (NYSE: UPS) ha aumentato il dividendo trimestrale ordinario del 9,6% passando da 0,52\$ per azione a 0,57\$ sulle azioni in circolazione di classe A e classe B. Il dividendo è dovuto al 7 marzo 2012 per gli azioni registrati al 21 febbraio 2012.

"UPS ha ottenuto importanti risultati nel 2011 nonostante la volatilità del contesto globale ", ha affermato Scott Davis, Presidente e CEO di UPS. "Si ritiene che nel 2012 il cash flow sarà più forte ed evidentemente le decisioni prese oggi dal Consiglio riflettono questa previsione".

APPENDIX J

Translation Delivered by the Participant of the Second Pair

Il Consiglio di Amministrazione di UPS comunica i dividendi

Il Consiglio di Amministrazione di UPS ha confermato un dividendo trimestrale ordinario di 0.52\$ per azione sulle azioni in circolazione di classe A e classe B.

Il dividendo è pagabile al 1 giugno 2011 per gli azionisti registrati al 16 maggio 2011.

In precedenza, il Consiglio di UPS aumenta il dividendo trimestrale ordinario dell'11% al livello attuale di 0,52\$ per azione. Il dividendo di UPS è più che triplicato dal 2000, quando si attestava allo 0,17\$ per azione. L'azienda ha aumentato o mantenuto costante il proprio dividendo ogni anno per oltre quattro decenni.

APPENDIX K

Translation Delivered by the Participant of the Third Pair

Il Consiglio di Amministrazione di UPS aumenta i dividendi dell'11 per cento a 0,52\$ per azione.

Il Consiglio di Amministrazione annuncia un incremento nelle previsioni degli utili

Il Consiglio di Amministrazione di UPS (NYSE: UPS) oggi ha aumentato il dividendo trimestrale ordinario dell'11% passando da 0,52\$ per azione a 0,47\$ sulle azioni in circolazione di classe A e classe B. Il dividendo è dovuto al 2 marzo 2011 per gli azionisti registrati al 14 febbraio 2011.

"Crediamo che il 2011 sarà un anno importante per UPS, e ci impegneremo per aumentare significativamente i dividendi per gli azionisti", ha affermato Scott Davis, Presidente e CEO di UPS. "Si ritiene che il cash flow sarà più forte ed evidentemente le decisioni prese oggi dal Consiglio riflettono questa previsione".

APPENDIX L

Post-edited Text Delivered by the Participant of the First Pair

UPS: dividendo trimestrale in aumento del 10 per cento

Annunciate le previsioni sui guadagni, forte flusso di cassa

Il Consiglio di Amministrazione di UPS (NYSE: UPS) ha aumentato oggi il dividendo trimestrale regolare del 9,6 per cento a 0,57 dollari per azione da 0,52 dollari su tutte le azioni di classe a e b in circolazione. Il dividendo sarà pagabile il 7 marzo 2012 agli azionisti che risultano alla chiusura delle attività il 21 febbraio 2012.

"UPS ha registrato una performance positiva nel 2011, nonostante un contesto globale instabile," ha dichiarato Scott Davis, presidente e CEO di UPS. "Prevediamo di registrare un forte flusso di cassa e chiaramente la decisione odierna del Consiglio riflette tale prospettiva."

APPENDIX M

Post-edited Text Delivered by the Participant of the Second Pair

Il Consiglio di UPS dichiara il proprio dividendo

Oggi il Consiglio di Amministrazione di UPS (NYSE: UPS) ha dichiarato un dividendo trimestrale regolare di 0,52 dollari per azione su tutte le azioni di Classe A e B rimaste in sospeso.

Il dividendo è pagabile a partire dall'1 Giugno 2011, agli azionisti registrati il 16 maggio 2011.

All'inizio di quest'anno, il Consiglio di UPS ha incrementato il dividendo trimestrale regolare dell' 11% al livello attuale di 0,52 dollari per azione. Il dividendo di UPS è più che triplicato dal 2000, quando si attestava a 0,17 dollari per azione. La società ha aumentato o mantenuto stabile il proprio dividendo ogni anno per più di quattro decenni.

APPENDIX N

Post-edited Text Delivered by the Participant of the Third Pair

Il Consiglio di Amministrazione di UPS aumenta il dividendo dell'11% portando il guadagno di ogni azione a \$ 0,52

L'amministrazione dell'azienda prevede una crescita robusta nei prossimi mesi

Il Consiglio di Amministrazione di UPS (NYSE: UPS) ha aumentato oggi il dividendo trimestrale regolare dell'11% portando il guadagno di ogni azione a \$ 0,52 (dai precedenti 0,47 \$) su tutte le azioni in circolazione di categoria A e B. Il dividendo verrà distribuito il 2 marzo 2011 agli azionisti registrati il 14 febbraio 2011.

"Siamo convinti che il 2011 sarà un anno importante per UPS e ci impegniamo ad aumentare in modo significativo i dividendi degli azionisti", ha detto il presidente di UPS nonché CEO, Scott Davis. "Il fatturato dovrebbe aumentare e senza dubbio la decisione odierna del Consiglio sembra riflette tale proiezione."

APPENDIX O

Rankings Obtained by the Participant of the First Pair Working within the CAT Setting¹⁵

- 4- Equivalent retrieval
- 1- Comprehension of the target-language term
- 3- Contextualisation
- 2- Equivalent monitoring

APPENDIX P

Rankings Obtained by the Participant of the Second Pair Working within the CAT Setting

- 3-Equivalent monitoring
- 4-Equivalent retrieval
- 2-Comprehension of the target-language term
- 1-Comprehension of the source-language term

APPENDIX Q

Rankings Obtained by the Participant of the Third Pair Working within the CAT Setting

- 3-Equivalent monitoring
- 4-Comprehension of the target-language term
- 2-Contextualisation
- 1-Equivalent retrieval

¹⁵ In appendices O-T, strategies are put in order of frequency. The number assigned to each of them corresponds to the ranking given by the participant and indicating perceived difficulty.

APPENDIX R

Rankings Obtained by the Participant of the First Pair Working within the PE Setting

- 4- Contextualisation
- 3- Equivalent retrieval
- 2- Equivalent monitoring
- 1- Comprehension of the source-language term

APPENDIX S

Rankings Obtained by the Participant of the Second Pair Working within the PE Setting

- 4- Contextualisation
- 2-Comprehension of the source-language term
- 3-Equivalent retrieval
- 1-Comprehension of the target-language term

APPENDIX T

Rankings Obtained by the Participant of the Third Pair Working within the PE Setting

- 4- Contextualisation
- 1-Comprehension of the source-language term
- 2-Equivalent monitoring
- 3-Equivalent retrieval