Trainee translators' use of digital technologies: An exploratory-descriptive study based on key-logging

Laura BRUNO - lbruno@unc.edu.ar
Faculty of Languages, National University of Cordoba, Argentina
Paula ESTRELLA - paula.estrella@unc.edu.ar
Faculty of Mathematics, Astronomy, Physics and Computer Science, National University of Cordoba, Argentina

Abstract

This paper presents an exploratory-descriptive study that investigates the patterns in the user activity during translation evaluations with free digital technologies such as online resources, a CAT tool and a word processor. The study bases its analyses on a combination of qualitative and quantitative data collected by key-logging in a natural learning context. These were random samples made from evaluations in technical translation of undergraduate students. A wide range of data including keyboard logs, mouse logs, time used, no-key activity and screenshots was analysed. The findings demonstrate that all participants fulfilled the translation evaluations by carrying out the basic functions of a free CAT tool within which we considered the source text loading, the translation of segments, the tag management, the glossary usage, the dictionary usage, and the target text generation. Most of the trainee translators reproduced the visual aspect of the source text in the target text. The amount of time the students allocated to the usage of a free CAT tool, a free word processor and free online resources for documentation and terminology management led to time differences among them. The results show that after undergoing a specific gradual training, pre-trained translators can complete a whole computer-assisted translation process during evaluation sessions.

Keywords: translation, gradual training, free CAT tools, digital technologies, key-logging

1. Introduction

The research underpinning this paper derives from a master thesis in Translation Studies that was carried out at the Faculty of Languages, National University of Cordoba, Argentina. In such work, the use of digital technologies, including online resources, CAT tools and word processors, has been the phenomenon under study due to the recognised importance within the training in specialized translation. The role played by translation aids has always been crucial since they are the almost-compulsory condition to solve a translation assignment in order to ensure acceptable levels of quality, at least in terms of specialized terminology or text genres. Above all, this may be one of the reasons that makes the difference when the trainee translator is informed about the current technologies, the appropriate ways to use them, and the specific translation process phases during which they have to be used. However, the patterns that emerge when trainee translators process their translations vary in many respects. Some of the factors involved can be source text comprehension, target text production, translation skills, source text difficulty, time pressure, and even instrumental skills.

This paper presents and discusses the results of an exploratory-descriptive study designed to analyse how four trainee translators used digital technologies, on the basis of key-logging data collected from translations carried out during evaluation sessions. Thus, the following research questions were raised and answered: (i) How is the user interaction with the basic functions of the free CAT tool? (ii) How is the physical structure of the target text compared to the source text? (iii) How much time does the usage of digital technologies take?

1.1. The CAT tool usage

OmegaT is a free and open source multiplatform Computer Assisted Translation (CAT) tool (Smolej, 2019) based on translation memory (TM) technology. First, participants were trained in how to use this technology. Afterwards, the user interaction with the basic functions of OmegaT was explored. The process of data gathering was conducted during evaluation sessions in the natural learning context of trainee translators. The user interaction required the user to perform six actions previously trained: (i) loading the source text (ST) to the Source Folder of the TM; (ii) translating the file one segment at a time; (iii) validating the tags before generating the translated file; (iv) activating the glossary and the dictionary; and finally, (v) generating the target text (TT), and (vi) recovering it from the /target/ subfolder of the TM.

1.2. The visual presentation

The visual aspect of the target text refers to the time span beginning with the opening of the translated text in the word processor LibreOffice, and ending with the revision of inconsistencies between ST and TT regarding the physical structure of texts or what Mossop (2014) defines as the visual aspect of texts. Among many other highlights of translation memories, it needs to be mentioned that the formatting information present in the source file is reproduced in the target file. However, it is common for trainee translators to be tempted to

amend the presentation by making unnecessary changes or to be even unaware of transferred errors from the TM (O'Brien et al., 2017) affecting the visual aspect of the text, as for example, at the following levels: layout (e.g. spacing, bulleted lists); typography (e.g. font type, font size, font effect); and organization (e.g. without traces of the source text, title layout).

1.3. The time distribution

Prior to our study, in an empirical study with a sample of 18 professional translators, Hvelplund (2017) observes that resource consultation takes almost 20 per cent of the time of the overall translation task, and concludes by noting the importance of digital resources on the translation process. In our study, time distribution refers to the amount of time allocated by trainee translators during the translation process to the usage of digital technologies, including the CAT tool, the word processor and the online resources.

2. Background

A summary of key concepts across the literature is provided in this section. First of all, we define the digital technologies in order to avoid misconceptions as various definitions exist for the term. Secondly, we describe the framework for the integration of the digital technologies into the didactics of the subject Technical Translation. Finally, we deal with the methodological approach employed which allowed the research questions to be answered.

2.1. Digital technologies

For the purpose of this paper, we will follow the definition of digital technologies proposed by Berrocoso et al. (2010), who describe them, by opposition to traditional technologies, as follows: (i) versatile because they are usable in different ways; (ii) unstable in the sense that they change rapidly with time, (iii) and opaque due to the fact that their internal functioning is hidden to the user.

In our study, we distinguish two broad categories of digital technologies: (i) tools or computer programs and (ii) online resources that trainee translators used to produce their translations. Table 1 below presents the digital technologies classified according to production stages.

Table 1: Categories of digital technologies

Process	Tools	Resources
Documentation management	browsers	metasearch and search engines
Terminology management	terminology tools	online terminological resources
Translation	CAT tools word processors	online linguistic resources

2.1.1. The translation tools

As it is well known, a CAT tool is a computer program used to translate texts and it is usually known as TM or translation memory. During the study, trainee translators used the translation memory OmegaT, a free open source application written in Java which works on a wide variety of software platforms, e.g. Windows, Mac, and Linux. This TM stores the translations and proposes possible translations from similar segments registered in the translation memory files thanks to its fuzzy matching feature. The user interface contains 6 panes among which there is an Editor pane where the user types and edits the translation while the Match pane displays the most similar segments from the translation memory and the Glossary pane displays the terms found for items in the segments to be translated. Once trainee translators were ready to view the final product, they exported the translated files, open them in the word processor LibreOffice and viewed the translation in the final format.

2.1.2. The online resources

As regards the online consultation, we distinguish three groups of resources: (i) information resources (i.e. metasearch and search engines) to solve thematic problems, (ii) terminological resources to solve terminological problems, and (iii) online linguistic resources. By way of example, Table 2 below presents the resources classified according to the phases of the terminology management.

Table 2: Phases of terminology management

Phases	Resources
Search for semantic information (L1)	English monolingual specialized resources (encyclopaedias, dictionaries, books, vocabularies)
Search for equivalent terms	English-Spanish bilingual specialized resources (technical glossaries and dictionaries, term banks)
Search for semantic information (L2)	Spanish monolingual specialized resources (encyclopaedias, dictionaries, books, vocabularies)
Search for pragmatic information	Monolingual or bilingual specialized resources (parallel corpuses)

2.2. The instrumental competences in translation competence training

Competence-based training is a recent pedagogical trend that stems from the evolution of previous models such as the learning goals training. In 2015, Hurtado Albir recognizes its theoretical foundations in social constructivism (Kiraly, 2000) and moves towards an integrated model of teaching, learning and evaluation in translation. Based on PACTE's models of translation competence and translation competence acquisition, Hurtado Albir redefines the

learning objectives and proposes six categories of competences in translation training that are applicable to all learning levels, types and modalities.

In our study, we explore the instrumental competences as defined by Hurtado (2015, p.262) as "Instrumental competences: managing documentary resources and an array of tools to solve translation problems". We also found that Gamero Pérez and Hurtado Albir (1999, p.142) propose objectives for teaching specialized translation that are directly related to technologies, and are formulated by the authors as follows:

Master the tools for the technical and scientific translator:
- Know and know how to use documentation sources
- Know how to use electronic specialized dictionaries
- Know how to access and work on the Internet
- Know how to use computer applications for translation
- Develop a critical spirit and know how to evaluate the sources consulted.

Focusing on this direction, we designed learning objectives and contents to guide the gradual training with digital technologies.

2.3. Key-logging and the user interaction

Although there are different techniques used to study the cognitive processes during the elaboration of a translation, key-logging is considered a non-intrusive technique, which allows saving every mouse and keyboard keystroke made by the translator in real time.

These collected data are known as User Activity Data and the most commonly used programs for recording this type of data in the field of Translation Studies are Translog (Carl, 2012) and InputLog (Leijten, 2013). In 2015, Lafuente developed an open source computational tool called ResearchLogger, to provide a free alternative but with the same features offered by its competitors. The tool was tested in a controlled environment research, which was focused on the study of some cognitive aspects in translation students at the National University of Cordoba. The results showed that the use of a translation memory complicated decision-making, which was reflected in increased pauses made by the participant. Also, under time pressure, participants reduced the time spent on orientation and revision phases.

Hvelplund (2017) conducted an empirical study on professional use of digital resources during the translation process. The researcher worked with eye tracking data and screen recording data from 18 professional translators. The findings demonstrated that consultation of digital resources accounted for a considerable amount (about 20%) of the total time spent by translators on a translation task.

Therefore, in order to gain insight into how trainee translators use digital technologies when translating and how digital technologies are integrated into the didactics of the subject Technical Translation, our methodological basis rests on some translation process researches previously reported in the literature, where the focus is on user interaction with technologies mainly by using key-logging.

3. Research design

The methodology of our study included data collection, tabulation and interpretation of key-logging data (e.g. screen, keyboard and mouse recording) as well as target file data. The user interaction with the CAT tool was explored by analysing the patterns of usage of the translation tool according to a required basic level of expertise. The visual aspect of the translated text was investigated by examining the procedures carried out by participants once the resulting translation was extracted from the TM. The time distribution was explored by examining the amount of time allocated to the usage of digital technologies throughout the translation process.

3.1. Participants and data gathering

The participants constitute a random sample of 4 trainee translators attending the subject Technical Translation (English-Spanish) in the Translation Training Course at the Faculty of Languages, National University of Cordoba, Argentina. These undergraduate students had no previous experience in the use of digital technologies for translating. The primary source of data gathering at our disposal were evaluations on translation provided by translation files, log files, full screenshots and screenshots around the mouse pointer. At this point, it is worth mentioning that the small sampling number responds to the exploratory-descriptive nature of the study design, time-space limitations, and the sizeable processing of a huge volume of data.

3.2. Recording software

Data of the translation process were collected during evaluation sessions by ResearchLogger (Estrella et al., 2017). This data-gathering tool is an open source keylogger that non-intrusively records in real time every keyboard and mouse activity made by the trainee translator when using several computer programs either in Linux or Windows. Unlike other existing tools, ResearchLogger provides logs, screenshots, and a high level of data granularity coming from online digital resources, CAT tools and word processors. In this way, it is possible to reconstruct, step by step, the procedures carried out by the participants, identifying patterns when solving arising problems, levels of expertise at which they can use technologies (i.e. basic functions or advanced functions), work preferences (e.g. always using the mouse instead of keyboard shortcuts), time length required for using one technology or another, or interleaving patterns between technologies used. The latest developments of the tool include the partial automation of results into a spreadsheet containing partial analysis of the logs (Grijalba et al., 2018).

4. Findings and analysis

In this section we present the results obtained from the analysis carried out to explore the user interaction with a CAT tool, the visual aspect of the target text and the amount of time allocated to the usage of digital technologies.

4.1. User interaction with CAT tool

Table 3 below summarises the results obtained from the exploration of the user interaction with the basic functions of OmegaT, the free CAT tool with which participants were trained prior to evaluation sessions.

All trainee translators generating the collected data started their translation process by performing the same functions of opening the TM and loading the ST. Three of the four participants translated all the untranslated segments, except for a sample where only one segment (a subtitle of the ST, to be specific) was left untranslated. This unmanaged function may be interpreted as a decision taken by the trainee translator, possibly due to the similarity with the following segment (first sentence of the paragraph) in the text and the intention of joining both segments. The resulting unnecessary omission is likely to be related to what many authors have referred to as the constraints imposed by translation memory systems regarding the lack of a global vision of the text (Gil & Pym, 2006). Regarding tags, half of the participants accomplished the tool management while the other half showed different patterns of usage: namely, deliberate omission of tags in the Editor pane and no tag validation at all.

In reviewing the exploration of the glossary usage, the majority of the samples allowed the visualization of terms in the Glossary pane, which indicated that the users created the glossary file and updated it manually for use in OmegaT during the evaluation sessions. Unfortunately, there was no data available to account for the non-usage of a glossary in one of the samples despite 67 full screenshots and 1127 screenshots around the mouse pointer taken with ResearchLogger being explored. The exploration of the dictionary usage revealed that only one participant managed to have StarDict, a general dictionary, activated in the dictionary pane. We might infer that the remaining participants could have disregarded its usage due to the usefulness of such dictionary for translating technical texts.

Finally, results showed that all trainee translators generating the collected data translated all the segments in the Editor pane, and performed the functions of generating and then retrieving the TT from the /target/ subfolder of the TM.

Table 3: User interaction with CAT tool
(S=sample; managed=1; unmanaged=0; no data available=NA)

Basic functions	S1	S2	S3	S4
Source text loading	1	1	1	1
Translation of segments	1	0	1	1
Tag management	1	0	1	0
Glossary usage	1	NA	1	1
Dictionary usage	NA	0	1	0
Target text generation	1	1	1	1

4.2. Visual aspect of target text

Table 4 below summarises the results obtained from the exploration of the visual aspect of the target text after recovering the translated file from the /target/ subfolder of the TM. The majority of the participants maintained most of the aspects of the visual presentation of the ST in the TT. This decision can be interpreted as a decision that attempts to preserve the superstructure that corresponds to technical genres.

Table 4: Visual aspect of target text
(managed=1; unmanaged=0; no data available=NA)

	S1	S2	S3	S4
Spacing	1	1	1	0
Bulleted lists	0	1	1	1
Font type	1	1	1	1
Font size	1	1	0	1
Font effect	1	1	1	1
Traces of Source text	1	1	0	1
Titles layout	0	1	0	1

4.3. Time distribution

Table 5 below summarises the results obtained from the exploration of the time allocated to the usage of digital technologies during the trainees' translation process. There were significant differences in the amount of time that participants spent when using different digital technologies during their translation process. The time allocated to CAT tool use constituted half of the translation process duration, while the use of the word processor demanded the minimum amount of time. This can be attributed to the high amount of time that trainee translators dedicate to producing a target text with the quality of a final draft when using a CAT tool. Finally, the online resource consultation also showed time differences among participants, even though the amount of time allocated by the majority of the participants was lower than the one for CAT tool use.

Table 6 below lists the online resources used in translations, the number of trainees that accessed to those resources and the total number of consultations.

Table 5: Time allocated to digital technologies usage (in minutes)

	S1	S2	S3	S4
Duration of translation process	31'	42'	50'	69'
OmegaT	15,27'	21,70'	23,95'	19,74'
LibreOffice	3,99'	7,32'	3,34'	17,24'
Online Resources	8,72'	12,96'	21,55'	29,00'

No keyboard activity	3,00'	0'	1,16'	3,00'

Table 6: Types of online resources, number of samples and number of consultations

Resource type	Samples	Consultations
ENGLISH MONOLINGUAL DICTIONARIES		
- Oxford Dictionaries	1	1
ENGLISH-SPANISH BILINGUAL DICTIONARIES, TERM BANKS, OTHER		
- WordReference Dictionary	2	7
- IATE	3	9
- TERMIUM PLUS	1	1
- ProZ	1	1
- Linguee	3	16
ENCYCLOPAEDIAS / ENGLISH MONOLINGUAL DICTIONARIES		
- Wikipedia	2	3
SEARCH ENGINES		
- Google	4	10
- Google images	1	1
TOTAL	4	50

5. Conclusions

The research described in this paper has been centered on a key-logging study designed to explore the usage of digital technologies during the translation process of four trainee translators.

Among the first conclusions to be drawn, we can mention the clear-cut description obtained from the process performed by the pre-trained students in a natural learning context. This was mainly possible due to the adoption of a mixed-method design which included the use of ResearchLogger.

The overall findings focused on the research questions conclude that all participants did a complete computer-assisted translation process after undergoing a specific gradual training with digital technologies. In particular, we can say that all the students pre-trained in the usage of OmegaT fulfilled the translation evaluation, despite some minor problems that did not hamper them to move forward onto the next stage, i.e. the revision of the visual aspect of the target text. Regarding the exploration of the time allocated to the usage of digital technologies, the results were conclusive, and showed time differences among participants, as initially expected. The CAT tool use was allocated approximately half of the time of the total duration of the translation

process, whereas LibreOffice registered the minimum times in all cases. Finally, the online resource consultation times ranged between 28 and 43 per cent in the overall translation process. This is worthy of consideration since students produce their own technology interleaving along the translation process.

In the light of what has been previously said, we conclude that the research findings provide us with formal understanding of the phenomenon studied which leads us to innovative didactic solutions for the integration of digital technologies in the subject Technical Translation. To consolidate the new knowledge on the technology-mediated translation process of trainee translators, our future efforts should focus on the improvement of the competence-based training by (i) the development of a set of learning objectives and contents exclusively related to programs, such as word processors and browsers, so that students optimize usage and consultation times by applying successful strategies; and (ii) the incorporation of an evaluation system with rubrics to assess the use of digital technologies.

For future research purposes, the same study may certainly shed further and in-depth data on the person-computer interaction if other sources of data extraction are added like eye-tracking and retrospective thinking aloud techniques.

Referencias bibliográficas

Berrocoso, J. V., Arroyo, M. D. C. G., & Sánchez, R. F. (2010). Enseñar y aprender con tecnologías: un modelo teórico para las buenas prácticas con TIC. *Teoría de la Educación. Educación y Cultura en la Sociedad de la Información*, *11*(1), 203-229.

Carl, M. (2012). Translog-II: A program for recording user activity data for empirical reading and writing research. *Proceedings of the Eighth International Conference on Language Resources and Evaluation,* 4108-4112. Istanbul, Turkey: Department of International Language Studies and Computational Linguistics.

Estrella, P., Lafuente, R. & Bruno, L. (2017). Broadening the scope of translation process research with ResearchLogger. *Proceedings of the 10th Leipzig International Conference on Translation & Interpretation Studies Translation 4.0 – Translation in the digital age (LICTRA),* 51-52. Leipzig, Germany: Institute of Applied Linguistics and Translatology.

Gamero Pérez, S., & Hurtado Albir, A. (1999). La enseñanza de la traducción especializada. In: Hurtado, A. (Ed). *Enseñar a traducir. Metodología en la formación de traductores e intérpretes,* 139-195. Madrid: Edelsa

Gil, J. R. B., & Pym, A. (2006). Technology and translation (a pedagogical overview). In Pym, A., Perekrestenko A., Starink, B. (eds.), *Translation Technology and its Teaching, Intercultural Studies Group*, Universitat Rovira i Virgili, Tarragona.

Grijalba, W. A., Castillo, J. A., Campos, R. A., & Estrella, P. (2018). Process-based assessment of computer science students. *2018 XIII Latin American Conference on Learning Technologies (LACLO),* 363-370. IEEE.

Hurtado Albir, A. (2015). The acquisition of translation competence. Competences, tasks, and assessment in translator training. *Meta: Journal des traducteurs/Meta: Translators' Journal*, *60*(2), 256-280.

Hvelplund, K. T. (2017). Translators' use of digital resources during translation. *HERMES-Journal of Language and Communication in Business*, (56), 71-87.

Kiraly, D. (2000). *A social constructivist approach to translator education; Empowerment from theory to practice*. Manchester: St. Jerome Publishing.

Lafuente, R. A. (2015). *Keylogging para el estudio de los procesos cognitivos del traductor* [Bachelor's thesis]. National University of Cordoba, Argentina. https://rdu.unc.edu.ar/handle/11086/2826?locale-attribute=es

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, *30*(3), 358-392.

Mossop, B. (2014). *Revising and editing for translators*. Oxon & New York: Routledge.

O'Brien, S., Ehrensberger-Dow, M., Hasler, M., & Connolly, M. (2017). Irritating CAT tool features that matter to translators. *Hermes: Journal of Language and Communication in Business, 56*, 145-162.

PACTE (2003). Building a translation competence model. In Alves, Fabio (ed.). *Triangulating translation: Perspectives in Process Oriented Research*, 43-66. Amsterdam: John Benjamins

Smolej, V. (n.d.). *OmegaT - Guía de usuario*. Retrieved May 21, 2022, from https://omegat.sourceforge.io/manual-standard/es/index.html