**The influence of written translation experience on cognitive load during text-level translation: EEG and behavioral measurements**

Matthias Kobi1, Michael Boos1, Stefan Elmer1,2\* & Lutz Jäncke1,3\*

1Division Neuropsychology, Department of Psychology, University of Zurich, Binzmühlestrasse 14/25, 8050 Zurich, Switzerland.

2Computational Neuroscience of Speech & Hearing, Department of Computational Linguistics, University of Zurich, Andreasstrasse 15, 8050 Zurich, Switzerland

3University Research Priority Program (URPP) “Dynamics of Healthy Aging”, University of Zurich, Andreasstrasse 15/2, 8050 Zurich, Switzerland.

\*shared last authorship

E-mail addresses:

MK: [matthias.kobi@psychologie.uzh.ch](mailto:matthias.kobi@psychologie.uzh.ch)

MB: m.boos@psychologie.uzh.ch

SE: [stefan.elmer@uzh.ch](mailto:stefan.elmer@uzh.ch)

LJ: [l.jaencke@psychologie.uzh.ch](mailto:l.jaencke@psychologie.uzh.ch)

Correspondence to:  
Matthias Kobi or Stefan Elmer  
Institute of Psychology  
Division Neuropsychology  
University of Zurich  
Binzmühlestrasse 14/25  
8050 Zurich, Switzerland

Date: 10. September 2022

# Abstract

Research on translation at the text level and the influence of expertise on written translation is still underexplored. Furthermore, it is still unclear whether electroencephalography (EEG) metrics can be used as sensitive measures to assess cognitive workload during translation and whether the degree of workload is modulated as a function of expertise. Accordingly, we combined EEG and behavioral indices while three groups of participants with varying levels of translation expertise completed three subtasks of increasing complexity, namely reading, copying, and translation. Additionally, we assessed whether non-native English text excerpts influence cognitive load parameters compared to edited English inputs. The results showed that professional translators were characterized by the lowest workload, as reflected by reduced frontal theta power in the reading task. Furthermore, the professionals outperformed student translators and multilingual controls in reading comprehension and showed higher fluency but not accuracy ratings in the translation outputs. Finally, we provided evidence indicating that the translation of non-native English texts was more demanding than edited English, as reflected by the generally lower accuracy ratings of the translations. Taken together, these results pave the way toward a better understanding of cognitive workload in association with written translation and of the influence of non-native texts on the translation process.

**Keywords:**

Cognitive load, frontal midline theta, EEG, professional translators, written translation, language expertise, English as lingua Franca

**Highlights**

-

# Introduction

With the ongoing globalization and the associated requirement to communicate in non-native languages, there is increasing interest in better understanding the cognitive and neural underpinnings of foreign language mastery. However, such a need to use foreign languages in everyday life is caused not only by expanded migration flows across the globe but also due to a growing number of collaborations between different countries in science, economy, politics, and several other fields. Furthermore, since several of these partnerships rely on written communication, skilled translators play an increasingly important role in the exchange process. Nevertheless, an important aspect that has often been neglected in previous studies on multilingual language processing is the type of English used for written communication (e.g., Albl-Mikasa, 2013; Albl-Mikasa et al., 2017). In fact, with the fast proliferation of English as a lingua franca (ELF), several authors commonly use English as a second language for communication purposes. Hence, professional translators are increasingly dealing with non-native English texts, which are often difficult to understand due to ambiguities, incoherences, and imprecisions. Accordingly, ELF might be associated with higher cognitive demands than Standard English due to additional plausibility checks and compensation loops needed for appropriate and stylish translations (Ehrensberger-Dow et al., 2020). Drawing on this background, a better understanding of the cognitive demands associated with written language translation and the mechanisms underlying language processing in professional translators and untrained bilinguals becomes increasingly relevant in several disciplines.

Most of the previous studies on language translation and foreign language processing focused on single features at the word level, including cognate status, level of concreteness and frequency, interlingual homographs, as well as language switching costs (Christoffels et al., 2013; Costa et al., 2000; de Groot, 1992; García et al., 2014; von Studnitz & Green, 2002). Otherwise, the few studies that focused on translation at the text level provided evidence of the influence of translation direction on behavioral metrics. More specifically, forward translation (FT) from the first (L1) to a second language (L2) is typically found to have shorter response times than backward translation (BT) from L2 to L1 (Muñoz et al., 2018). In addition, neuroimaging studies found that BT and FT are partly based on different neural mechanisms as reflected by divergent patterns of brain activity (García & Ibáñez, 2016; Klein et al., 1995; Quaresima et al., 2002; Rinne et al., 2000). In particular, FT is thought to require more cognitive effort and to place more demands on executive control than BT, possibly because during FT, L1 has to be more strongly suppressed to avoid interferences from the dominant language(García et al., 2016)*.*

Although bilingualism and translation studies constitute a relatively young branch of research, a handful of them found that professional translators show changes in neuronal pathways and are characterized by facilitation in lexical processing compared to non-translator bilinguals (Beatty-Martínez & Dussias, 2017; Carl & Kay, 2012; Christoffels et al., 2013; García et al., 2014; Grabner et al., 2007; Ibáñez et al., 2010; Jost et al., 2018; Pérez et al., 2022). In particular, García et al. (2014) examined the impact of translation expertise on reading aloud in L1 and L2, as well as on BT and FT, while translating noun pairs that were manipulated in concreteness and cognate status. The authors tested 36 native Spanish speakers with a high English (L2) proficiency and divided the participants into three groups based on their translation expertise, namely first-year translation students, senior-year translation students, and professional translators with field experience. The results showed that concreteness and cognate status effects on response times were manifested in all three groups. Moreover, reading in L1 was faster than in L2 and generally faster than translating. However, the authors could not provide evidence for an advantage of BT compared to FT. Translation expertise also affected lexical processing, and translators with field experience as well as senior-year students responded faster than first-year translation students in all tasks, even though the level of L2 proficiency was comparable and the influence of translation expertise was more pronounced in word translation than in reading. In a further study by Ibáñez et al. (2010), the authors addressed the influence of formal translation training on lexical access and language switching by means of two reading experiments. In both experiments, the authors examined twelve translators with more than two years of experience as well as twelve bilinguals matched in age, language fluency, and working memory capacity. In this context, the participants had to read sentences in L1 (Spanish) or L2 (English), which contained randomly placed cognate words (orthographically similar or identical words in L1 and L2) or non-cognate control words. Furthermore, the authors manipulated the sequential order of the languages in that half of the sentences were preceded by the other language (switching trials), whereas the other half were not (non-switching trials). The sentences were visually presented word-by-word, and the participant’s self-paced reading speed was taken as an index of processing time. Furthermore, in experiment 1, the participants had to repeat the presented sentence aloud before continuing to read the next sentence, whereas there was no repetition task in experiment 2. The results of experiment 1 revealed that in the translators’ groups, cognate words were processed faster than non-cognates. In contrast, bilinguals did not show any significant difference regarding the cognate status of words. Interestingly, reading time was also generally slower for translators than for bilinguals. This effect was interpreted as indicating that the translators used a different strategy of simultaneously activating both languages which may have led to higher attentional demands and slower processing time. However, experiment 2 showed that both groups were faster at processing cognate words in L2, which might represent a non-selective activation of the non-target language. Finally, the evaluation of language switching showed that in experiment 1, professional translators were not influenced by the switching condition, while bilinguals showed switching costs only when changing from L2 to the dominant L1. However, since neither bilinguals nor translators showed switching costs in experiment 2, the authors concluded a remarkable ability of professional translators to co-activate two languages during comprehension and production and constantly switch between those two languages. In a further interesting study, Carl and Kay (2012) examined the focus of attention during written translation. In particular, they investigated gaze behavior and typing activity in a sample of 12 professional and student translators (L1 = Danish and L2 = English) while translating an English text of 160 words into Danish (BT). Since the source text was presented in the upper part of the monitor while the target text was visible in the lower part, the authors were able to evaluate both typing and translation speed as well as fixation behavior on the source or target text. The results showed that during translation, professionals simultaneously focused their gaze on the source text while producing the target text. In contrast, the students sequentially shifted their attention from one text to the other. Although typing speed was comparable for both groups, the typing behavior of the students was more fragmented and characterized by more and longer pauses compared to professionals.

Reading and translation are multistep processes that rely on distributed brain areas involved in feature extraction (i.e., recognizing visual patterns and letters), sublexical (onsets, rhymes, syllables, morphemes), lexical (length, word frequency, familiarity, orthographic and phonological neighborhood) and semantic processing (lexicality, concreteness/imageability, meaningfulness), syntactic analyses (see Balota et al. (2006) for a review), as well in higher cognitive functions (Indefrey & Levelt, 2004). Although depending on the task structure, mental workload during reading and translation can vary, based on the cognitive load theory (Sweller, 2010). The theory assumes a close relationship between working memory capacity and cognitive effort. In this vein, processing new information needs a higher working memory capacity which has been shown to be reflected in a higher cognitive load (Antonenko et al., 2010). Moreover, there is evidence showing that working memory capacity and cognitive load are inversely related, suggesting that cognitive load is driven by space occupied in working memory (Mills et al., 2017). In contrast to subjective workload measurements, objective brain metrics are highly advantageous because they have been shown to directly correlate with workload (Hogervorst et al., 2014). Furthermore, among several psychophysiological indices, including electrodermal activity (Kohlisch & Schaefer, 1996; Reimer & Mehler, 2011), eye movements (Brookings et al., 1996; Veltman & Gaillard, 1998), pupil size (Hampson et al., 2010; Porter et al., 2007), heart rate (Vogt et al., 2006) and heart rate variability (Aasman et al., 1987; Hancock et al., 1985), electroencephalography (EEG) has been shown to be the most reliable one as it directly enables to measure cognitive load (Swerdloff & Hargrove, 2020) with high sensitivity (Hogervorst et al., 2014; Mills et al., 2017). For example, using a battery of different cognitive tests, Berka and colleagues (2007) could show that EEG indices of workload correlated with both subjective and objective performance data and hence can be used as a robust tool for assessing online cognitive load. In this context, numerous studies reported increased event-related synchronization (ERS) in frontal theta power with increasing cognitive demands induced by parametrically modulating working memory load (i.e., Berka et al., 2007; Borghini et al., 2014; Cavanagh & Frank, 2014; Gevins & Smith, 2003; Holm et al., 2009; Jensen & Tesche, 2002; Sammer et al., 2007; So et al., 2017). In contrast, parietal alpha power has been shown to be inversely related to cognitive load and to desynchronize (ERD) with increasing cognitive demands (Gevins et al., 1997; Klimesch, 2012; Stipacek et al., 2003; Vassileiou et al., 2018).

ERD and ERS metrics have also been used in the context of language processing tasks (Grabner et al., 2007; Pérez et al., 2022). For example, in a recent EEG study by Pérez and colleagues (2022), the authors focussed on ERS and ERD while 27 early bilinguals (L1 = Spanish, L2 = English, high L2 proficiency) were engaged in L1 and L2 reading tasks and performed FT and BT at the word level. In particular, the participants were instructed to read the single words aloud as quickly as or to translate them and press a specific keyboard key as soon as they were ready to respond. The results showed that the FT task was performed slower than the BT one. Moreover, in an early time window in the range of 0-300 ms, FT was associated with higher theta power (4-7 Hz) at frontal electrodes as well as with lower fronto-parietal connectivity in the lower-alpha band (<10 Hz). Furthermore, greater behavioral differences between FT and BT were reflected by more pronounced power in the frontal theta power. In contrast, in a later latency window (300-600 ms), the authors revealed reduced lower-beta power (14-20 Hz) at centro-posterior electrodes as well as upper-beta power (21-30 Hz) at centro-frontal scalp sites. In a further EEG study, Grabner et al. (2007) focussed on ERD and ERS during a language translation task with different difficulty levels induced by low- and high-frequency English words. With this purpose in mind, Grabner and colleagues examined 13 advanced translation and interpreting students with German as L1 and English as L2 while the participants were asked to find the appropriate German translation of single words and typewrite the respective translations. The results showed that low-frequency words were translated slower and induced a higher parietal theta ERS (4-7 Hz, 300-600 ms) as well as an increased frontal upper-alpha ERD (10-13 Hz, 300-500 ms) compared to high-frequency words. Furthermore, successfully translated low-frequency words showed stronger ERD in the lower-alpha (7-10 Hz) and upper-alpha (10-13 Hz) band compared to not translated words, and beta ERD (20-30 Hz) at around 400 ms was more pronounced for high- than low-frequency words.

This study aimed to shed more light on the influence of written translation expertise on backward translation using an ecologically valid setting. Drawing on this background, we examined a relatively large sample of professional translators, translation students, and bilinguals while the participants completed a reading, a copying as well as a translation task at the sentence level. In particular, we assessed translation expertise based on self-reports and L2 proficiency and measured task-related cognitive workload using psychometric and behavioral indices in association with EEG. Furthermore, we used authentic English texts written for an international research conference by non-native speakers as ELF version and their edited equivalents in Standard English (EdE). We expected that professional translators would demonstrate shorter reading times in the reading task and produce more output texts during the copying and translation tasks compared to translation students and bilinguals without training in translation. In the same vein, we expected that professional translators would outperform the student and bilingual cohorts in the translation tasks and that this advantage should be manifested in higher output ratings of fluency and accuracy. We also hypothesized that translating would generally induce a higher cognitive load than copying and reading. This effect is expected to be expressed in increased theta and weaker alpha power. Furthermore, based on their expertise, we assumed that professional translators would show lower cognitive workload during all tasks than the other two groups, as reflected in neurophysiological metrics and perceived task difficulties. Regarding the differences in the text input, we expected that ELF text versions would lead to a higher cognitive load than the EdE versions and that ELF processing is associated with lower output ratings of fluency and accuracy.

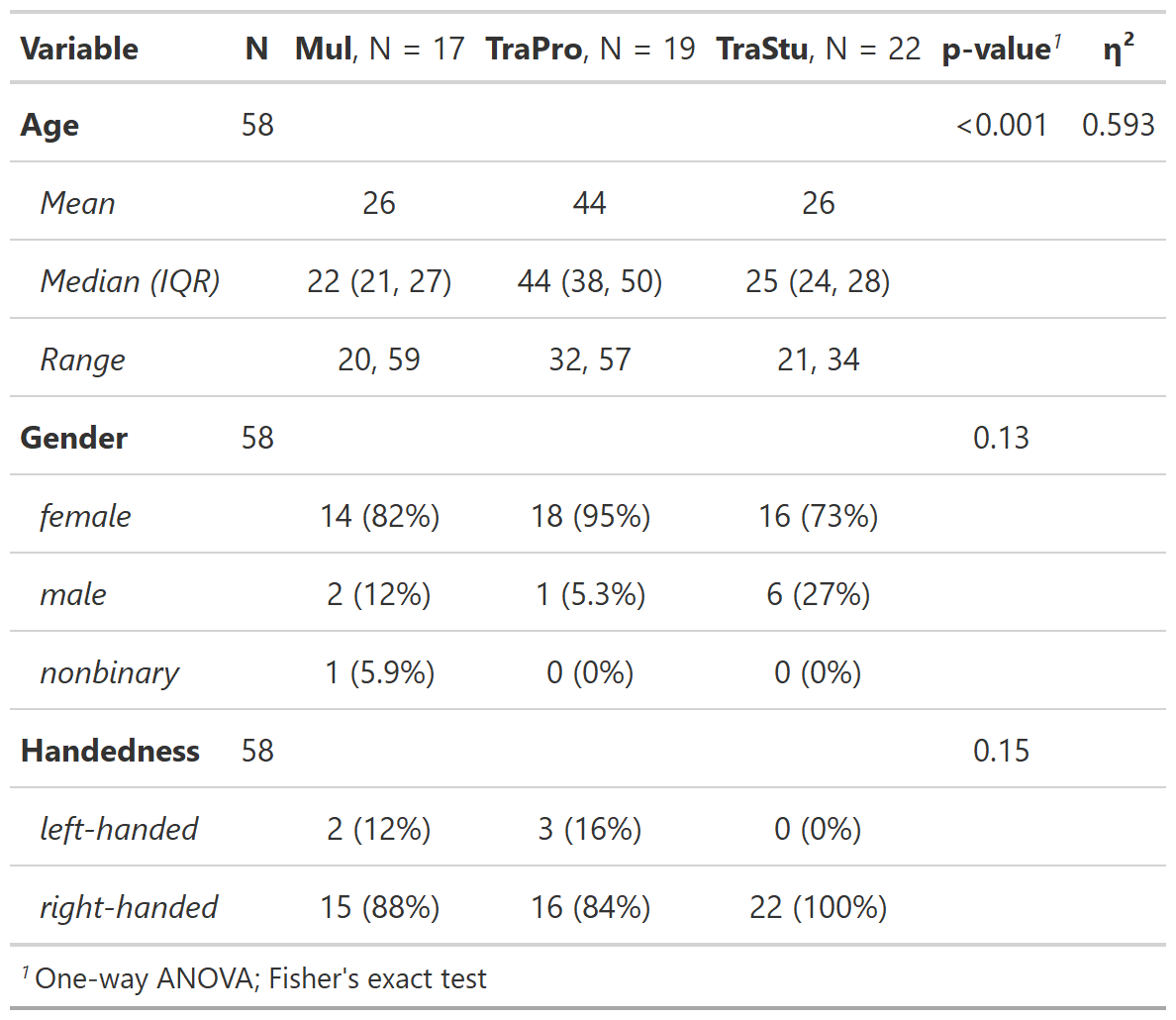
# Methods

## Participants

We analyzed the data of three groups of 58 native German (L1) participants (48 females, one nonbinary) with a professional English (L2) background, namely professional written translators (TraPro, N = 19), trainee translators (TraStu, N = 22), and multilingual controls (Mul, N = 17). The multilingual controls consisted of English language and literature studies students, or high school teachers of English as a foreign language. All participants were required to use L2 in their daily routines, and the primary direction of translating for professional and student translators was from L2 to L1. Obviously, since we recruited participants with varying levels of professional experience, the groups could not be matched regarding age. However, the three groups did not differ in respect of gender or handedness (Annett, 1970, see Table 1 for participant overview), and all participants had a normal or corrected-to-normal vision. Two participants reported using medicaments for diabetes, two for high blood pressure, and two reported concussions that occurred longer than five years before testing. The experiment lasted approximately four hours and was rewarded with cash. Fourteen participants were excluded from the analysis because of failure to follow the instructions of the experiment (N = 6), medication (N = 4, anti-depressants or Ritalin), and noisy or missing data (N = 4). The study was carried out according to the principles of the declaration of Helsinki and approved by the Swiss National Science Foundation ethics committee.

### Table 1

*Sample characteristics*



*Note.* Mul = multilingual control group, TraPro = professional translators, TraStu = translation students.

## Psychometric measurements and questionnaires

Every participant completed a short English language test (https://www.sprachtest.de/ einstufungstest-englisch) to assess L2 proficiency. This online procedure lasted about 15 minutes and consisted of 13 vocabulary, grammar, listening, and reading comprehension questions. The maximum score of the test was 40. Furthermore, we collected data on the age of L2 acquisition and experience in translating and interpreting (cumulative number of training hours and cumulative number of training hours per day since the age of 17) using a language background questionnaire. To assess working memory capacity, participants completed both a visual and an auditory 3-back task comprising 60 letter stimuli and20 target stimuli. The order of the tasks was pseudorandomized across the groups. The N-back data were analyzed using d-primes (d’) which were calculated as the difference between the z-transformed hit and false alarm rates (Hautus et al., 2021), reflecting a measure for the sensitivity of the performance in yes/no task that is uninfluenced by response bias (Stanislaw & Todorov, 1999). Additionally, we evaluated the general cognitive capabilities of the participants using a short version of the WAIS (Wechsler Adult Intelligence Scale) test battery (Waldmann, 2008). This short version was composed of four subtests: number-symbol associations, detection of commonalities, the mosaic test, and digit span forward and backward. Based on standardized T-values, the four subtests have been shown to sensitively reflect general intellectual abilities (Waldmann, 2008).

## Stimulus material

In this experiment, we used two original English abstracts submitted to an international research conference. Because the authors of the abstracts were non-native English writers, their original texts were regarded as ELF stimuli. In a next step, the two abstracts were adapted to edited-to-standard English (EdE) versions by professional translators of the Zurich University of Applied Sciences (ZHAW). Only a few changes were made to keep the texts as close as possible to the original while generating grammatically correct sentences and overall better readability. This translation procedure resulted in four different text stimuli: text 1 (ELF, original, 746 words in 25 sentences), text 1 (EdE, 750 words in 27 sentences), text 2 (ELF, original, 776 words in 37 sentences), and text 2 (EdE, 769 words in 37 sentences).

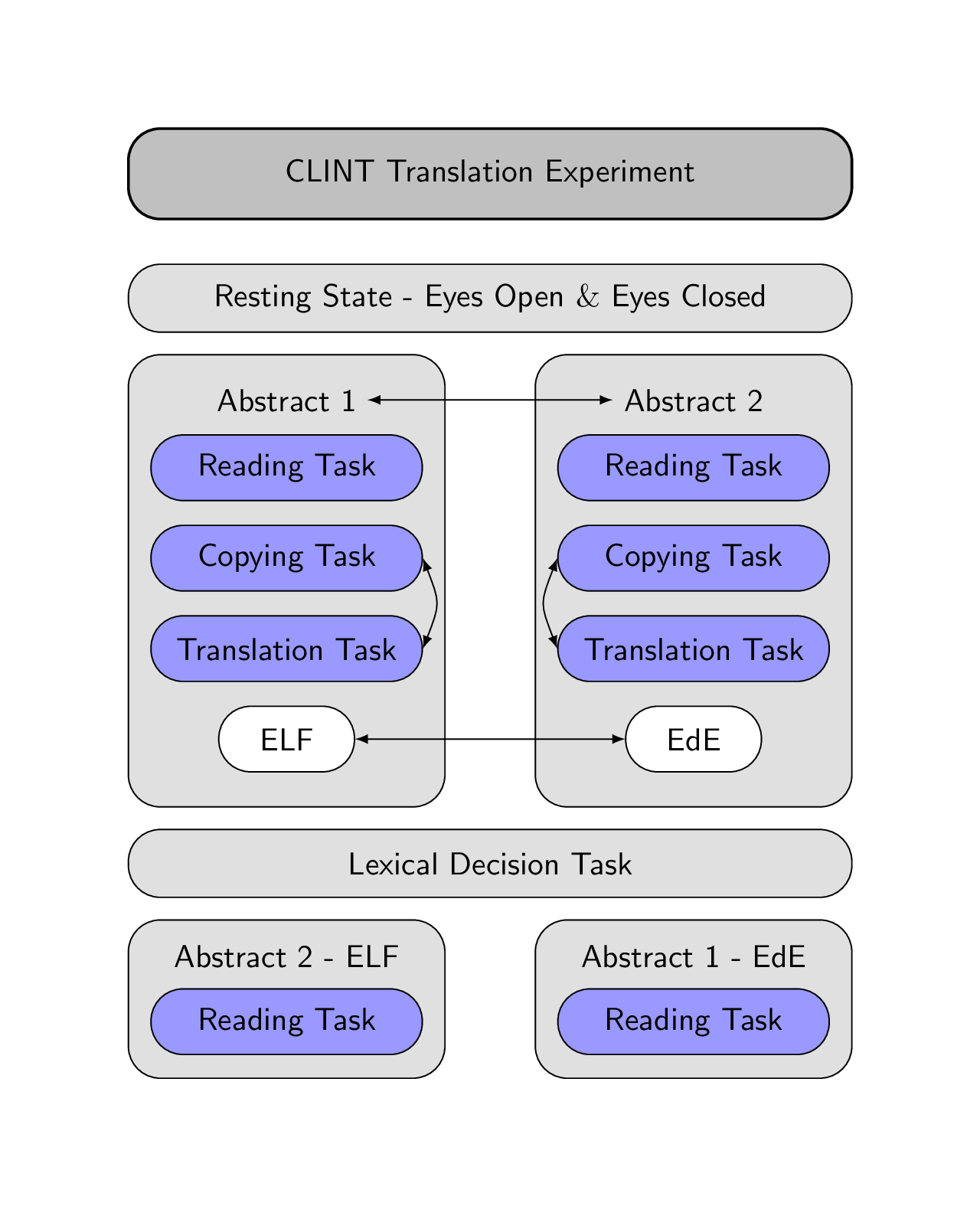
## Experimental procedure

First, participants started with the reading task, followed by the copying and translation tasks of the first abstract. In the reading task, the text was presented sentence by sentence, and participants were instructed to read at a self-paced speed through a button press. Subsequently, we asked the participants to judge the difficulty of the task, and the answer was collected through mouse press on a 10 cm horizontal bar (easy on the left, difficult on the right). Additionally, we asked five control questions to check whether participants read and understood the text. The questions were in a multiple-choice format, and participants had to choose one out of three answer possibilities by pressing a corresponding response key. Second, in the copying task, participants were asked to copy the presented sentences, and therefore, the generated output was an English (L2) transcription. After completing a sentence, participants could move on to the following one by pressing the “Enter” button. Third, in the translation task, the presented sentences had to be translated to German (L1), reflecting a backward translation (BT) process. Again, after completion of a sentence, participants had to press “Enter” to continue with the following sentence. After the translation task, we asked the participants again how difficult they judged the task, using the same procedure described above.

In all tasks, the words included in the sentences were separated by double spacing and double lines. In the reading, copying, and translation task, the sentences that had to be processed were displayed in the upper part of the monitor, while the participants’ answers in the copying and translation task were presented in the lower part. The reading task duration differed based on the participants’ self-paced reading. However, the copying and translating tasks were limited to five minutes each. After working on the first abstract, participants continued using the same procedure with the second abstract.

### Figure 1

*Experimental design*



*Note.* Arrows indicate randomizations in the experiment.

In the experiment, we randomized the order of the abstracts (text 1, text 2), the English versions (ELF, EdE), and the copying and translation tasks across participants (Figure 1). Accordingly, each participant only processed an abstract in one version but not in the other one. In particular, if the first abstract was in ELF, then the second one was in EdE, and vice versa. Since the copying and translation task duration was limited to 5 minutes each, participants did not process the whole text but always started from the beginning and worked through the text sentence by sentence. However, it was ensured that no sentence was used twice in the copying and translation task. At the beginning of the experiment, instructions for the task were presented on the computer screen, and to become confident with the keyboard, participants had to copy a sentence that contained all possible special symbols from the abstract.

## Data acquisition

After written informed consent was obtained, participants completed all psychometric measurements. Afterward, they were prepared for the behavioral and EEG data acquisition, which took place in a light-dimmed Faraday cage where the participants were seated approximately 70 cm in front of a 24-inch monitor. The participants were instructed to relax and stay as still as possible during the EEG measurements, while a chin rest ensured a stable head position. The experiment was programmed in MATLAB 2016b using the Psychophysics Toolbox Version 3 extension (Kleiner et al., 2007) for behavioral data acquisition. Furthermore, we recorded high-density EEG data at a sampling rate of 500 Hz with a bandpass filter of 0.1-100 Hz using the EGI 300 Geodesic EEG system with a 128-channel HydroCel Geodesic Sensor Net (HCGSN) (Electrical Geodesics, Eugene, Oregon). Before recording, each electrode was double-checked to ensure a good contact on the scalp, and impedances were kept below 40 kΩ. This procedure was repeated after the EEG resting state, the processing of abstracts 1 and 2, and the lexical decision task. The recording reference electrode was Cz.

## Behavioral data processing

The preprocessing of the behavioral data was done using R (version 3.6.3, https://www.r-project.org/). For the reading task, we evaluated the percentage of the correct answered control questions per text as well as the average reading duration per sentence, which was adjusted for the different lengths of the texts. Furthermore, we analyzed the perceived difficulty as the distance in cm from 0 (easy) for both the reading and translation tasks. Therefore, higher values indicate a more pronounced perceived difficulty of the task. Regarding the keystroke data in the copying and translation tasks, we evaluated the total amount of chars typed during the 5 minutes, which can be regarded as a measure of efficiency in both the copying and translation tasks (Ehrensberger-Dow & Massey, 2014). Furthermore, we retrieved the percentage of deletions, namely pressing the “backslash” on the keyboard, relative to the total number of chars typed for both tasks. Finally, we analyzed the output generated by each participant in the translation task by rating the fluency (0: incomprehensible – 5: flawless German) and the accuracy ( 0: no meaning – 5: all meaning) per sentence (Koehn & Monz, 2006). For the rating, we fully randomized the sentences from both texts and conditions and all participants. Three independent raters (three German native translators with a master’s degree in translating, English as primary work language, and experience in rating target text material) rated the fluency first and subsequently the accuracy of all sentences. For the accuracy rating, the translation output was compared to a reference translation template provided by three native German translators with a master’s degree in translation and English as the primary work language. Then, the sentence ratings were averaged per condition (ELF vs. EdE) and both texts to calculate an intraclass correlation coefficient (ICC) using the irr package (Version 0.84.1, https://cran.r-project.org/web/packages/irr/) in R. Applying a 2-way mixed-effects model of the type “consistency” and a mean-rating (k=3) revealed an ICC(C,3) = 0.575 (95%-confidence interval = 0.421 – 0.694) for the fluency rating and an ICC(C,3) = 0.909 (95%-confidence interval = 0.875 – 0.934) for the accuracy rating. The ICC for the fluency rating likely reflected moderate reliability, whereas the ICC for the accuracy rating demonstrated an excellent reliability (Koo & Li, 2016). Finally, we averaged the three raters to generate a mean rating score for fluency and accuracy, and this value was further used in the statistical analyses.

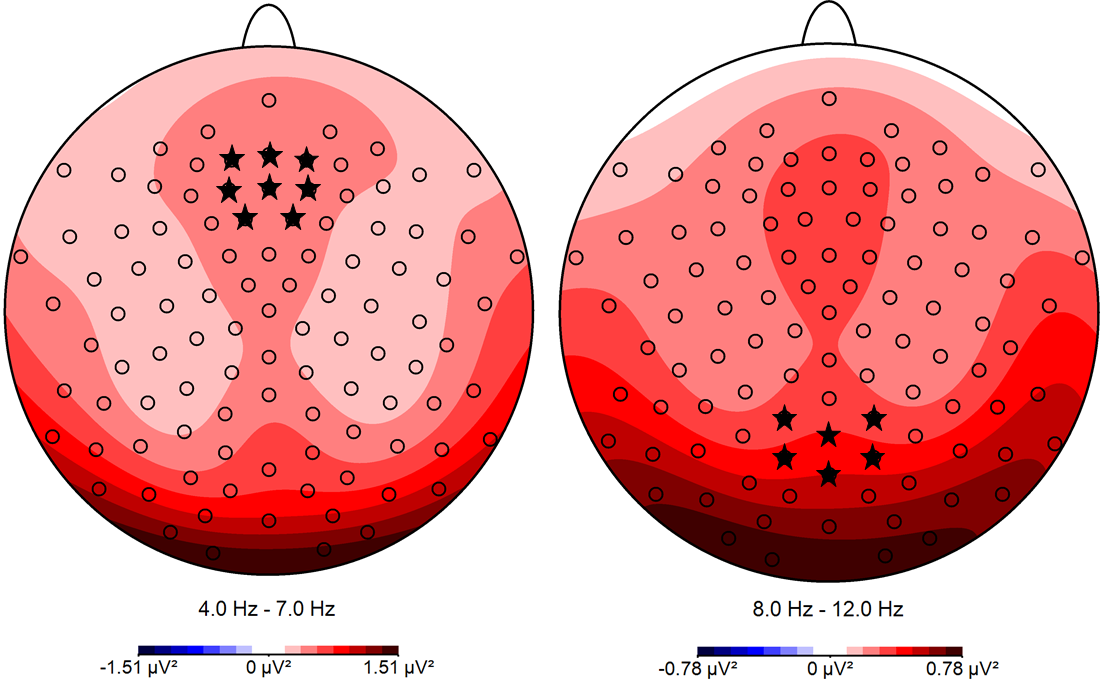
## EEG data processing

The data were processed using MATLAB (2018), EEGLAB (version 2021\_0, Delorme & Makeig, 2004), and Brain Vision Analyzer (version 2.2.0, BrainProducts, Munich, Germany). For EEG data preprocessing, we used the Automagic toolbox implemented in MATLAB (v.2.5, Pedroni, Bahreini, & Langer, 2019), which is a pipeline for automatic EEG data cleaning, using the following options. First, the number of EEG channels was reduced to 105 by discarding channels lying on the neck and face. Second, we applied the PREP pipeline for bad channel detection with the minimum variance set to 1. Third, we used the ICLabel approach with a temporary 2 Hz high-pass filter for artifact correction to remove muscle, eye, heart, and channel noise components with a probability threshold higher than 0.8. Fourth, we selected eleven fronto-polar electrodes for the electrooculogram (EOG) regression as proposed by Automagic. Fifth, power line noise at 50 Hz was removed using the ZapLine method (de Cheveigné, 2020), eliminating five components. Sixth, we applied a 0.1 Hz high-pass and a 30 Hz low-pass filter. Finally, bad channels were reconstructed through spherical interpolation, and we applied the detrending algorithm to remove slow drifts. This procedure was applied independently for each task to avoid preprocessing unnecessary noise the during pauses in between the tasks.

Further preprocessing of the EEG data was executed in the Brain Vision Analyser version 2.0.2 (BrainProducts, Munich, Germany). First, we re-referenced the data to an average reference montage, and segmented the EEG into the different task-related segments of maximum 5 minutes duration. Second, we used an automatic raw data inspection to mark bad time windows indicating remaining artifacts that were not removed by Automagic. Third, we divided the data into segments of 2 s length without overlaps, in which data segments marked as bad intervals were skipped. Forth, a fast Fourier transform (FFT) with a Hanning window (Length = 10%) was applied to all remaining segments. The resulting FFT transforms were averaged per participant and condition before exporting the power values for the theta (bandwidth from 4 to 7 Hz) and alpha (bandwidth from 8 to 12 Hz) bands. Based on the voltage distribution of the grand average across the reading, copying, and translation tasks, we analyzed theta power at 8 frontal channels (E4, E5, E10, E11, E12, E16, E18, and E19) and alpha power at 6 parietal channels (E61, E62, E67, E72, E77, and E78, see Figure 2 for electrode positions). Finally, we averaged the power for both frontal and parietal electrode pool and frequency band, and these values were used for statistical analyses.

### Figure 2

*Mean topographical voltage distribution maps for theta (left) and alpha (right) band across reading, copying and translation tasks and all participants.*

**

*Note.* The channels selected for statistical analyses are marked with \*.

## Statistical analyses

For the analysis of the psychometric and questionnaires data, we used one-way ANOVAS for continuous variables and Fisher’s exact test for non-continuous variables implemented in the gtsummary package (<https://cran.r-project.org/web/packages/gtsummary/>). For continuous variables, generalized eta squared was calculated as a measure of the effect size using the lsr package (<https://cran.r-project.org/web/packages/lsr/>). Statistical analyses for the behavioral and neurophysiological data were performed using Linear Mixed Models (LMM) implemented in the lme4 package (Version 1.1-23, https://cran.r-project.org/web/packages/lme4/). For model-fitting, we used a bottom-up strategy starting with the null model and added fixed effects for our target variables. We reported the estimates of the fixed effects in the resulting model indicated by β. In general, we used three levels for task (reading, copying, and translation task), two levels for text (text1 and text2), two levels for condition (EdE and ELF) as well as three levels for group (TraPro, TraStu, and Mul).

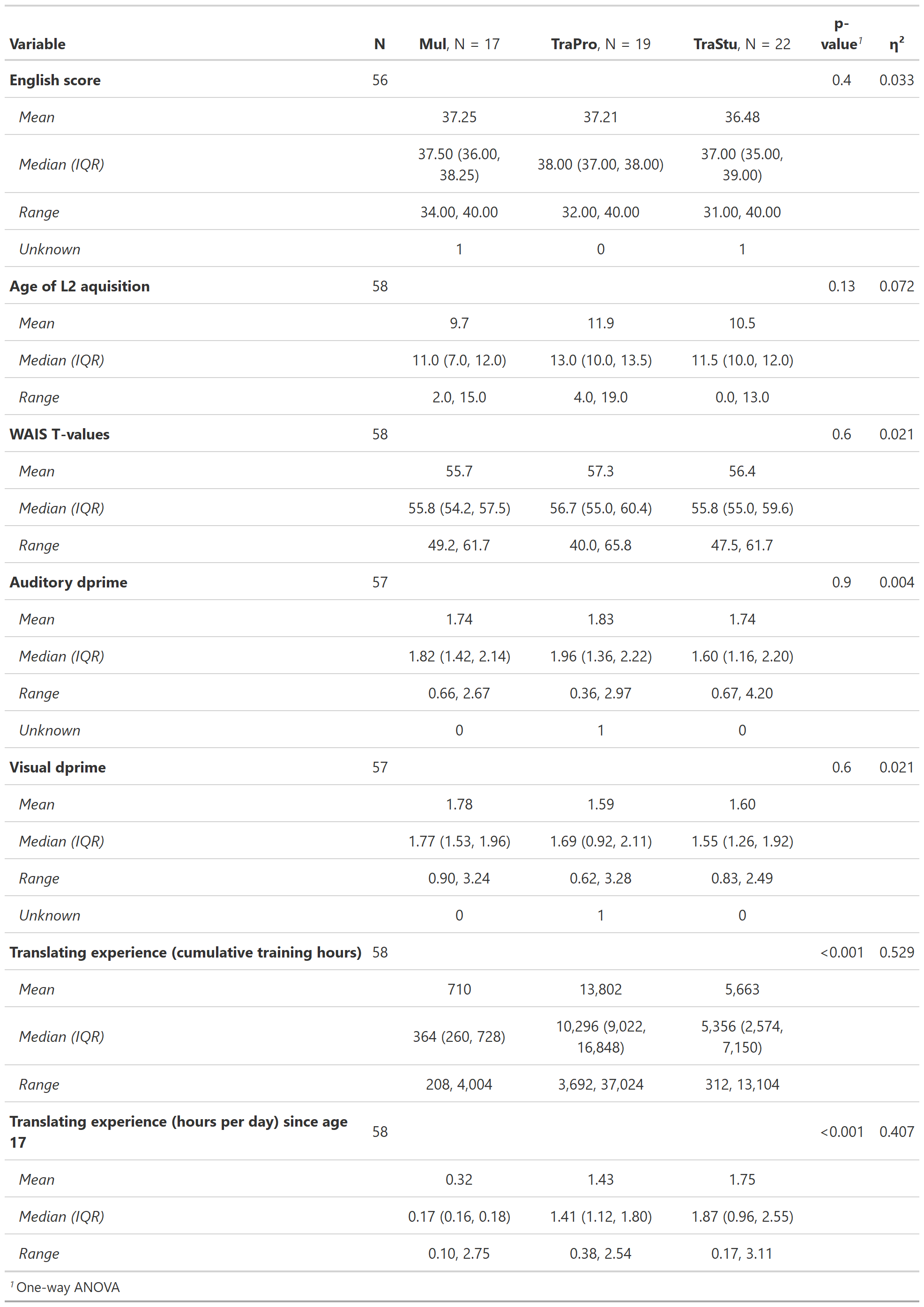
# Results

Psychometric measurements and questionnaires**.**

The three groups did not differ regarding English score, age of L2 acquisition, WAIS T-values, auditory d’, and visual d’ (see Table 2). However, as expected, the groups differed regarding cumulative number of training hours in translating and interpreting, as well as in the cumulative number of training hours per day since the age of 17. The results of the psychometric data and questionnaires are summarized in Table 2.

### Table 2

*Results of the psychometric data and questionnaires.*

****

*Note.* Mul = multilingual control group, TraPro = professional translators, TraStu = translation students

## Behavioral results

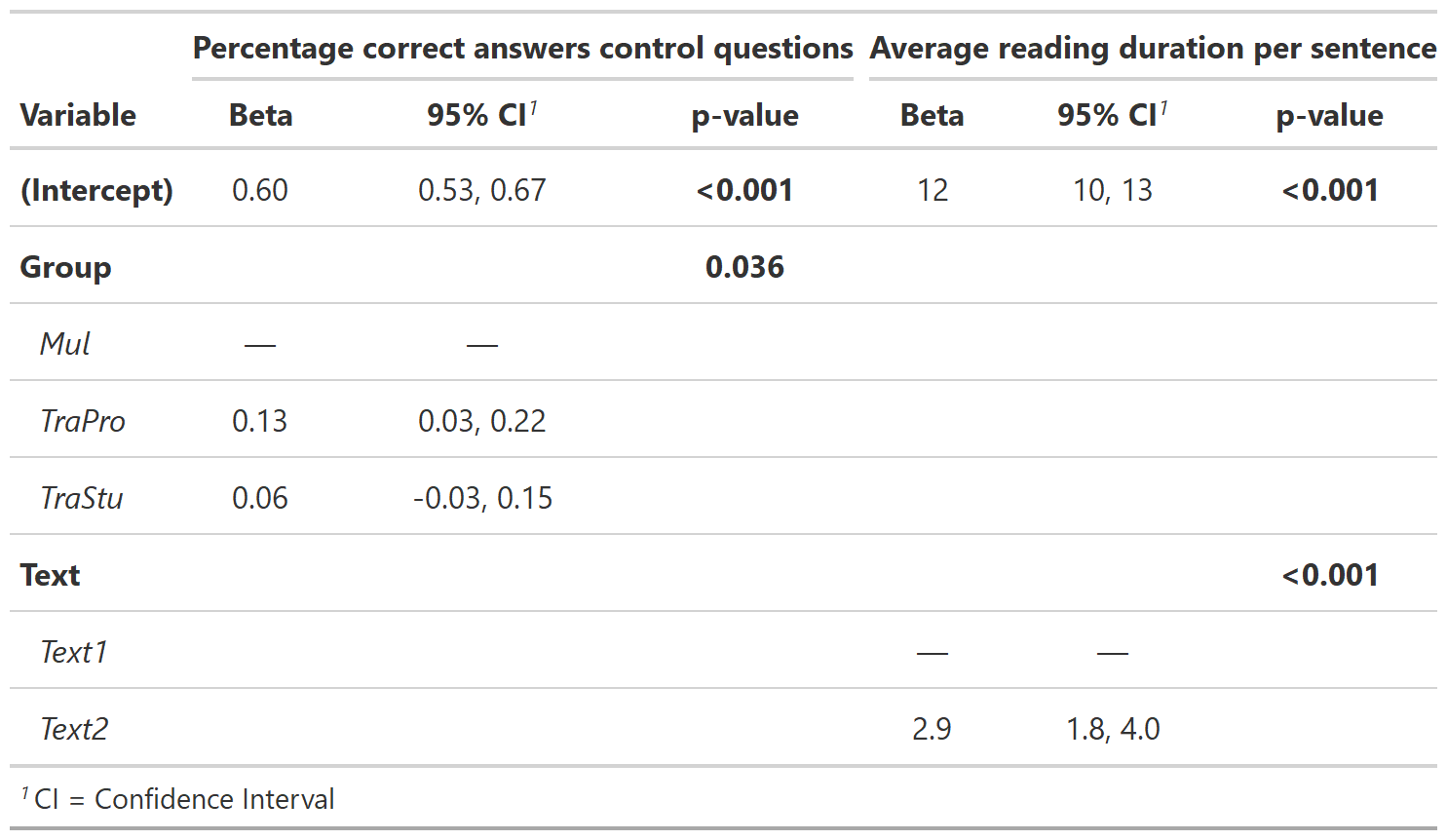
### Reading task

For the analysis of the average reading duration per sentence, including a fixed effect for text significantly improved model fit (∆Χ2 = 23.559, *p* < 0.001), indicating a longer reading duration for the second text (*β* = 2.900) compared to the first one. Introducing a fixed effect for condition (∆Χ2 = 0.072, *p =* 0.789) as well as for group (∆Χ2 = 3.540, *p =* 0.170) did not significantly improved model fit. Thus, the average reading duration per sentence was best predicted by text.

For the analysis of the percentage of correct answers to the control questions, including a fixed effect for text (∆Χ2 = 1.89, *p =* 0.169) as well as for condition (∆Χ2 = 0.074, *p =* 0.785) did not significantly improved model fit. Introducing a fixed effect for group significantly improved model fit (∆Χ2 = 6.619, *p* = 0.037), indicating a more accurate responses for the TraPro (*β* = 0.126) and TraStu (*β* = 0.059) compared to the Mul. Thus, the percentage of correct responses to the control questions was best predicted by group. All LMM fixed effects of behavioral measurements in the reading task are summarized in Table 3.

### Table 3

*Summary of the LMM statistics of the reading task.*

**

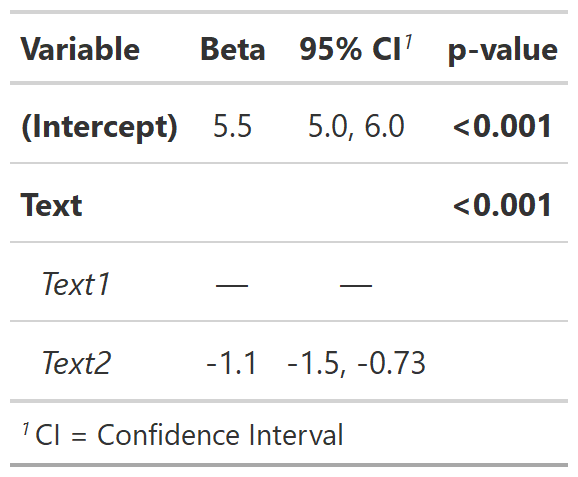
*Note.* Mul = multilingual control group, TraPro = professional translators, TraStu = translation students.

### Perceived difficulty of the reading and translation task

For the analysis of the perceived difficulty of the reading and translation task, including a fixed effect for task (∆Χ2 = 0.006, *p =* 0.941) did not significantly improve model fit. However, introducing a fixed effect for text significantly improved model fit (∆Χ2 = 28.166, *p* < 0.001), indicating that the second abstract was perceived as less difficult (*β* = -1.128) compared to the first one. Adding a fixed effect for condition (∆Χ2 = 0.001, *p =* 0.993) as well as for group (∆Χ2 = 3.125, *p =* 0.230) did not significantly improve model fit. Thus, the perceived difficulty of the reading and translation task was best predicted by text. The summary of all LMM fixed effects of the perceived difficulty is summarized in Table 4.

### Table 4

*Summary of the LMM statistics of the perceived difficulty.*



### Keystroke data

For the analysis of the total number of chars typed in the copying and translation tasks, including a fixed effect for task significantly improved model fit (∆Χ2 = 239.63, *p <* 0.001), indicating that the total number of chars typed was lower in the translating task (*β* = -490.28) compared to the copying task. Introducing a fixed effect for text (∆Χ2 = 0.299, *p =* 0.585), condition (∆Χ2 = 1.062, *p =* 0.303) as well as for group (∆Χ2 = 1.192, *p =* 0.551) did not significantly improved model fit. Thus, the total number of chars typed in the copying and translation task was best predicted by task.

For the analysis of the percentage of deletions in the copying and translation task, including a fixed effect for task significantly improved model fit (∆Χ2 = 93.565, *p <* 0.001), indicating that the percentage of deletions was higher in the translating task (*β* = 0.052) compared to the copying task. Introducing a fixed effect for text significantly improved model fit (∆Χ2 = 6.110, *p =* 0.013), revealing that the percentage of deletions was lower for the second abstract (*β* = -0.011) compared to the first abstract. Adding a fixed effect for condition significantly improved model fit (∆Χ2 = 4.067, *p =* 0.044) showing that the percentage of deletions was lower for the ELF version (*β* = -0.001) compared to the EdE version. Including a fixed effect for group (∆Χ2 = 1.254, *p =* 0.534) did not significantly improved model fit. However, modeling an interaction between task and text significantly improved model fit (∆Χ2 = 5.745, *p =* 0.017), whereas the interaction between text and condition (∆Χ2 = 0.116, *p =* 0.733), and task and condition did not (∆Χ2 = 0.069, *p =* 0.793). The interaction between task and text reflected a lower difference in the percentage of deletions between the copying and translation tasks for the second abstract (*β* = -0.021). Thus, the percentage of deletions in the copying and translation task was best predicted by task, text, condition, and the interaction between task and text. All LMM fixed effects of the keystroke data are summarized in Table 5.

### Table 5

*Summary of the LMM statistics of the keystroke data.*

**Ein Bild, das Text, Tisch enthält.

Automatisch generierte Beschreibung**

*Note.* EdE = edited English, ELF = English as lingua franca.

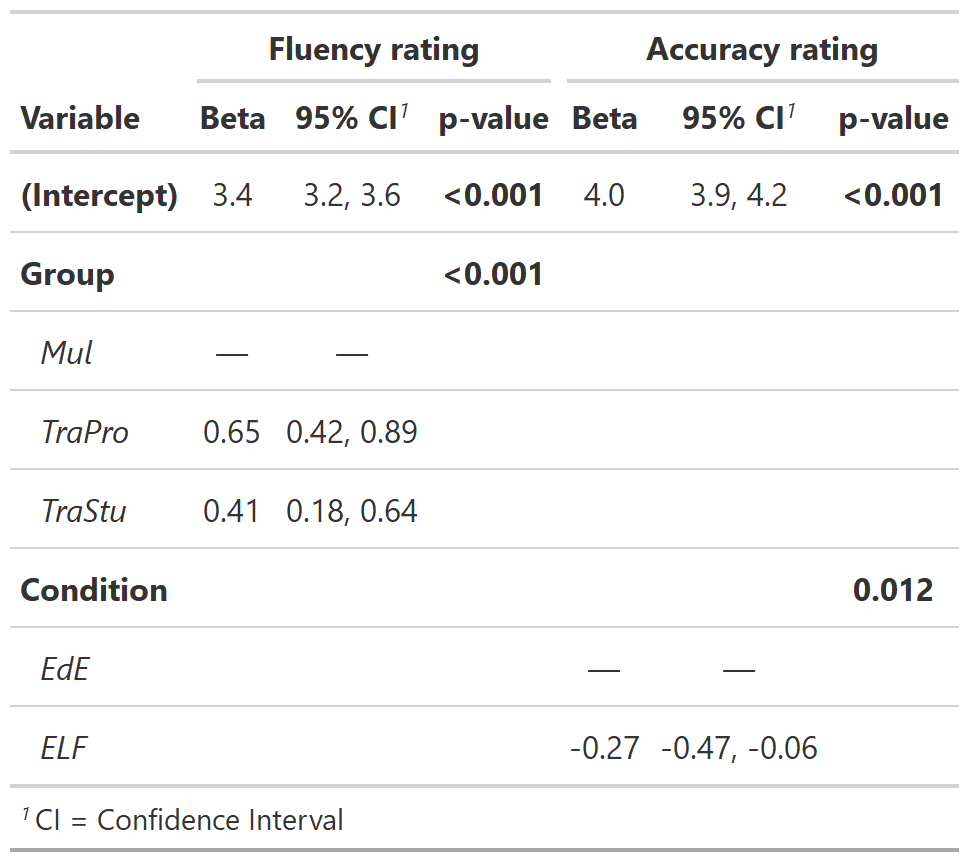
### Translation task

For the analysis of the fluency rating, including a fixed effect for text (∆Χ2 = 0.018, *p =* 0.893) as well as for condition (∆Χ2 = 0.035, *p =* 0.852) did not significantly improved model fit. Introducing a fixed effect for group significantly improved model fit (∆Χ2 = 25.768, *p <* 0.001), indicating that the translations of the group TraPro (*β* = 0.65) and TraStu (*β* = 0.41) were rated to be more fluent compared to the Mul group. Thus, the fluency rating was best predicted by group.

For the analysis of the accuracy rating, including a fixed effect for text (∆Χ2 = 0.171, *p =* 0.679) did not significantly improve model fit. Introducing a fixed effect for condition significantly improved model fit (∆Χ2 = 6.314, *p =* 0.012), indicating that the translations of the ELF version (*β* = -0.266) were rated to be less accurate compared to the EdE version. Adding a fixed effect for group (∆Χ2 = 1.850, *p =* 0.396) did not significantly improved model fit. Thus, the accuracy rating was best predicted by condition. All LMM fixed effects of the translation task are summarized in Table 6.

### Table 6

*Summary of the LMM statistics of the translation task.*

****

*Note.* EdE = edited English, ELF = English as lingua franca, Mul = multilingual control group, TraPro = professional translators, TraStu = translation students.

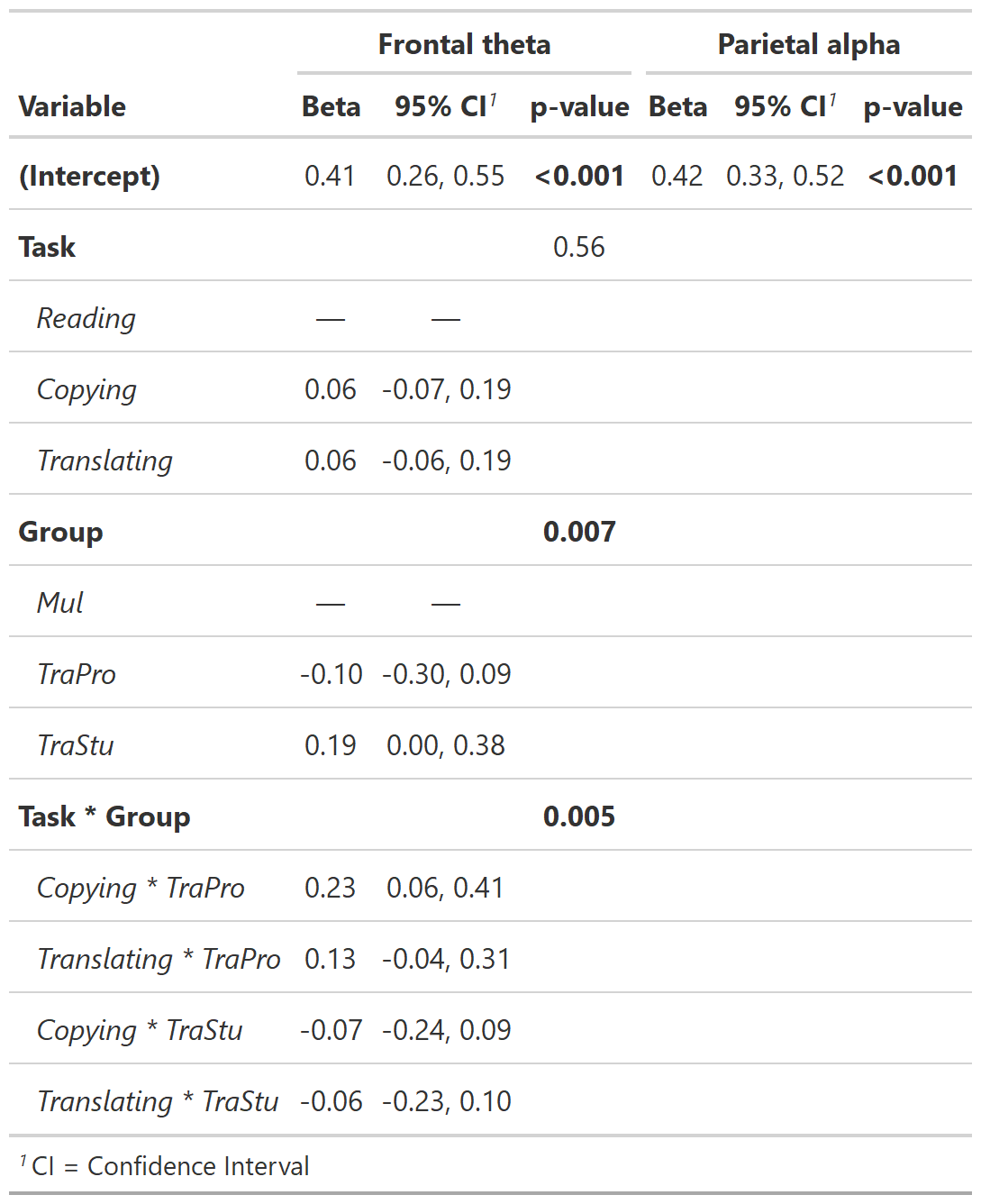
## EEG results

For the analysis of frontal theta power, including a fixed effect for task significantly improved model fit (∆Χ2 = 9.646, *p =* 0.008), indicating higher theta activity in the copying (*β* = 0.106) and translating tasks (*β* = 0.081) compared to the reading task. Introducing a fixed effect for text (∆Χ2 = 1.408, *p =* 0.235), condition (∆Χ2 = 0.310, *p =* 0.577) as well as for group (∆Χ2 = 3.640, *p =* 0.162) did not significantly improved model fit. However, adding an interaction between task and group significantly improved model fit (∆Χ2 = 14.623, *p =* 0.006). In the reading task, the groups of TraPro revealed lower theta power (*β* = -0.105) compared to the Mul group, whereas the group of TraStu revealed higher theta power (*β* = 0.186). In the copying task, the TraPro (*β* = 0.070) and TraStu (*β* = 0.070) were reflected by higher theta power compared to the Mul group. In the translating task, the TraPro (*β* = -0.030) showed lower theta power compared to the Mul group, whereas the TraStu group (*β* = 0.070) revealed higher theta power. Thus, the frontal theta activity was best predicted by task, group, and the interaction between group and task.

For the analysis of parietal alpha power, including a fixed effect for fixed effect for task (∆Χ2 = 4.223, *p =* 0.121), text (∆Χ2 = 0.650, *p =* 0.420), condition (∆Χ2 = 0.544, *p =* 0.461) as well as group (∆Χ2 = 0.174, *p =* 0.917) did not significantly improved model fit. Thus, the parietal alpha power was best predicted by the null model. All LMM fixed effects of the EEG parameters are summarized in Table 7 and the results for the theta power visualized in Figure 3.

### Table 7

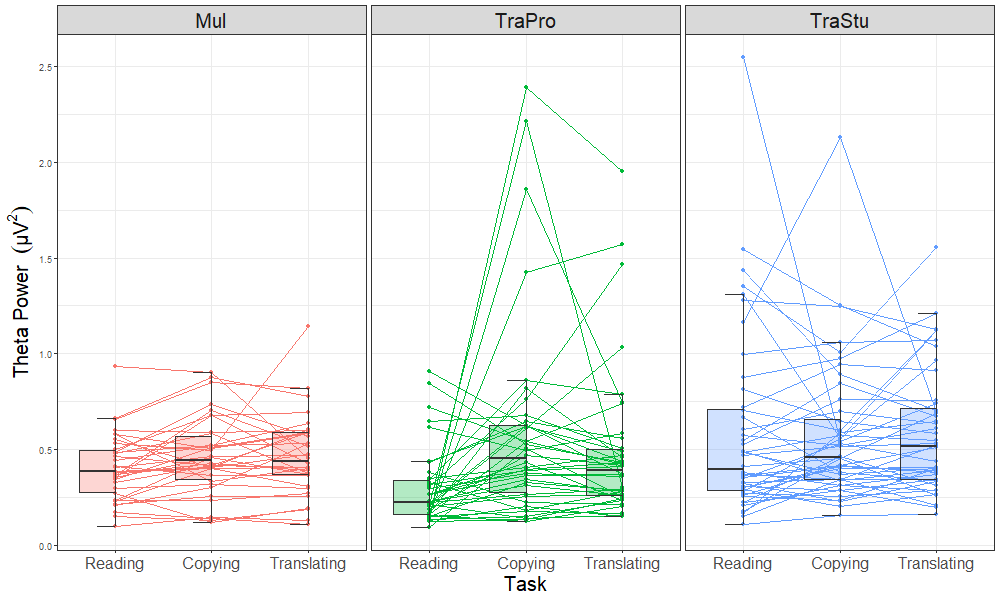
*Summary of the LMM statistics of the EEG data.*

**

*Note.* Mul = multilingual control group, TraPro = professional translators, TraStu = translation students. Intercept values reflect the Mul group in the reading task.

### Figure 3

*Boxplots and individual values of frontal theta power in the reading, copying, and translation tasks across all groups*

**

*Note.* Mul = multilingual control group, TraPro = professional translators, TraStu = student translators. Each participant is represented by two values, one from each abstract. Lines between points represent values of the same participant.

# Discussion

In the present EEG study, we assessed cognitive workload parameters as a function of written translation expertise using three different tasks, namely reading, copying, and backward translation. Furthermore, we provided the first neurophysiological assessment of text-level translation that also considers the cognitive demands associated with processing non-standard English inputs. The evaluation of average reading duration per sentence only revealed an effect of *text*, indicating that the second text was more complex to read than the first one. In line with this argumentation, the reading duration per sentence was longer in the second text, even though the number of words per sentence was smaller in the second one (text 1: 29.92, text 2: 20.89). Interestingly, against our expectation, we did not observe a difference in reading speed between the three groups. In fact, we expected that TraPros would need less time to complete the same amount of text compared to the less experienced TraStus and Mul. A possible explanation for the absence of group differences in reading speed could be the self-paced character of our reading task. Another hypothesis is that although task instructions were the same for all participants, the three groups could have used different reading strategies. In particular, since the participants were aware that they would have to translate the texts in the next task, it is conceivable that the TraPros not only read the texts but also prepared the respective translations, which is a time-consuming process (Macizo & Bajo, 2006). However, this interpretation somewhat contrasts with the examination of the control questions after the reading task, which revealed that the TraPro performed better than TraStu and Mul, irrespective of text or the text version. Taken together, these results might indicate that TraPros could gather more information during the same time, leading to a higher percentage of correct answers to the control questions.

The behavioral metrics of the copying and translating task included the overall estimation of efficiency, quantified by the total number of chars typed, the percentage of deletions, as well as the ratings of the fluency and accuracy of the text output generated in the translating task. In line with our expectation, the overall output was much lower in the translating compared to the copying task, as indicated by the reduced efficiency. This result was not surprising because transcribing a text from English to English is much less demanding than translating from English to German. Interestingly, we did not observe an effect of *group,* although we expected TraPro would be more efficient than TraStu and Mul, especially in the translating task. This finding was surprising considering previous workplace studies showing that TraPros commonly outperform beginners and master students in translation regarding the number of keystrokes (Ehrensberger-Dow & Massey, 2014). In contrast, the eye-tracking study of Carl and Klay (2012) revealed that professional translators did not differ from student translators in the number of characters typed per production unit. However, professionals were able to focus their gaze on the source text while students shifted their attention more often from the source to the target text. Moreover, translator students’ typing behavior was more fragmented and hence associated with slower performance. Nevertheless, we may speculate that the absence of an effect of expertise on efficiency in our data might be due to the particular setting we used in association with EEG. Even though we attempted to apply a realistic and ecologically valid setting where the participants translated sentence by sentence (like the often-used Translog software), they had to work with an unfamiliar keyboard, an EEG cap on the head, a chin rest to ensure a stable head position while at the same time producing texts within a MATLAB script that does not exactly behave as other typewriting programs. For example, with the procedure we used, it was impossible to mark parts of the text with the mouse to modify it. Thus, the performance benefits of TraPro could be limited to the familiar environment of their keyboard or workplace studies.

Furthermore, the percentage of deletions, which is the number of backslashes relative to the number of keystrokes, can be used as an index task complexity. The more complex a task is, the more often a person is expected to use the backslash. Therefore, it is not surprising that the percentage of deletions was higher in the translating task than in the copying task. Interestingly, the text input type impacted the percentage of deletions, with the ELF condition characterized by fewer deletions than the EdE condition, contradicting our expectation that non-native text inputs are more challenging to transcribe or translate. Accordingly, we may speculate that the participants were more challenged by the ELF condition, which might have led to a reduced awareness of typos and lesser use of backslashes. Unfortunately, we could not to check if there were more remaining typos in the ELF compared to the EdE versions to support our claim.

The evaluation of the fluency rating of the text outputs yielded a main effect of *group* indicating a higher fluency of translated sentences in TraPro compared to TraStu and Mul. Even though the inter-rater reliability was moderate, possibly due to the challenge of rating fluency on a sentence level, our results accurately reflected the higher level of expertise in TraPro and TraStu. Otherwise, the accuracy rating did not reveal an effect of *group* as we would have expected it. However, this might have been related to the control participants being also highly proficient bilinguals. Hence, we speculate that the advantage in the performance of experienced translators might only be manifested when evaluating text-level translations. In our setting, the participants had to translate sentence-wise and only for 5 minutes. Such a procedure resulted in translating only a few sentences at the beginning of the text, which might not have favored professional translators. Typically, professional translators work repeatedly on the target text to ensure consistency. In addition, it is noteworthy that we instructed our participants not to delete large amounts of produced text if they were not satisfied with their translation. The reason for such an instruction was due to the limited duration of the translation task and the impossibility of using mouse clicks to navigate through the produced text and modify it. Finally, it is interesting to denote that our results indicated an overall lower accuracy when translating ELF input. Therefore, we provide first evidence that translating non-native text inputs might be more challenging, as reflected by a lower accuracy in the target texts. Based on our data, we can only speculate that the lower performance of our participants, including TraPros, was possibly driven by time pressure that prevented them from doing more plausibility checks and compensation loops (Ehrensberger-Dow et al., 2020).

The examination of perceived difficulty, which was assessed using a Likert scale, revealed a main effect of the text. This main effect originated from the fact that the second abstract was generally perceived as less challenging, whether written in ELF or EdE or whether the participants had to read or translate. This finding somewhat contradicts the longer reading times we revealed in the second abstract, which we interpreted as an indicator of task demands.

In the present study, we used EEG as an objective marker of workload assessment and expected frontal theta power to increase with workload, whereas parietal alpha power will decrease. We observed a significant interaction between group and task for frontal theta power but not for parietal alpha power. In the reading task, the TraPro group demonstrated the lowest theta power, whereas the TraStu group was associated with the highest one. With the finding that TraPros outperformed the other two groups in the control questions after the reading task, the lower theta power in the TraPro group could be interpreted as supporting the neural efficiency hypothesis by Haier and colleagues (Haier et al., 1988). Such effects of neural efficiency have mostly been found in frontal areas (Neubauer and Fink, 2009) and in tasks with low to moderate difficulty (Nussbaumer, Grabner, & Stern, 2015). Further evidence for this perspective comes from the translation task. TraPro had a slightly lower theta power than the other groups and outperformed them in terms of fluency rating of the output texts. Notably, our results also indicated an increase of frontal theta power from the reading to the copying and the translating task. However, this was only the case in the TraPro and Mul but not in the TraStu group, indicating higher task demands associated with producing an output text compared to only reading. Interestingly, we did not find an increase in frontal theta power in the translation compared to the copying tasks within the Mul group, casting some doubts on the sensitivity of frontal theta power to assess workload. In the TraStu group, there was also a minor increase from the copying task to the translating task. However, the TraPro group was additionally accompanied by a decrease in frontal theta power in the translating compared to the copying task, and this would be in line with the efficiency hypothesis mentioned above.

### Limitations

In the present study, we thought to use a realistic and ecologically valid setting to bridge the gap between research on single-word and text-level translation. However, as mentioned above, we could not implement a translation setting in Psychtoolbox that was identical to the familiar Translog software. Furthermore, professional translators are commonly used to improve their first draft, which was not possible in our setting. Moreover, the time limit of 5 minutes for translation did not allow to fully assess the performance advantages of professional translators at the text level. Another limitation is that based on the experimental approach we used, we were not able to disentangle the possible influence of age on behavioral and electrophysiological measurements. However, since expertise is directly related to age, it is challenging to determine the influence of this factor. Finally, it should be mentioned that a main challenge of research on bilingualism is the vast heterogeneity within the samples (Calvo et al., 2016).

### Conclusions

In the present study, we attempted to assess cognitive workload in three subtasks related to text-level translation by measuring three groups of participants with varying levels of expertise. Moreover, we also considered the cognitive demands associated with processing non-native source texts (ELF). Professional translators outperformed student translators and multilingual controls in the reading comprehension questions placed after the reading task, and this effect was accompanied by lower theta band power, although we did not find evidence for an influence of expertise on reading speed. Furthermore, our results showed that professional translators demonstrated higher fluency but not accuracy ratings in the translation task and were characterized by lower theta band power. Otherwise, we did not find evidence for an effect of expertise on typing speed in both the copying and the translation tasks. Importantly, we also provided first evidence indicating that translating ELF texts leads to a lower accuracy of the target texts in professional translators, student translators, and highly proficient bilinguals. However, this behavioral effect was not paralleled by distinctive neurophysiological metrics.

# Declarations of interest:

The authors declare no potential sources for conflicts of interest.

# Author contributions:

Matthias Kobi: Conceptualization, Methodology, Formal Analysis, Investigation, Data Curation, Project administration, Writing – Original draft

Michael Boos: Conceptualization, Methodology, Project administration, Writing – Review & Editing

Stefan Elmer: Conceptualization, Methodology, Funding acquisition, Writing – Review & Editing, Supervision

Lutz Jäncke: Funding acquisition, Writing – Review & Editing, Supervision

# Acknowledgement:

The research was supported by the Swiss National Science Foundation (SNF, grant nr. CRSII5\_173694 to Lutz Jäncke). We like to thank everyone of the University of Zurich and the Zurich University of Applied Sciences involved in the project, especially M. Motadel and D. Torosjan, who contributed to the introduction with their bachelor thesis.

# Data accessibility:

Behavioral data and analysis scripts are accessible on <https://github.com/mkobi89/LDT>. EEG raw and preprocessed data can be made available upon request to the authors.

# References

Aasman, J., Mulder, G., & Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors*, *29*(2), 161–170. https://doi.org/10.1177/001872088702900204

Albl-Mikasa, M. (2013). ELF speakers’ restricted power of expression: Implications for interpreters’ processing. *Translation and Interpreting Studies. The Journal of the American Translation and Interpreting Studies Association*, *8*(2), 191–210. https://doi.org/10.1075/tis.8.2.04alb

Albl-Mikasa, M., Guggisberg, S., & Talirz, F. (2017). (Source) texting ELF-native and non-native English speaker discourse production and conference interpreters’ preference for the native speaker. In *Translata II: «Translation Studies & Translation Practice», Innsbruck, Austria, 30 October-1 November 2014* (pp. 276–275). Peter Lang.

Annett, M. (1970). A classification of hand preference by association analysis. *British Journal of Psychology*, *61*(3), 303–321. https://doi.org/10.1111/j.2044-8295.1970.tb01248.x

Antonenko, P., Paas, F., Grabner, R., & Van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review*, *22*(4), 425–438. https://doi.org/10.1007/s10648-010-9130-y

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual Word Recognition: The Journey From Features to Meaning (A Travel Update). In *Handbook of Psycholinguistics* (pp. 285–375). Academic Press. https://doi.org/10.1016/B978-012369374-7/50010-9

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, *95*, 173–189. https://doi.org/10.1016/j.jml.2017.04.002

Berka, C., Levendowski, D. J., Lumicao, M. N. M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation Space and Environmental Medicine*, *78*(5), 232–244.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience and Biobehavioral Reviews*, *44*, 58–75. https://doi.org/10.1016/j.neubiorev.2012.10.003

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, *42*(3), 361–377. https://doi.org/10.1016/0301-0511(95)05167-8

Calvo, N., García, A. M., Manoiloff, L., & Ibáñez, A. (2016). Bilingualism and cognitive reserve: A critical overview and a plea for methodological innovations. *Frontiers in Aging Neuroscience*, *7*, 249. https://doi.org/10.3389/fnagi.2015.00249

Carl, M., & Kay, M. (2012). Gazing and typing activities during translation: A comparative study of translation units of professional and student translators. *Meta*, *56*(4), 952–975. https://doi.org/10.7202/1011262ar

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Christoffels, I. K., Ganushchak, L., & Koester, D. (2013). Language conflict in translation: An ERP study of translation production. *Journal of Cognitive Psychology*, *25*(5), 646–664. https://doi.org/10.1080/20445911.2013.821127

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

de Cheveigné, A. (2020). ZapLine: A simple and effective method to remove power line artifacts. *NeuroImage*, *207*. https://doi.org/10.1016/j.neuroimage.2019.116356

de Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1001–1018. https://doi.org/10.1037/0278-7393.18.5.1001

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Ehrensberger-Dow, M., Albl-Mikasa, M., Andermatt, K., Hunziker Heeb, A., & Lehr, C. (2020). Cognitive load in processing ELF: Translators, interpreters, and other multilinguals. *Journal of English as a Lingua Franca*, *9*(2), 217–238. https://doi.org/10.1515/jelf-2020-2039

Ehrensberger-Dow, M., & Massey, G. (2014). Cognitive ergonomic issues in professional translation. *The Development of Translation Competence : Theories and Methodologies from Psycholinguistics and Cognitive Science*, 58–86.

García, A. M., & Ibáñez, A. (2016). Processes and verbs of doing, in the brain. *Functions of Language*, *23*(3), 305–335. https://doi.org/10.1075/fol.23.3.02gar

García, A. M., Ibáñez, A., Huepe, D., Houck, A. L., Michon, M., Lezama, C. G., Chadha, S., & Rivera-Rei, Á. (2014). Word reading and translation in bilinguals: The impact of formal and informal translation expertise. *Frontiers in Psychology*, *5*, 1–14. https://doi.org/10.3389/fpsyg.2014.01302

García, A. M., Mikulan, E., & Ibáñez, A. (2016). A neuroscientific toolkit for translation studies. In R. Muñoz Martín (Ed.), *Reembedding translation process research* (pp. 21–46). John Benjamins. https://doi.org/10.1075/btl.128.02gar

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human–computer interaction. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 113–131. https://doi.org/10.1080/14639220210159717

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, *7*(4), 374–385. https://doi.org/10.1093/cercor/7.4.374

Grabner, R. H., Brunner, C., Leeb, R., Neuper, C., & Pfurtscheller, G. (2007). Event-related EEG theta and alpha band oscillatory responses during language translation. *Brain Research Bulletin*, *72*(1), 57–65. https://doi.org/10.1016/j.brainresbull.2007.01.001

Hampson, R. E., Opris, I., & Deadwyler, S. A. (2010). Neural correlates of fast pupil dilation in nonhuman primates: Relation to behavioral performance and cognitive workload. *Behavioural Brain Research*, *212*(1), 1–11. https://doi.org/10.1016/j.bbr.2010.03.011

Hancock, P. A., Meshkati, N., & Robertson, M. M. (1985). Physiological reflections of mental workload. *Aviation Space and Environmental Medicine*, *56*(11), 1110–1114.

Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2021). Detection theory: A user’s guide. In *Routledge*. https://doi.org/10.4324/9781003203636

Hogervorst, M. A., Brouwer, A. M., & van Erp, J. B. F. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in Neuroscience*, *8*, 322. https://doi.org/10.3389/fnins.2014.00322

Holm, A., Lukander, K., Korpela, J., Sallinen, M., & Müller, K. M. I. (2009). Estimating brain load from the EEG. *TheScientificWorldJournal*, *9*, 639–651. https://doi.org/10.1100/tsw.2009.83

Ibáñez, A. J., Macizo, P., & Bajo, M. T. (2010). Language access and language selection in professional translators. *Acta Psychologica*, *135*(2), 257–266. https://doi.org/10.1016/j.actpsy.2010.07.009

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *European Journal of Neuroscience*, *15*(8), 1395–1399. https://doi.org/10.1046/j.1460-9568.2002.01975.x

Jost, L. B., Radman, N., Buetler, K. A., & Annoni, J. M. (2018). Behavioral and electrophysiological signatures of word translation processes. *Neuropsychologia*, *109*, 245–254. https://doi.org/10.1016/j.neuropsychologia.2017.12.034

Klein, D., Milner, B., Zatorre, R. J., Meyer, E., & Evans, A. C. (1995). The neural substrates underlying word generation: A bilingual functional- imaging study. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(7), 2899–2903. https://doi.org/10.1073/pnas.92.7.2899

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., & Cornelissen, F. (2007). *What’s new in Psychtoolbox-3?* https://pure.mpg.de/rest/items/item\_1790332/component/file\_3136265/content

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007

Koehn, P., & Monz, C. (2006). *Manual and Automatic Evaluation of Machine Translation between European Languages 1 Evaluation Framework*. 102–121. http://www.statmt.org/wmt06/

Kohlisch, O., & Schaefer, F. (1996). Physiological changes during computer tasks: Responses to mental load or to motor demands? *Ergonomics*, *39*(2), 213–224. https://doi.org/10.1080/00140139608964452

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Macizo, P., & Bajo, M. T. (2006). Reading for repetition and reading for translation: Do they involve the same processes? *Cognition*, *99*(1), 1–34. https://doi.org/10.1016/j.cognition.2004.09.012

MATLAB. (2018). *version 9.5.0.944444 (R2018b)*. The MathWorks Inc.

Mills, C., Fridman, I., Soussou, W., Waghray, D., Olney, A. M., & D’Mello, S. K. (2017). Put your thinking cap on: Detecting cognitive load using EEG during learning. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 80–89). https://doi.org/10.1145/3027385.3027431

Muñoz, E., Calvo, N., & García, A. M. (2018). Grounding translation and interpreting in the brain: what has been, can be, and must be done. *Perspectives*, *27*(4), 483–509. https://doi.org/10.1080/0907676X.2018.1549575

Pedroni, A., Bahreini, A., & Langer, N. (2019). Automagic: Standardized preprocessing of big EEG data. *NeuroImage*, *200*, 460–473. https://doi.org/10.1016/j.neuroimage.2019.06.046

Pérez, G., Hesse, E., Dottori, M., Birba, A., Amoruso, L., Martorell Caro, M., Ibáñez, A., & García, A. M. (2022). The bilingual lexicon, back and forth: Electrophysiological signatures of translation asymmetry. *Neuroscience*, *481*, 134–143. https://doi.org/10.1016/j.neuroscience.2021.11.046

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *Quarterly Journal of Experimental Psychology*, *60*(2), 211–229. https://doi.org/10.1080/17470210600673818

Quaresima, V., Ferrari, M., Van Der Sluijs, M. C. P., Menssen, J., & Colier, W. N. J. M. (2002). Lateral frontal cortex oxygenation changes during translation and language switching revealed by non-invasive near-infrared multi-point measurements. *Brain Research Bulletin*, *59*(3), 235–243. https://doi.org/10.1016/S0361-9230(02)00871-7

Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: A field study and simulation validation. *Ergonomics*, *54*(10), 932–942. https://doi.org/10.1080/00140139.2011.604431

Rinne, J. O., Tommola, J., Laine, M., Krause, B. J., Schmidt, D., Kaasinen, V., Teräs, M., Sipilä, H., & Sunnari, M. (2000). The translating brain: cerebral activation patterns during simultaneous interpreting. *Neuroscience Letters*, *294*(2), 85–88. https://doi.org/10.1016/S0304-3940(00)01540-8

Sammer, G., Blecker, C., Gebhardt, H., Bischoff, M., Stark, R., Morgen, K., & Vaitl, D. (2007). Relationship between regional hemodynamic activity and simultaneously recorded EEG-theta associated with mental arithmetic-induced workload. *Human Brain Mapping*, *28*(8), 793–803. https://doi.org/10.1002/hbm.20309

So, W. K. Y., Wong, S. W. H., Mak, J. N., & Chan, R. H. M. (2017). An evaluation of mental workload with frontal EEG. *PLoS ONE*, *12*(4), 1–17. https://doi.org/10.1371/journal.pone.0174949

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stipacek, A., Grabner, R. H., Neuper, C., Fink, A., & Neubauer, A. C. (2003). Sensitivity of human EEG alpha band desynchronization to different working memory components and increasing levels of memory load. *Neuroscience Letters*, *353*(3), 193–196. https://doi.org/10.1016/j.neulet.2003.09.044

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory* (pp. 29–47). Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511844744.004

Swerdloff, M. M., & Hargrove, L. J. (2020). Quantifying cognitive load using EEG during ambulation and postural tasks. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 2849–2852). IEEE. https://doi.org/10.1109/EMBC44109.2020.9176264

Vassileiou, B., Meyer, L., Beese, C., & Friederici, A. D. (2018). Alignment of alpha-band desynchronization with syntactic structure predicts successful sentence comprehension. *NeuroImage*, *175*, 286–296. https://doi.org/10.1016/j.neuroimage.2018.04.008

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, *41*(5), 656–669. https://doi.org/10.1080/001401398186829

Vogt, J., Hagemann, T., & Kastner, M. (2006). The impact of workload on heart rate and blood pressure in en-route and tower air traffic control. *Journal of Psychophysiology*, *20*(4), 297–314. https://doi.org/10.1027/0269-8803.20.4.297

von Studnitz, R. E., & Green, D. W. (2002). Interlingual homograph interference in German-English bilinguals: Its modulation and locus of control. *Bilingualism: Language and Cognition*, *5*(1), 1–23. https://doi.org/10.1017/s1366728902000111

Waldmann, H. C. (2008). Kurzformen des HAWIK-IV: statistische Bewertung in verschiedenen Anwendungsszenarien. *Diagnostica*, *54*(4), 202–210. https://doi.org/10.1026/0012-1924.54.4.202