

# On post-editing effort: Testing translator perceptions

Translating Europe Workshop

“The changing profile of the translator profession: technology, competences and fit-for-market training”

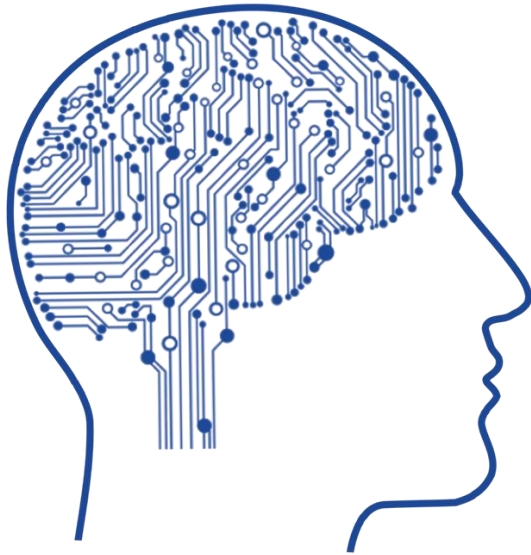
23-24 May 2018

Vilelmini Sosoni, Maria Stasimioti, Katia-Lida Kermanidis  
Ionian University



# Introduction

Machine Translation (MT) has been increasingly used in industrial translation production scenarios thanks to the development of neural machine translation (NMT) models and the improvement of MT output, especially at the level of fluency (Castilho et al., 2017). In particular, in an effort to speed up the translation process and reduce costs, MT output is used as raw translation to be subsequently post-edited by translators (Gaspari et al., 2015; Lommel and DePalma, 2016).



Companies and researchers report an increase in productivity when using post-editing (PE) compared to regular human translation (Aranberri, Labaka, et al. 2014; Plitt and Masselot 2010), while at the same time still delivering products of comparable quality (Garcia 2010; O'Curran 2014).

However, despite its growing popularity, PE still involves a number of open challenges, such as the development of fair pricing schemes and the effort experienced by translators during PE (Vieira, 2014: 188).

Translators remain sceptical towards PE and question its real benefits (Gaspari, 2014; Koponen, 2012).

# PE and Cognitive Processes

PE has been found to differ from both human translation and revision of human translation in terms of the cognitive processes and the practical goals and methods employed (Krings, 2001; O'Brien, 2002).

**Cognitive processes** refer to 'cognitive effort' or 'cognitive load' (Krings 2001, p. 179) and are taken to involve a combination of cognitive, technical and temporal effort measured usually with the use of eye tracking and keystroke logging data (Vieira, 2014).

Cognitive effort can be studied via gaze data, building on the eye-mind hypothesis by Just and Carpenter (1980). An increased number of fixations (Doherty, O'Brien, et al. 2010) or higher average fixation durations (Carl, Dragsted, Elming, et al. 2011) have been used as indicators of increased cognitive effort.

In eye tracking studies carried out by O'Brien (2007), Sharmin, Šparkov et al. (2008), Doherty, O'Brien et al. (2010) Carl, Dragsted, Elming et al. (2011), Koglin (2015), Nitzke and Oster (2016) and Daems, Vandepitte et al (2017), **postediting of RBMT or SMT output was found to be less cognitively demanding than human translation.**

In addition to the cognitive effort, **translators' attitudes and perceptions matter as well**, as it is important for translators to feel content and productive. Studies show that translators' attitudes towards MT and PE are mixed (see Guerberof, 2013) or negative (see Gaspari, Toral, et al. 2014; Moorkens and O'Brien, 2015). PE is perceived either as more effortful and time-consuming than human translation, or tedious and lacking creativity.

# The Aim

The present study reports on an attempt at tracing the cognitive processes involved in the translation and PE of NMT output and at evaluating the translators' perceptions of the effort required to carry out PE.

## The Research Questions



1. Does the effort experienced by translators for full PE of NMT output differ from the effort experienced for manual translation?
2. Does the actual post-editing effort match the users' perception of the effort and time required to carry out full PE of NMT output?

# Methodology and Experimental procedure



# Method – Answering the Research Questions

A **case study** involving **12 Greek professional translators** and focusing on the **English-Greek language pair** was carried out in March 2018.

Participants were asked to carry out two translation tasks and two PE tasks at the HUBIC Lab (<http://www.hubic-lab.eu/>) at the Athena Research Centre in Athens, Greece.

**Eye-tracking and keystroke logging** were used in combination with **questionnaires** in order to investigate **cognitive effort during human translation and PE** as well as **translators' perception of PE**.

# Call for participation

- **Call for participation:** A call for participation was sent to the members of the two biggest Greek Associations of professional translators, i.e. the Panhellenic Association of Translators (PEM) and Panhellenic Association of Professional Translation Graduates of the Ionian University (*PEEMPIP*) and was shared on social media.
- **Expression of interest:** Potential participants expressed their interest for participating in this study by filling in a Google form.
- **Instructions and guidelines:** They received an e-mail with details on the aim of the research and guidelines for each task along with some educational material.

# Questionnaires

✓ 12 professional translators



## Two Questionnaires:

- **Questionnaire A (profile and experience):** 34 questions. 22 closed-ended questions (multiple choice) and 12 open-ended questions.  
Filled in before the experiment.
- **Questionnaire B (feedback on translation and PE tasks):** 15 questions. 13 closed-ended questions (multiple choice) and 2 open-ended questions.  
Filled in after the experiment.

# Materials

- The source texts (STs) used in this study were short educational texts selected from OER Commons (<https://www.oercommons.org/>), which is a public digital library of open educational resources.
- Six 142/161-word excerpts were selected from various courses on Business Administration and Social Change. They all had comparable Lexile<sup>®</sup> scores (between 1300L and 1400L), i.e. they were suitable for 11th/12th graders. The titles of the courses were retained as context information for translators.
- The NMT was taken from Google Translate (output obtained March 24).

# Automatic metrics

Text	BLEU score
Text 3	51.33
Text 4	60.62

The Bilingual Evaluation Understudy Score (BLEU) is a metric for evaluating a machine-translated (generated) sentence to a reference sentence.

The approach works by counting matching n-grams in the candidate translation to n-grams in the reference text, where 1-gram or unigram would be each token and a bigram comparison would be each word pair. The comparison is made regardless of word order.

**The higher the score, the more and the larger n-grams are shared between the reference and the machine translation. A BLEU score offers more of an intuitive rather than an absolute meaning and is best used for relative judgments.**

# Guidelines

## ➤ Translation task:

Since all the participants in this study were professional translators, the only guideline provided to them was to produce a text with the same *skopos* as that of the original text and of publishable quality.

## ➤ Post-editing task:

Since not all of the participants had previous training and experience in post-editing, they received brief training on PE before executing this task.

The training included **a video on post-editing, a presentation, as well as specific PE guidelines that the participants were instructed to follow for reasons of consistency.**

The training material was sent to them five days before the execution of the tasks.

# PE Guidelines

**Guideline 1:** Retain as much raw MT translation/output as possible.

**Guideline 2:** The message transferred should be accurate.

**Guideline 3:** Fix any omissions and/or additions (at the level of sentence, phrase or word).

**Guideline 4:** Correct mistranslations.

**Guideline 5:** Correct morphological errors.

**Guideline 6:** Correct misspellings and typos.

**Guideline 7:** Fix incorrect punctuation if it interferes with the message.

**Guideline 8:** Correct wrong terminology.

**Guideline 9:** Fix inconsistent use of terms.

**Guideline 10:** Do not introduce stylistic changes.

# General guidelines and limitations/restrictions

- Your hair should not block your eyes.
- Do not wear mascara.
- Avoid touching your eyes (e.g. rubbing your eyes, removing and wearing eyeglasses, etc.).
- During the translation and PE tasks, look exclusively at the computer screen in front of you.
- Try to keep your head as steady as possible.
- External resources cannot be used.

A detailed consent form was signed by all participants, while all stored data were fully anonymized in accordance with Law 2472/97 (as amended by Laws 3783/2009, 3917/2011 and 4070/2012).

# Procedure

- The experiment consisted of one session for each participant. All sessions took place in March 2018.
- A warm-up task was completed for human translation before the translation task and a warm-up task for PE before the PE task, so that participants could get used to the environment, the tools, and the different types of tasks.
- The actual experimental tasks involved the translation of two texts and the PE of two texts.
- A combination of keystroke logging and eye tracking tools were used to register the translation and PE processes.

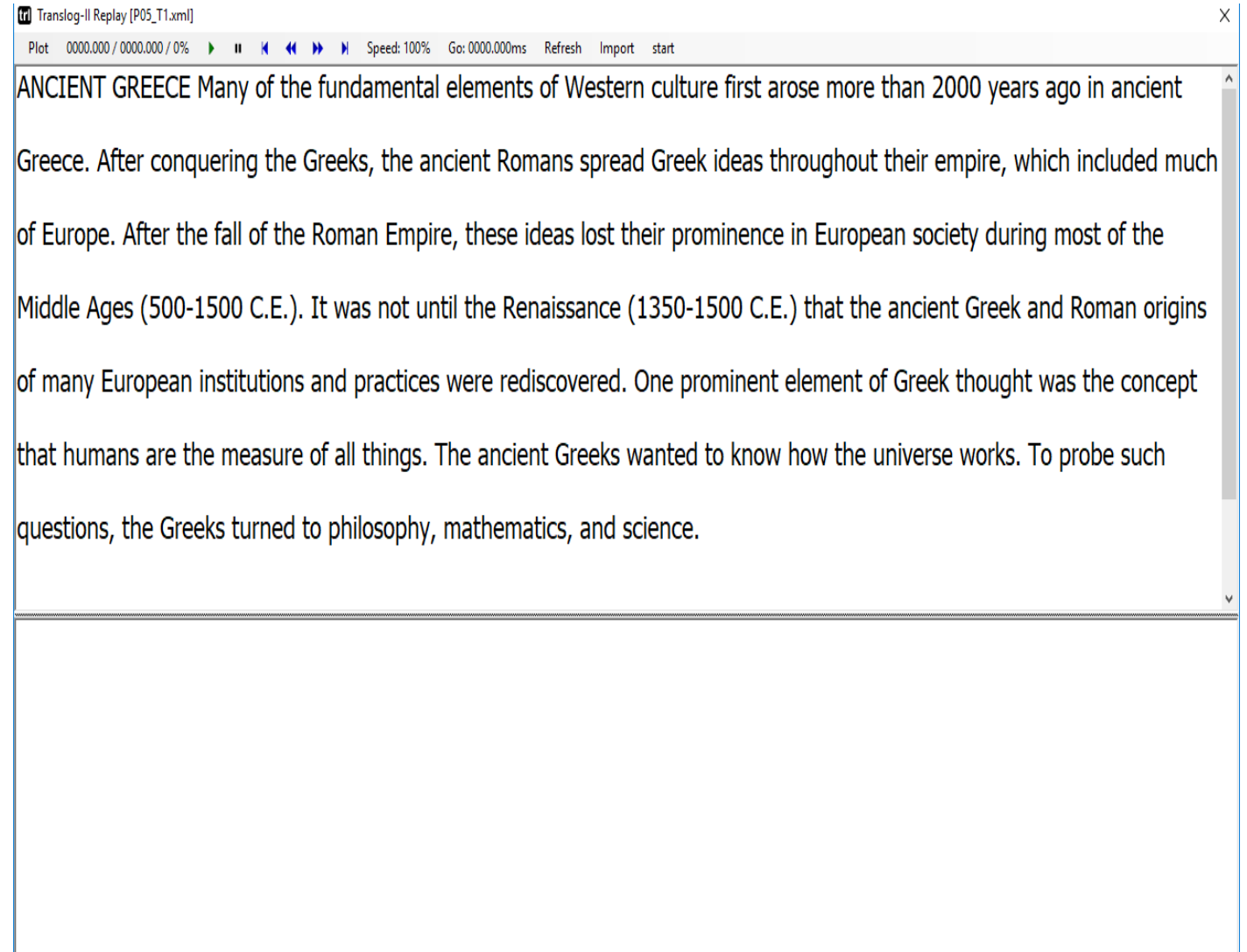
# Eye-tracking and Keystroke logging Apparatus

- A Tobii TX-300 eye tracker at the HUBIC Lab of the Athena Research Center in Athens was used to register the participants' eye movements.
- The texts were displayed in 17-point Tahoma font and double spacing on a Tobii TX Display (23") at 1920 x 1080 pixels. The average viewing distance aimed at was 50-60 cm from the screen.
- The software used as a translation and PE environment was Translog-II software. Translog II is a Windows-oriented program to record and study reading and writing processes on a computer. More specifically, Translog-II records user activity data (UAD), i.e. all the keystrokes (insertions, deletions, navigation, copy/cut-and-paste, return key and/or mouse operations) and gaze movements. The information is stored in an XML format and can be replayed or analyzed with Translog-II or analyzed in an external tool (Carl, 2012).

# Translation

This task was a traditional human translation assignment.

Participants were asked to provide their translation in a split-screen window. The ST was displayed at the top half of the screen and the translation at the bottom half.



The screenshot shows a software window titled "Translog-II Replay [P05\_T1.xml]". The window has a toolbar with various playback controls (play, stop, previous, next, speed) and a status bar showing "Plot 0000.000 / 0000.000 / 0%", "Speed: 100%", "Go: 0000.000ms", "Refresh", "Import", and "start". The main content area displays a text passage about Ancient Greece. The text is as follows:

ANCIENT GREECE Many of the fundamental elements of Western culture first arose more than 2000 years ago in ancient Greece. After conquering the Greeks, the ancient Romans spread Greek ideas throughout their empire, which included much of Europe. After the fall of the Roman Empire, these ideas lost their prominence in European society during most of the Middle Ages (500-1500 C.E.). It was not until the Renaissance (1350-1500 C.E.) that the ancient Greek and Roman origins of many European institutions and practices were rediscovered. One prominent element of Greek thought was the concept that humans are the measure of all things. The ancient Greeks wanted to know how the universe works. To probe such questions, the Greeks turned to philosophy, mathematics, and science.

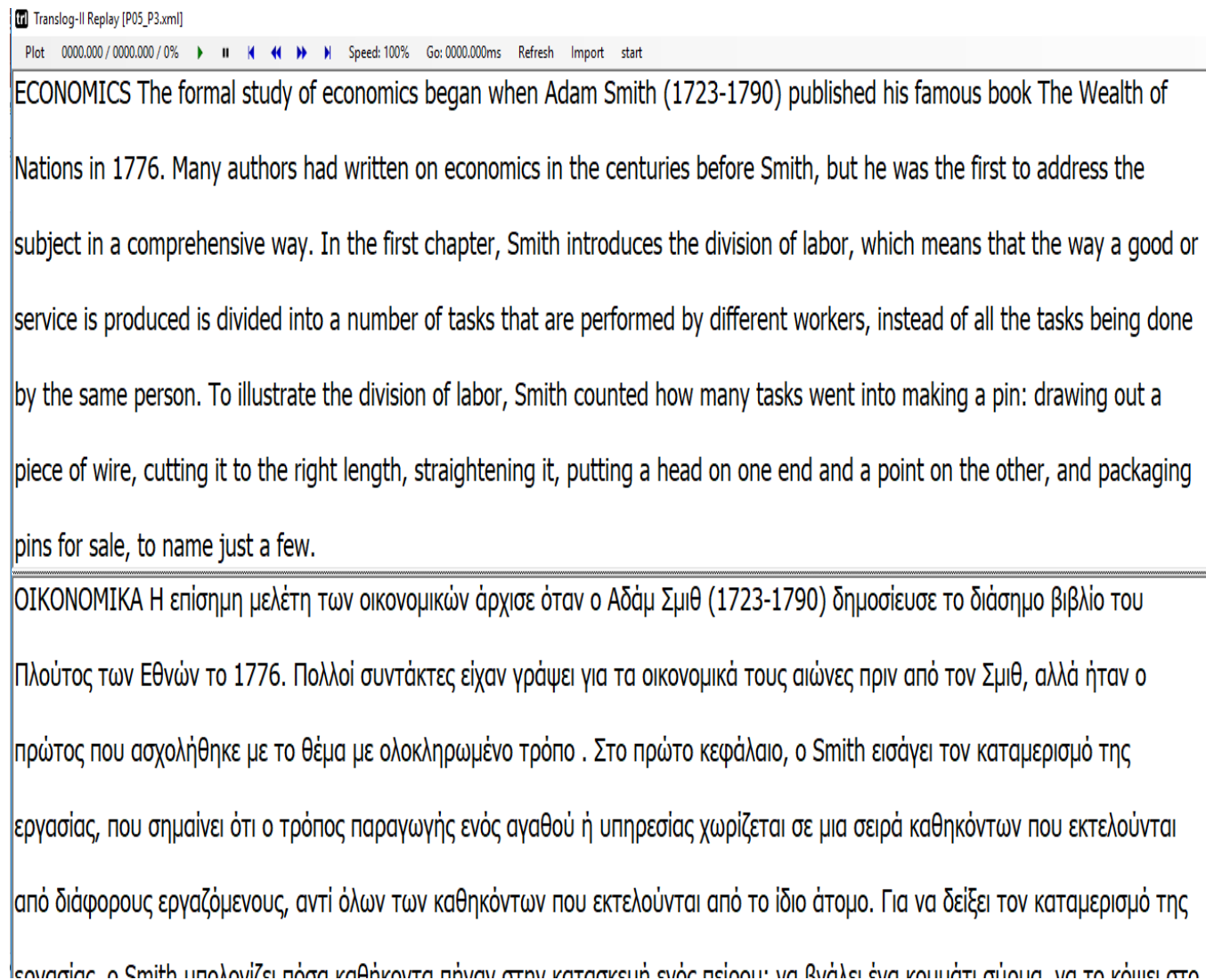
The bottom half of the window is currently empty, representing the area for the participant's translation.

# Post-editing NMT output

This task was a traditional PE assignment.

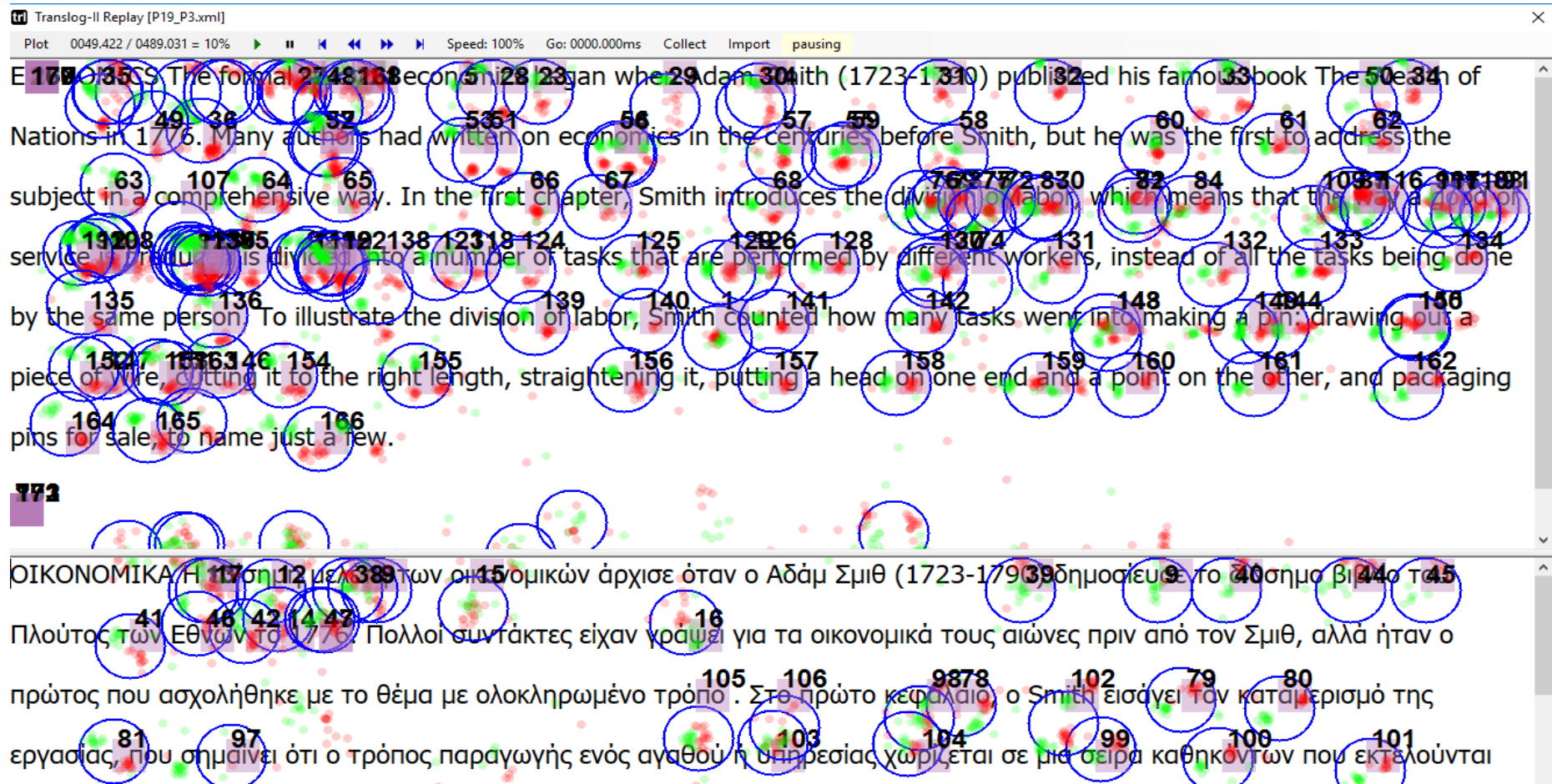
Participants were asked to post-edit the raw output generated by the NMT engine.

The ST was displayed at the top half of the screen and the translation at the bottom half. Translators worked directly on the translation.



The screenshot shows a software interface titled "Translog-II Replay [P05\_P3.xml]". At the top, there is a control bar with a "Plot" indicator showing "0000.000 / 0000.000 / 0%", playback controls (play, stop, back, forward), "Speed: 100%", "Go: 0000.000ms", and buttons for "Refresh", "Import", and "start". The main content area is split into two horizontal panels. The top panel displays the English source text: "ECONOMICS The formal study of economics began when Adam Smith (1723-1790) published his famous book The Wealth of Nations in 1776. Many authors had written on economics in the centuries before Smith, but he was the first to address the subject in a comprehensive way. In the first chapter, Smith introduces the division of labor, which means that the way a good or service is produced is divided into a number of tasks that are performed by different workers, instead of all the tasks being done by the same person. To illustrate the division of labor, Smith counted how many tasks went into making a pin: drawing out a piece of wire, cutting it to the right length, straightening it, putting a head on one end and a point on the other, and packaging pins for sale, to name just a few." The bottom panel displays the Greek translation: "ΟΙΚΟΝΟΜΙΚΑ Η επίσημη μελέτη των οικονομικών άρχισε όταν ο Αδάμ Σμιθ (1723-1790) δημοσίευσε το διάσημο βιβλίο του Πλούτος των Εθνών το 1776. Πολλοί συντάκτες είχαν γράψει για τα οικονομικά τους αιώνες πριν από τον Σμιθ, αλλά ήταν ο πρώτος που ασχολήθηκε με το θέμα με ολοκληρωμένο τρόπο . Στο πρώτο κεφάλαιο, ο Smith εισάγει τον καταμερισμό της εργασίας, που σημαίνει ότι ο τρόπος παραγωγής ενός αγαθού ή υπηρεσίας χωρίζεται σε μια σειρά καθηκόντων που εκτελούνται από διάφορους εργαζόμενους, αντί όλων των καθηκόντων που εκτελούνται από το ίδιο άτομο. Για να δείξει τον καταμερισμό της εργασίας, ο Smith υπολογίζει πόσα καθήκοντα πήγαν στην κατασκευή ενός πείρου: να βγάλει ένα κομμάτι σύρμα, να το κόψει στο"

# Translog II





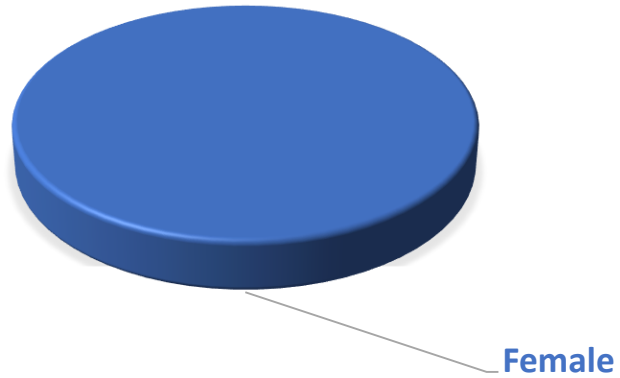
## The Participants

**All participants had normal or corrected to normal vision.**

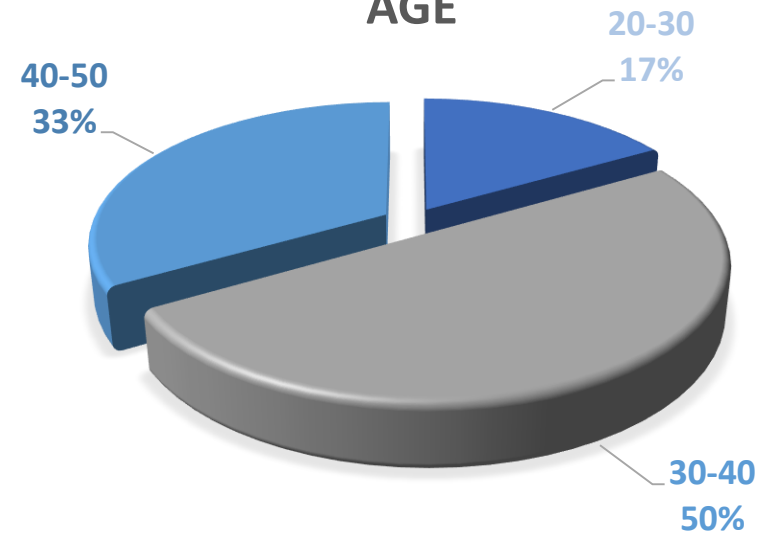
**Two wore contact lenses and one wore glasses, yet the calibration with the eye tracker was successful for all three.**

# Demographics/Participants' profile

GENDER

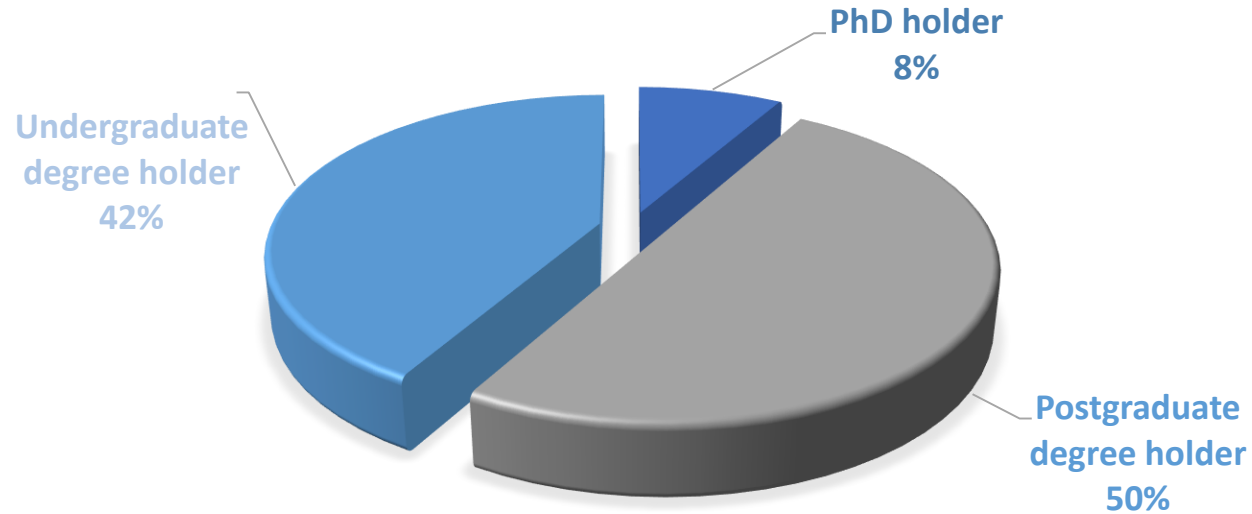


AGE

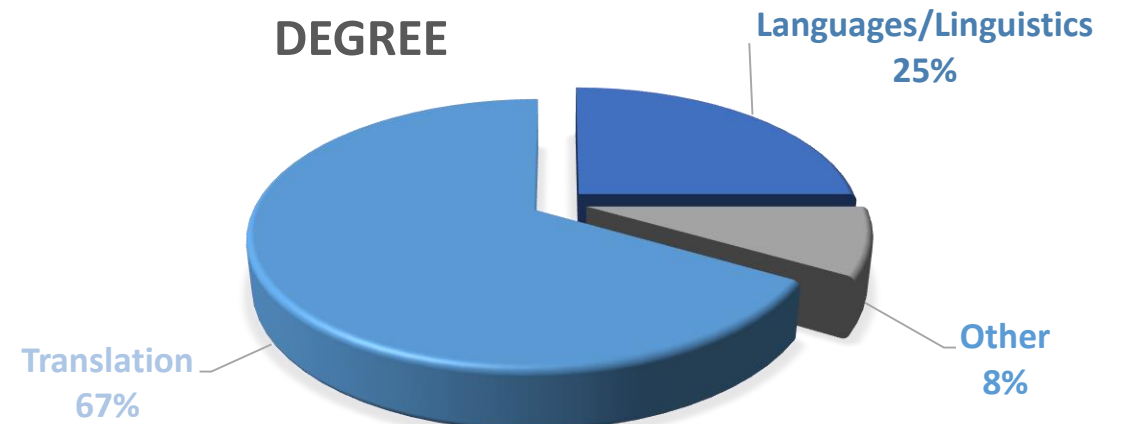


# Demographics/Participants' profile

EDUCATION LEVEL

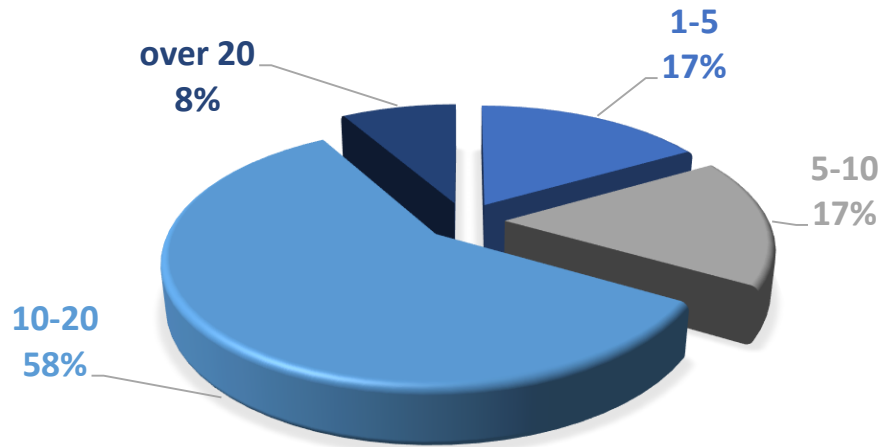


DEGREE

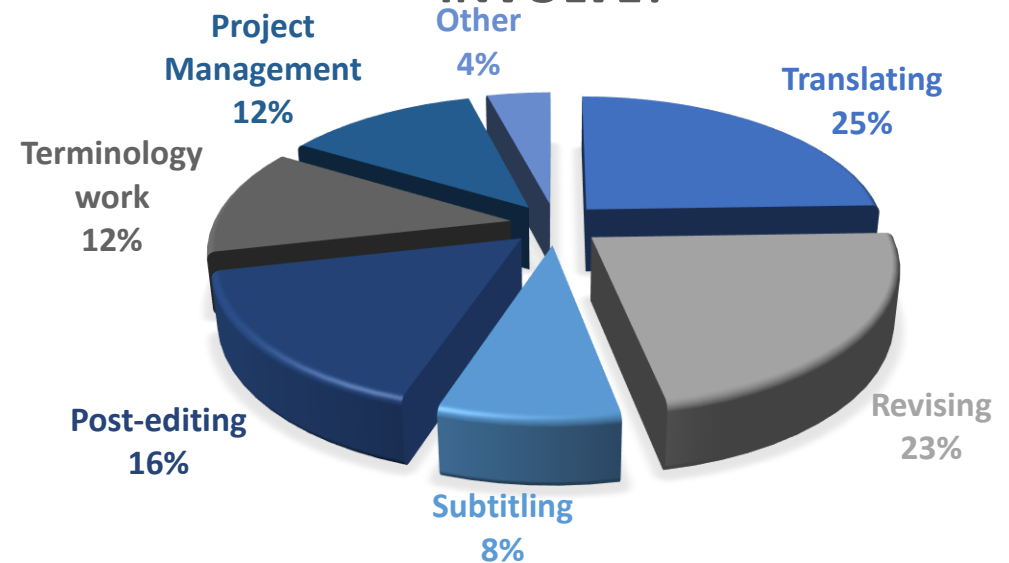


# Demographics/Participants' profile

YEARS OF EXPERIENCE IN TRANSLATION

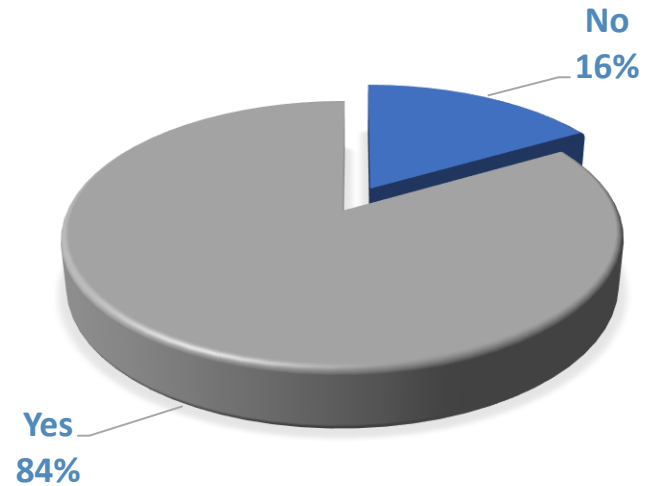


WHICH TASKS DOES YOUR WORK INVOLVE?

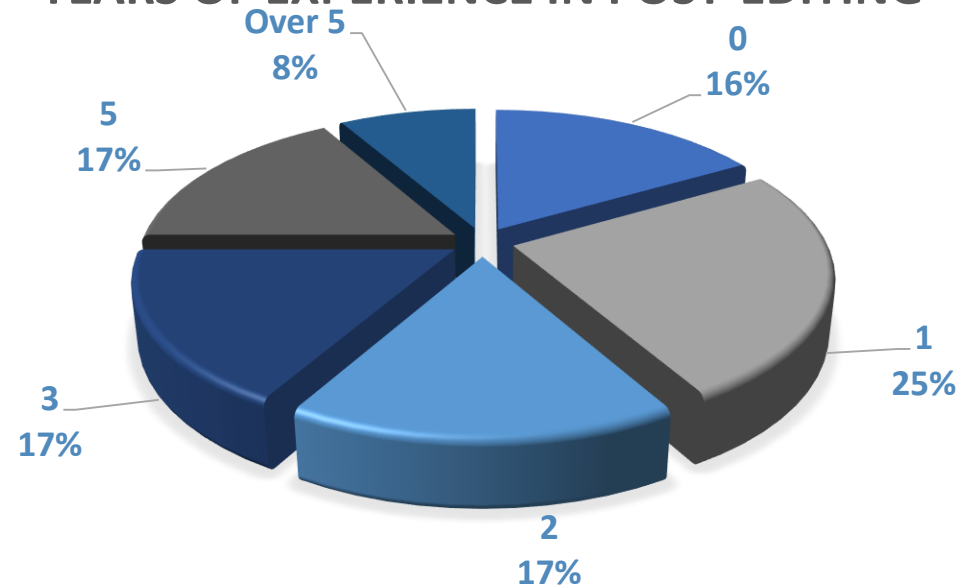


# Demographics/Participants' profile

DO YOU HAVE EXPERIENCE IN POST-EDITING?

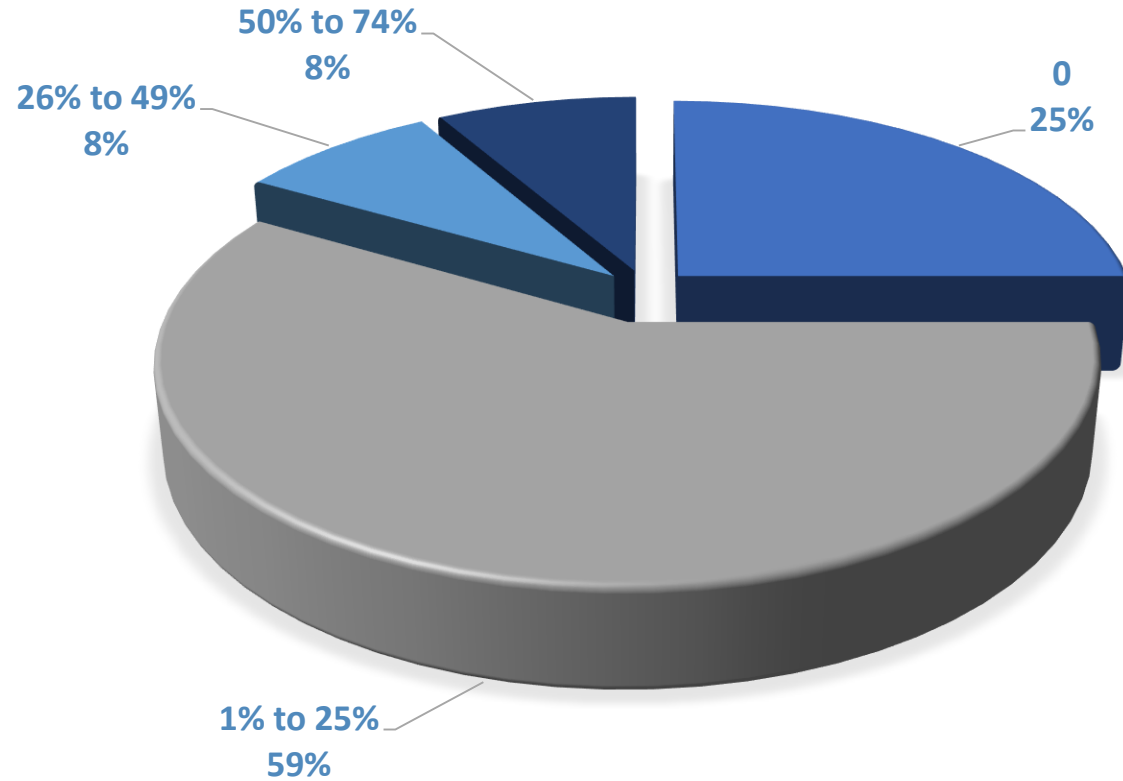


YEARS OF EXPERIENCE IN POST-EDITING



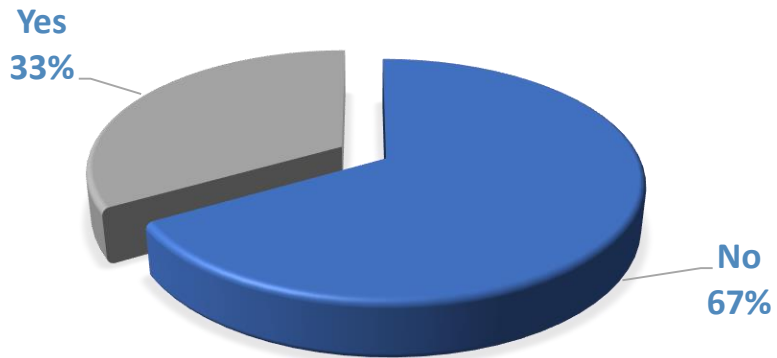
# Demographics/Participants' profile

WHAT PERCENTAGE DOES POST-EDITING MT OUTPUT REPRESENT IN YOUR WORK?

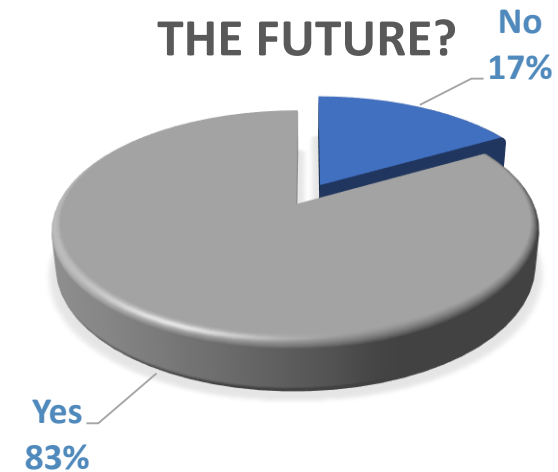


# Demographics/Participants' profile

HAVE YOU RECEIVED TRAINING IN POST-EDITING?

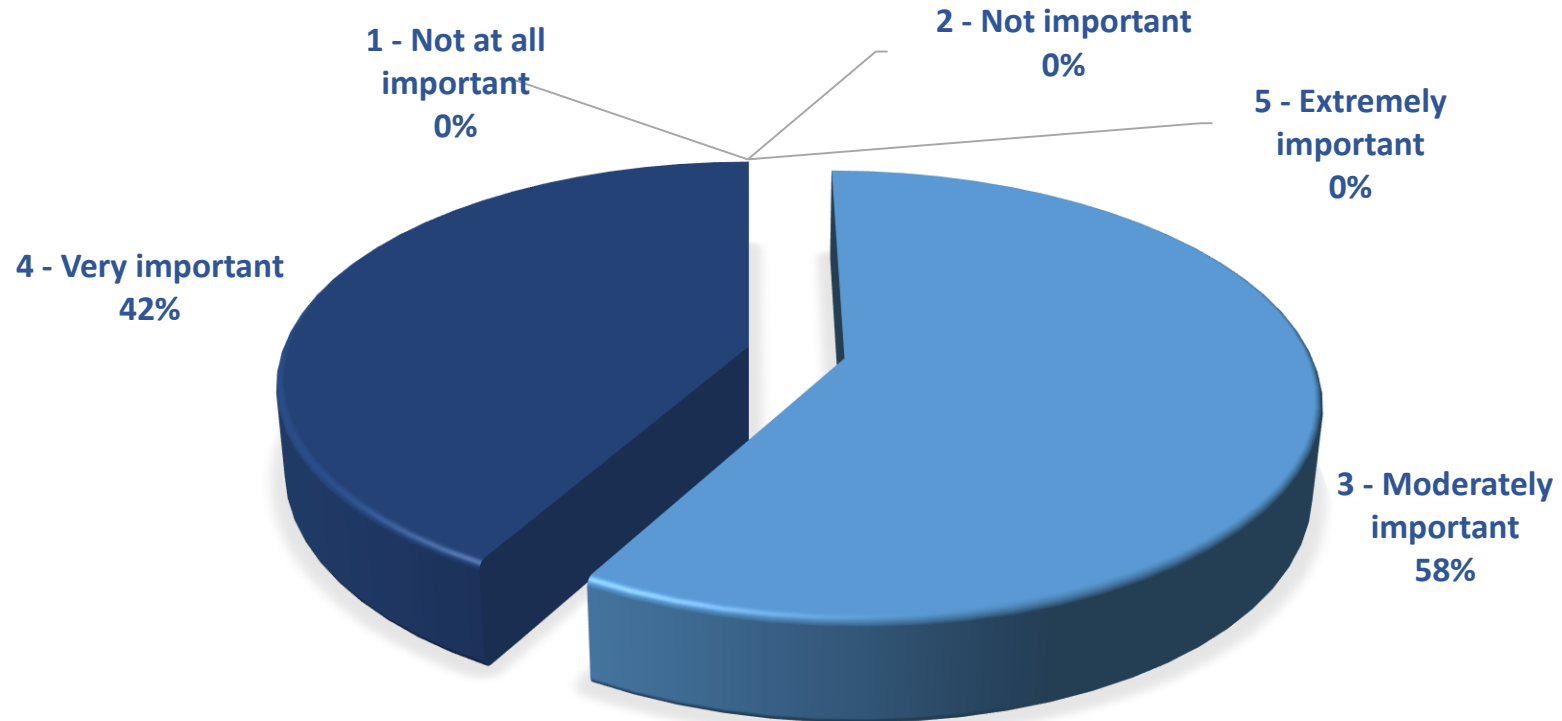


WOULD YOU BE INTERESTED IN ATTENDING A POST-EDITING COURSE IN THE FUTURE?



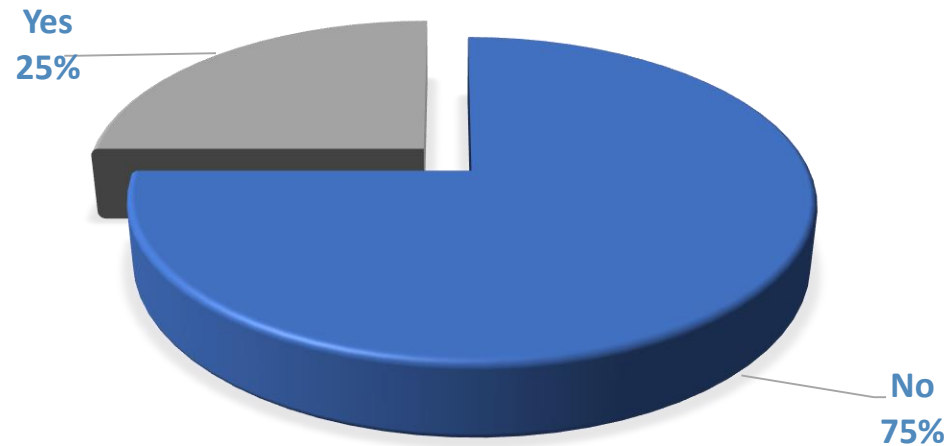
# Demographics/Participants' profile

HOW IMPORTANT DO YOU THINK IS TRAINING IN POST-EDITING?



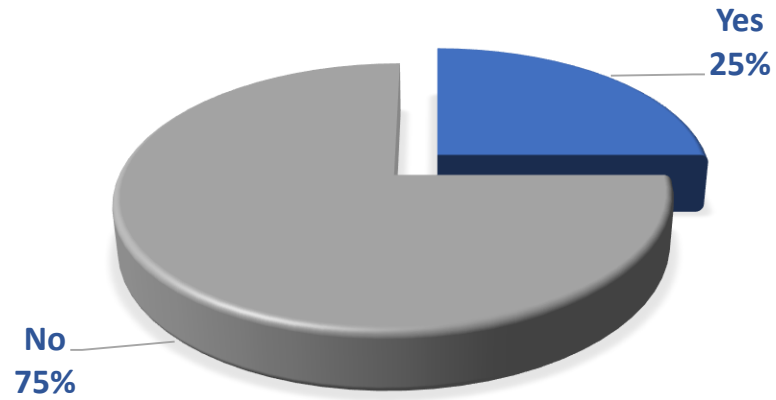
# Demographics/Participants' profile

DO YOU USE MACHINE TRANSLATION IN YOUR  
CAT TOOLS?

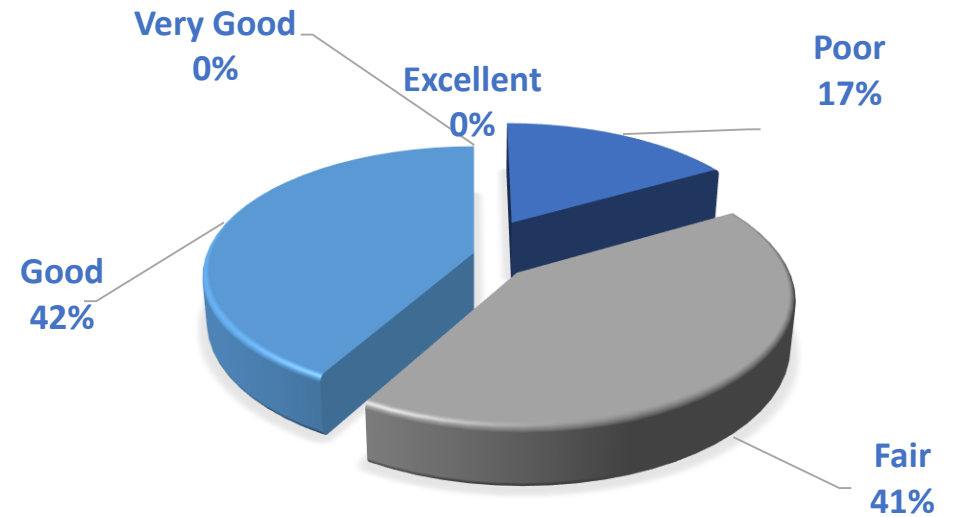


# Demographics/Participants' profile

## DO YOU USE GOOGLE TRANSLATE

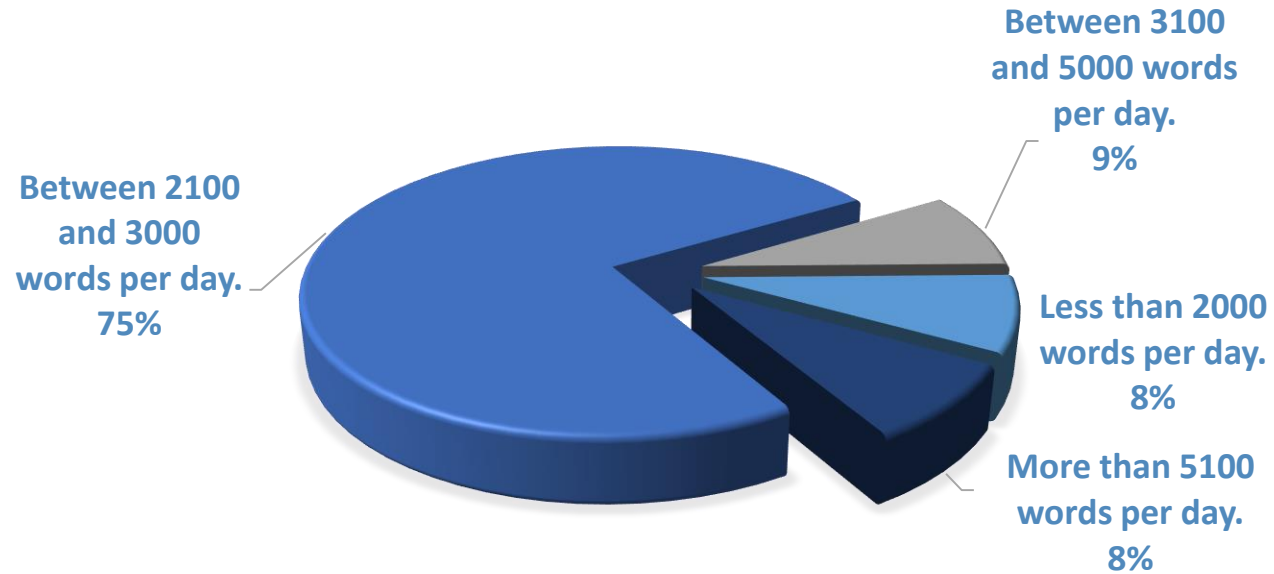


## FROM YOUR EXPERIENCE, THE QUALITY OF GOOGLE TRANSLATE IS

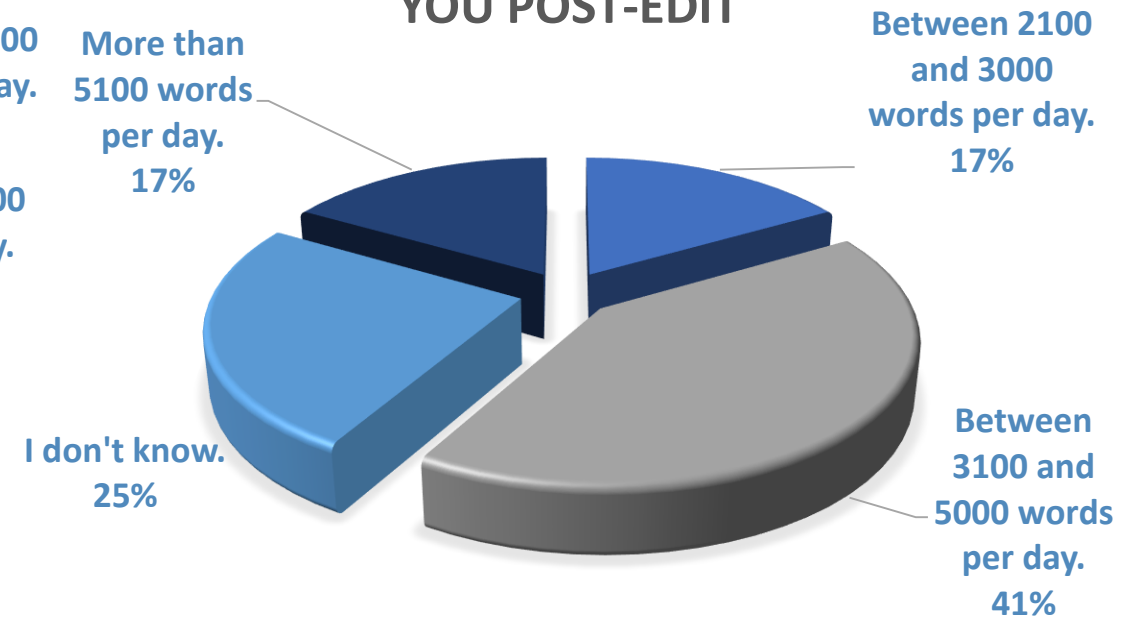


# Demographics/Participants' profile

## AVERAGE DAILY THROUGHPUT WHEN YOU TRANSLATE FROM SCRATCH



## AVERAGE DAILY THROUGHPUT WHEN YOU POST-EDIT



# Findings and analysis



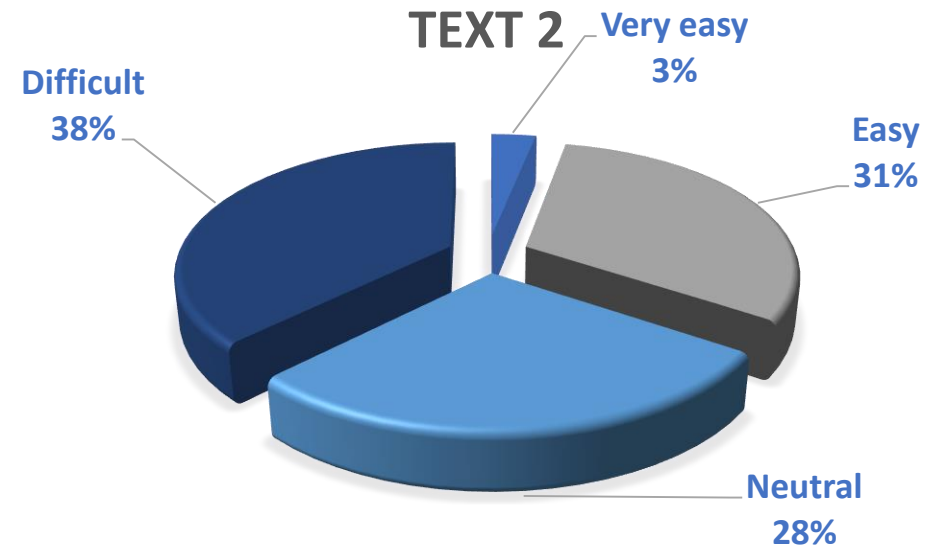
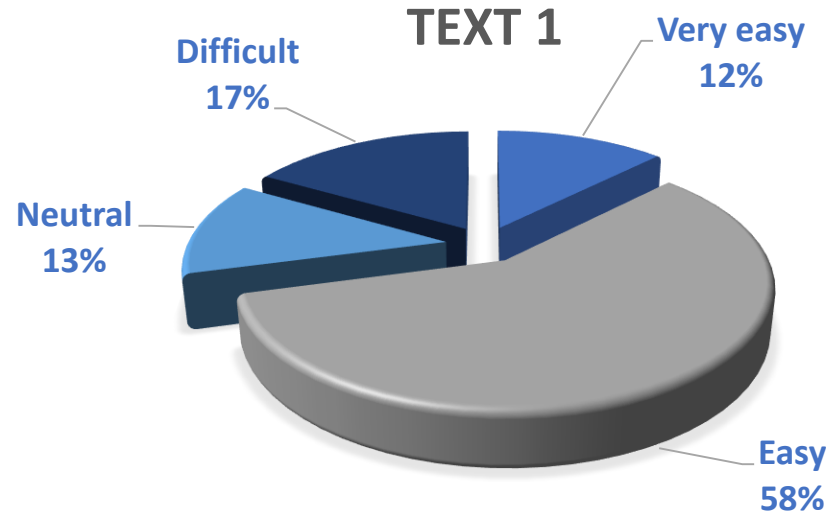
# Questionnaire A/Perception of post-editing



# Questionnaire A/Perception of MT

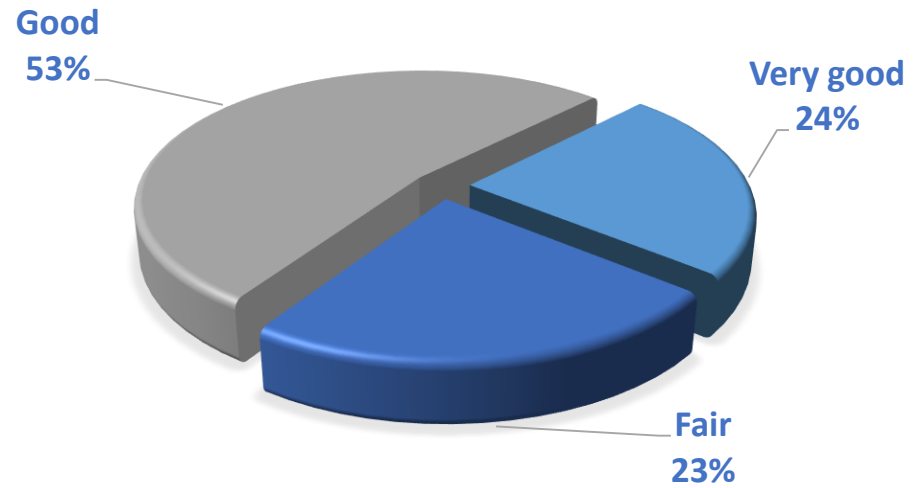


# Questionnaire B/ST difficulty

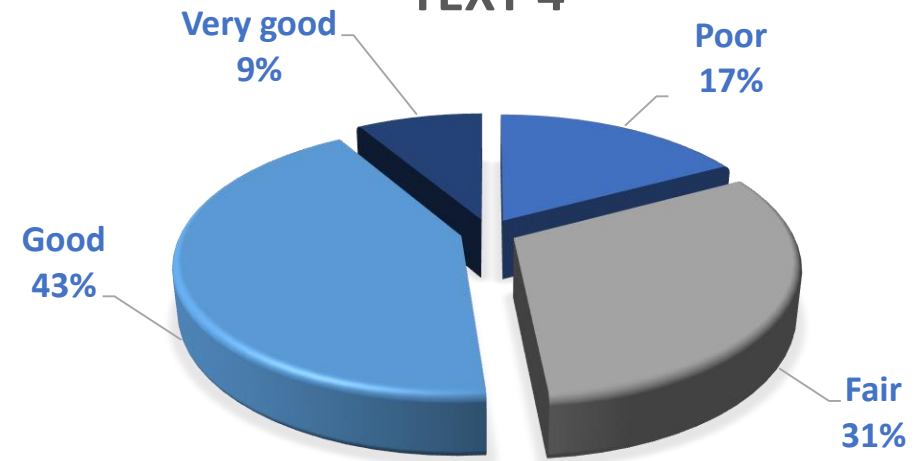


# Questionnaire B/Quality of MT raw output

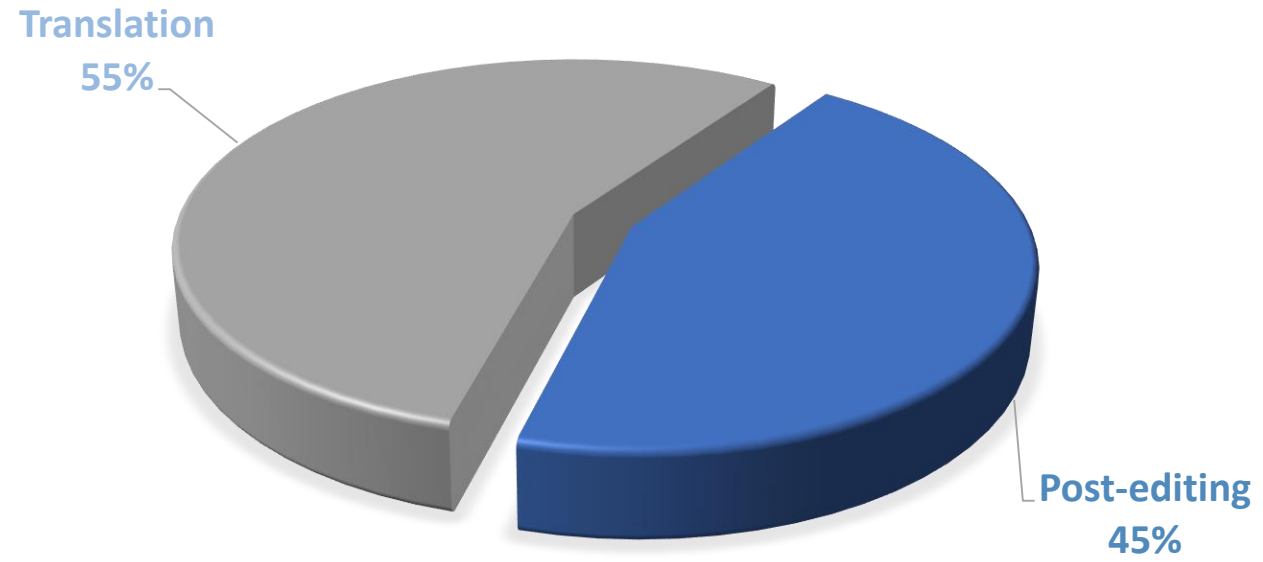
TEXT 3



TEXT 4

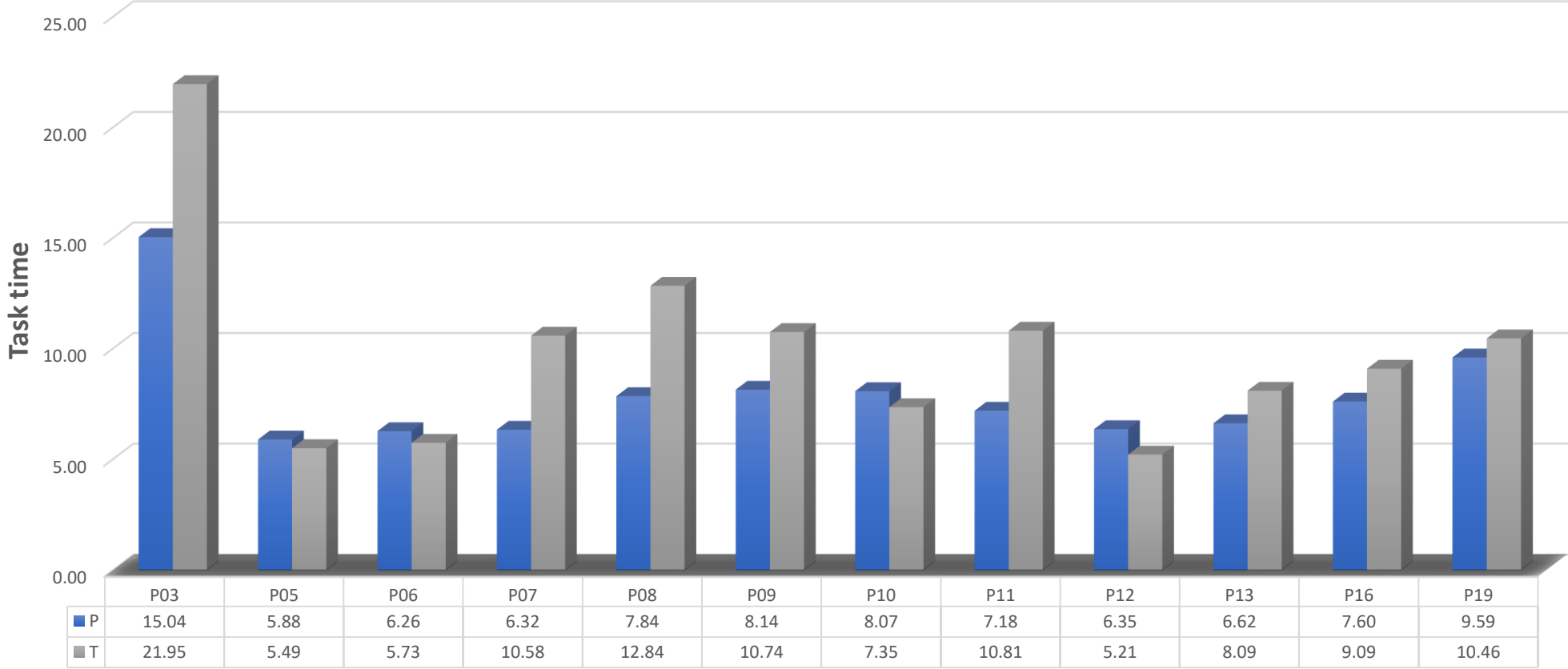


## TEMPORAL EFFORT AVERAGE TASK TIME



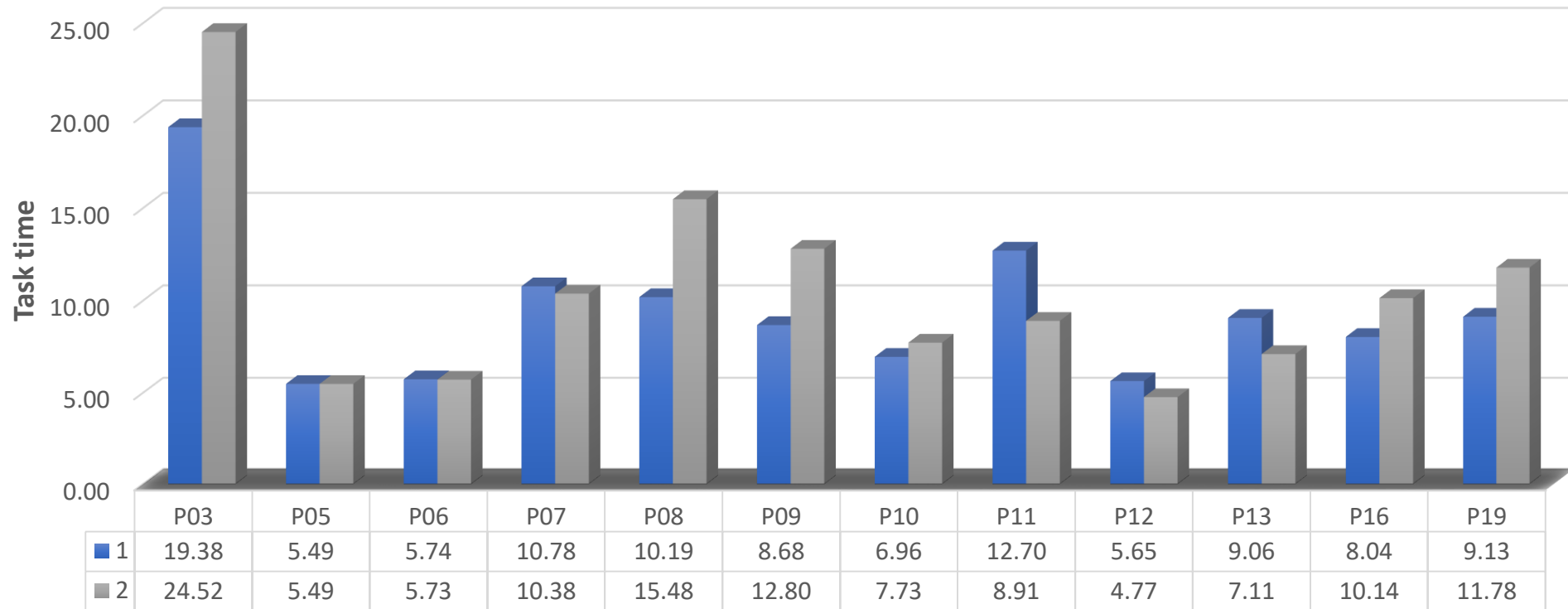
# Temporal Effort

## Average time per participant per task



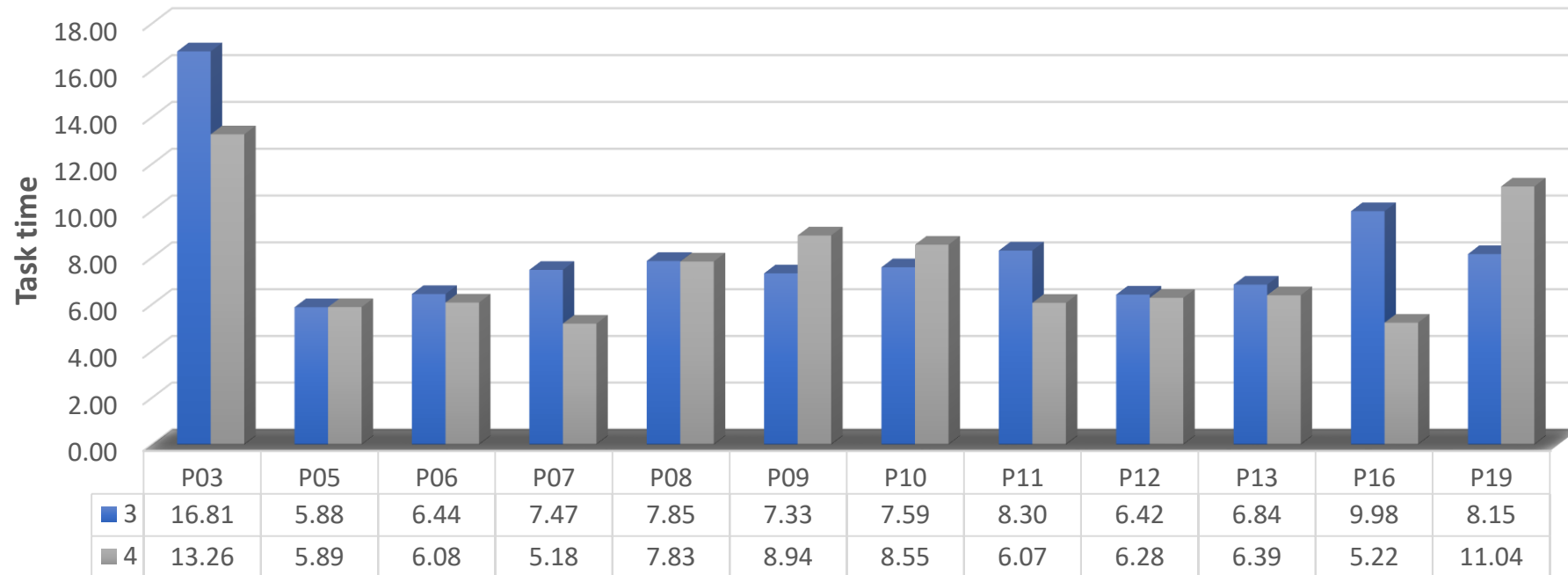
## Temporal Effort

### Average task time per participant per text in the translation task

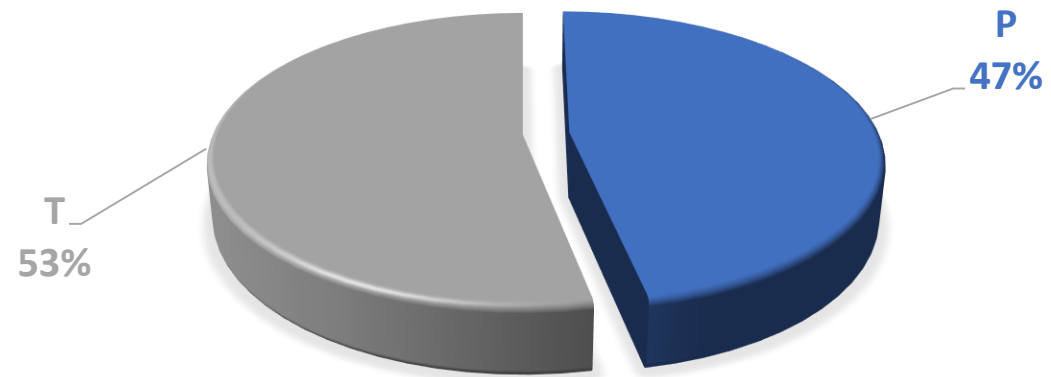


## Temporal Effort

Average task time per participant per text in the post-editing task

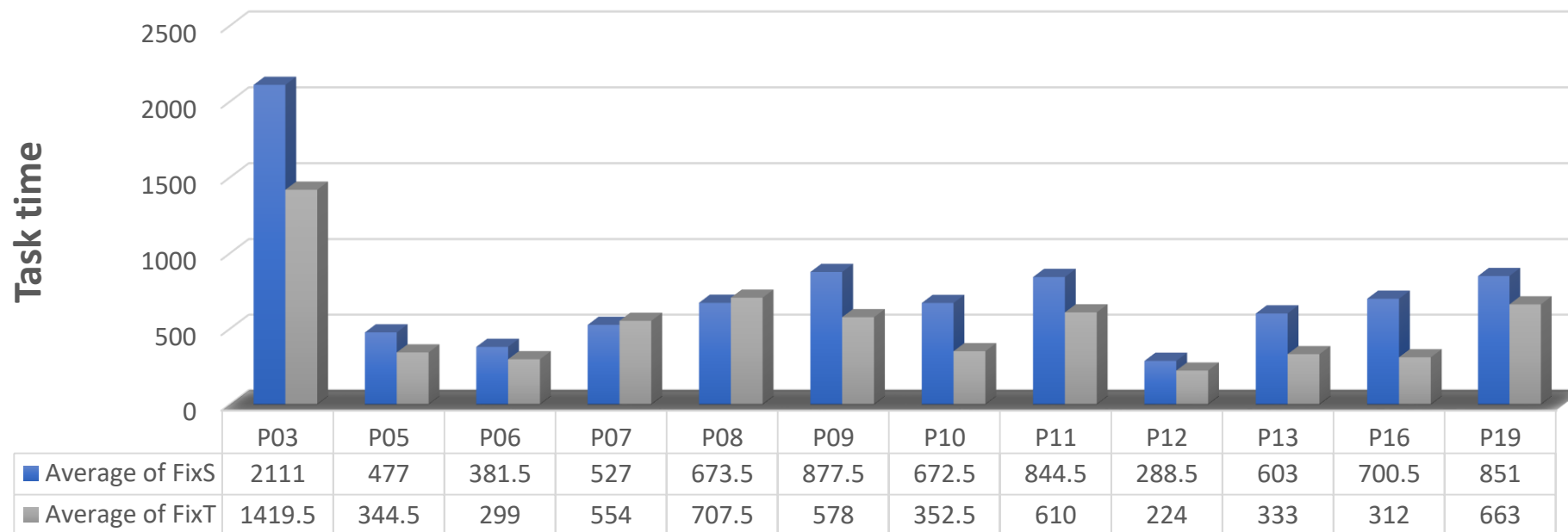


## COGNITIVE EFFORT AVERAGE FIXATIONS PER TASK



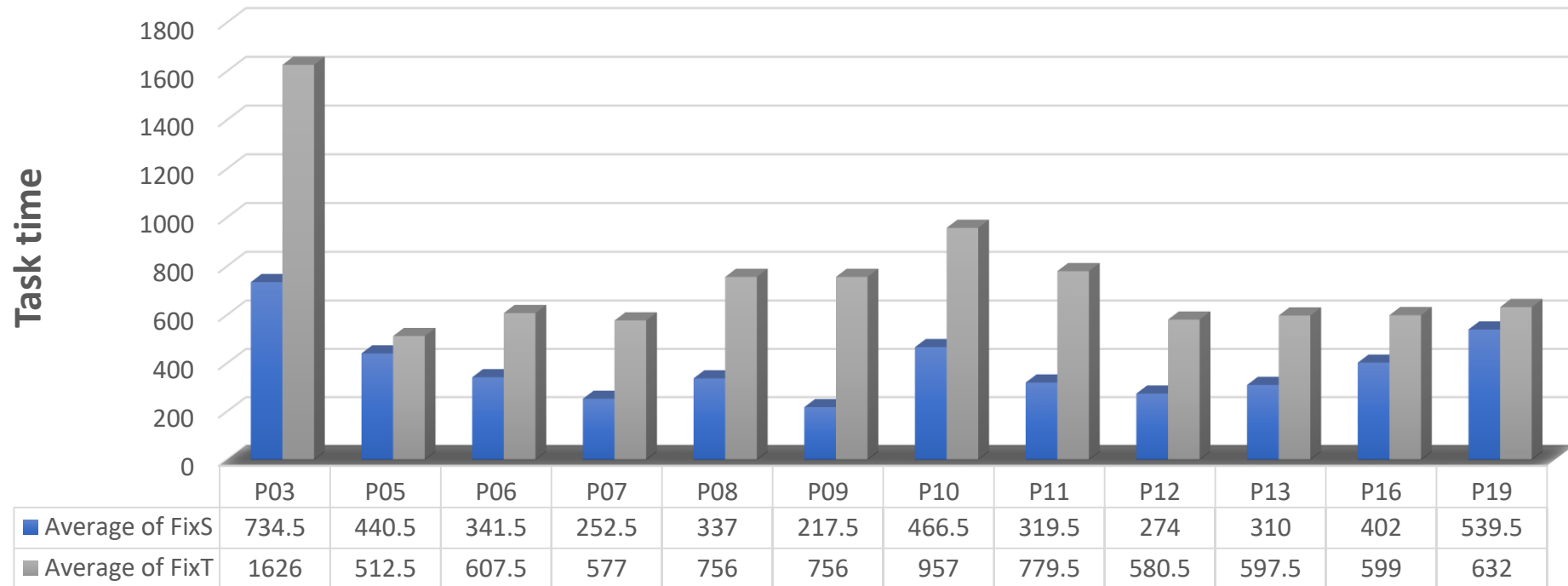
## Cognitive Effort

### Average fixations per participant in ST and TT areas in the translation task

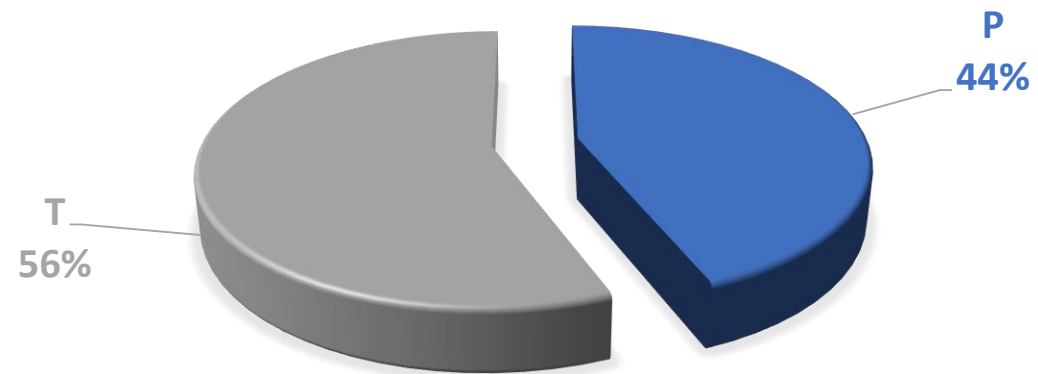


## Cognitive Effort

Average fixations per participant in ST and TT areas in the post-editing task

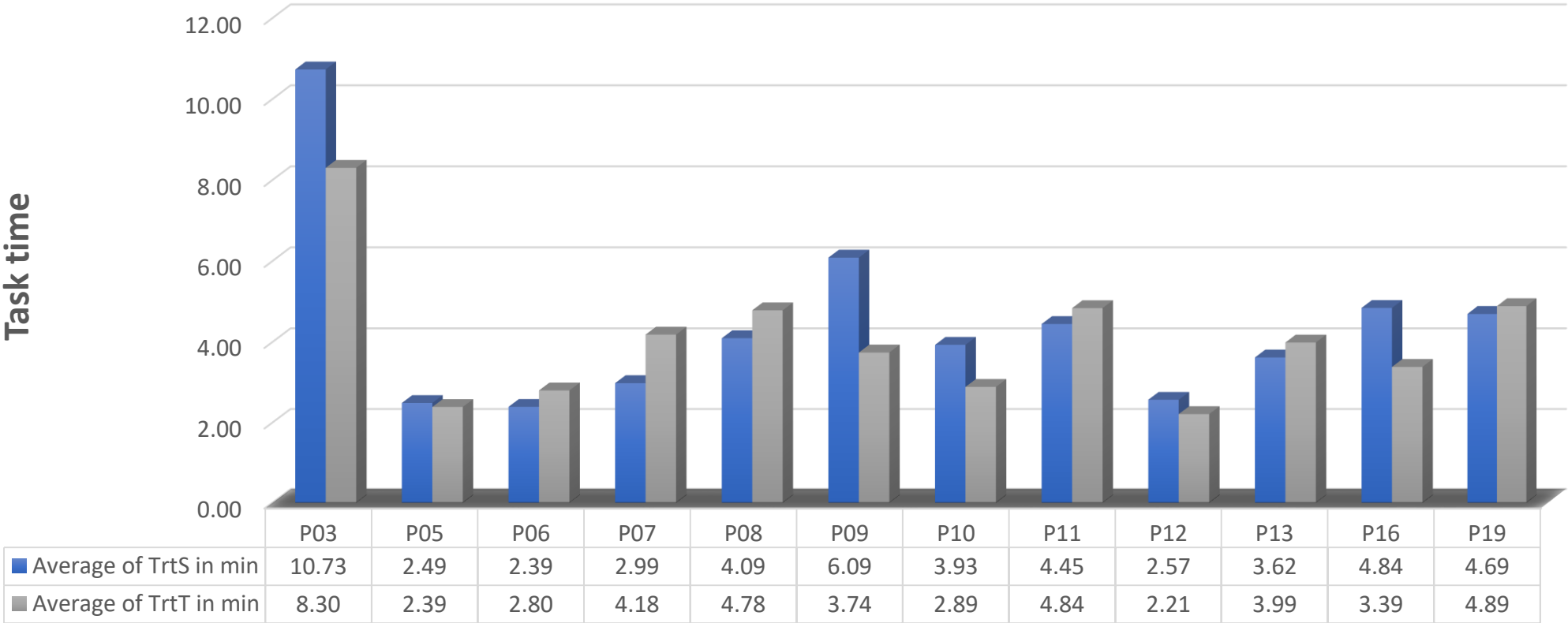


## COGNITIVE EFFORT AVERAGE GAZE TIME PER TASK



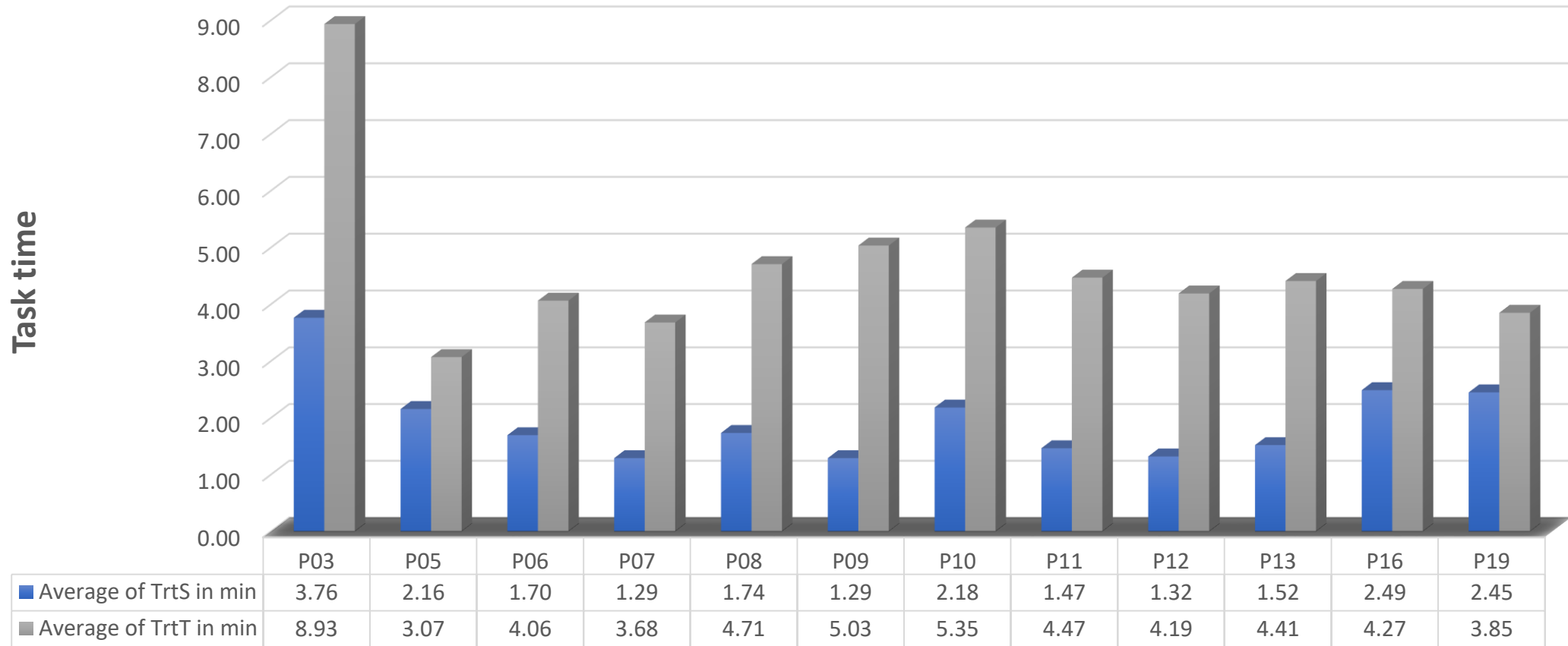
# Cognitive Effort

## Average gaze time per participant in source (ST) and target (TT) areas in the translation task

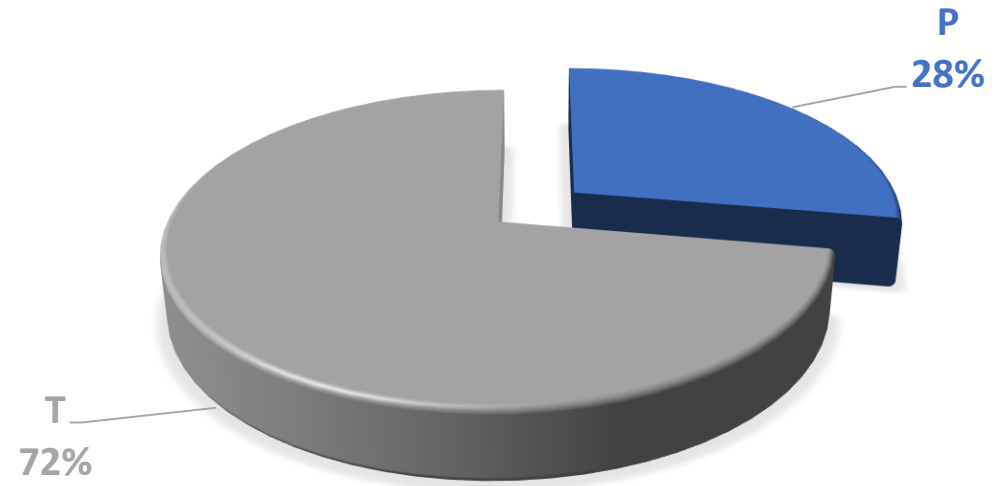


## Cognitive Effort

Average gaze time per participant in source (ST) and target (TT) areas in post-editing task

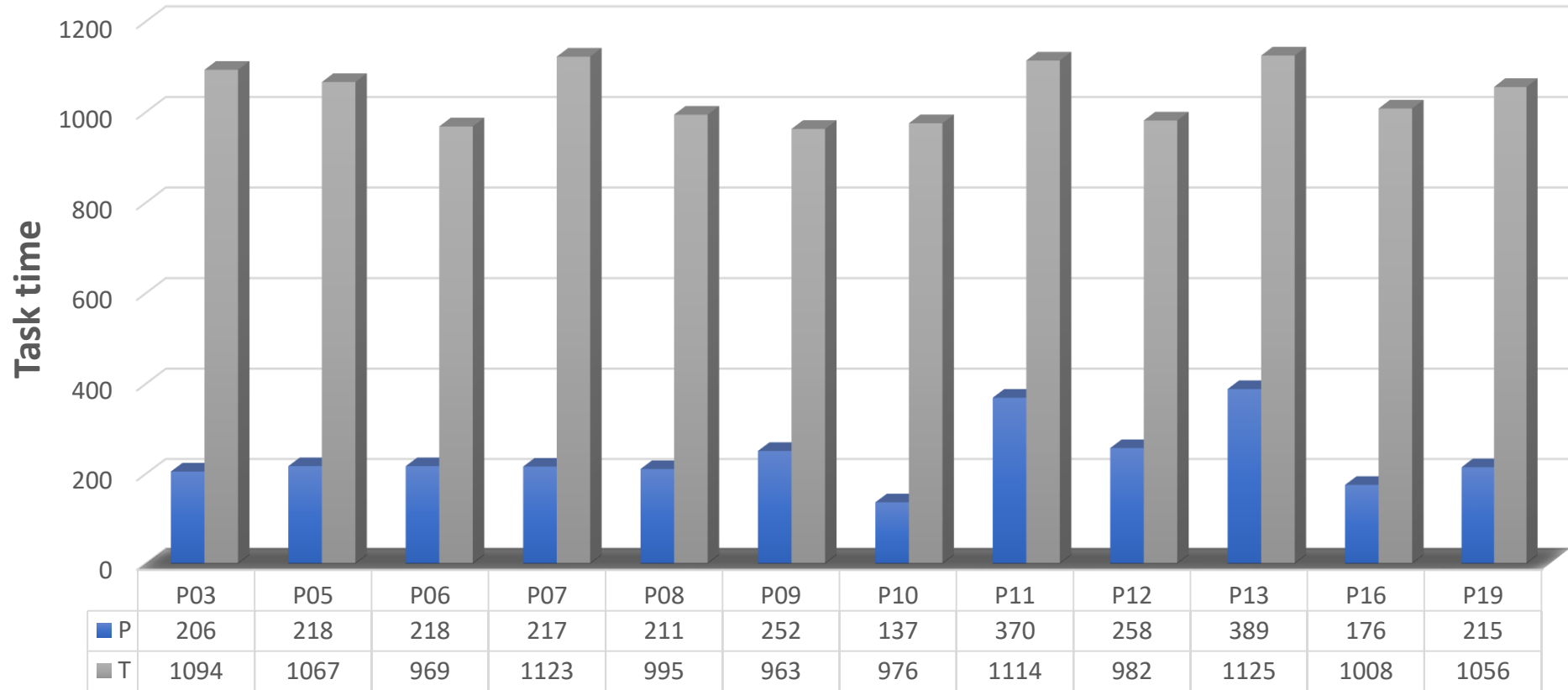


# TECHNICAL EFFORT AVERAGE KEYBOARD ACTIVITY PER TASK



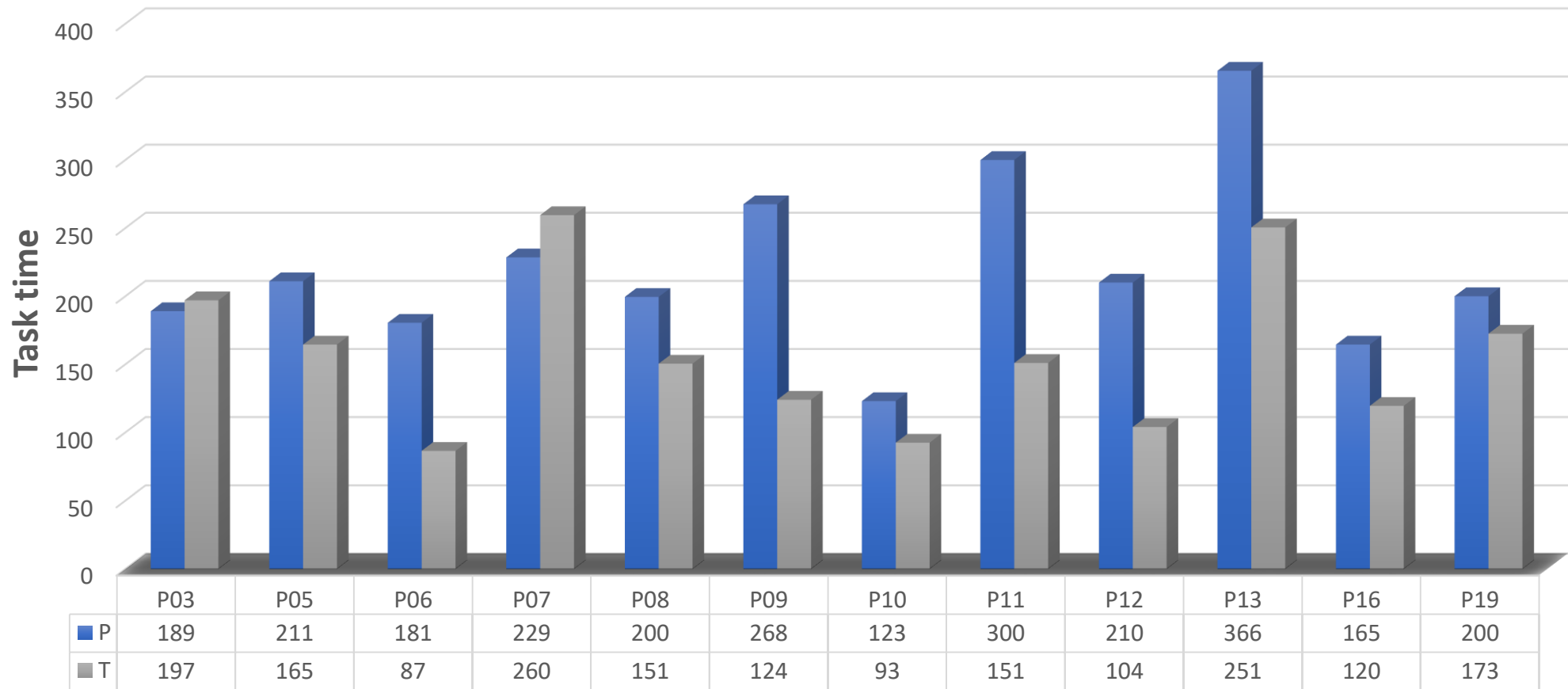
## Technical Effort

### Average number of insertions per participant per task



## Technical Effort

### Average number of deletions per participant per task



# Conclusions and future work

## Answering the Research Questions

Although the sample is small when the length of the texts and the number of participants are taken into account, our initial study indicates that **the PE of NMT output by professional translators is less cognitively demanding than human translation.**

**Translators' perceptions are mixed.** Translators are often negatively disposed towards MT, although MT and PE can constitute useful tools not only for speeding up the translation process, but also for reducing the cognitive effort.

It is our intention in the future to build on this research by increasing sample sizes and target languages, by comparing the effort experienced by professionals with that by novice translators and by complementing the results with mark-up error analysis.

Thank you!  
Dziękuję Ci!  
Ačiū!

[sosoni@ionio.gr](mailto:sosoni@ionio.gr)

[Stasimioti.maria@gmail.com](mailto:Stasimioti.maria@gmail.com)