**An examination of how translation experience influences cognitive load in a real-world situation using EEG**

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

1Division Neuropsychology, Department of Psychology, University of Zurich, Zurich, Switzerland.

2University Research Priority Program (URPP) “Dynamics of Healthy Aging”, University of Zurich, Zurich, Switzerland.

\* Shared last authorship

Keywords:

Cognitive load, translation, language expertise, EEG

E-mail addresses:

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

MB: m.boos@psychologie.uzh.ch

SE: s.elmer@psychologie.uzh.ch  
LJ: 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: 27 March, 2022

**Abstract**

**Introduction**

*Importance of bilingualism*

With the ongoing globalization in the past decades, the interest in research of language mastery has vastly increased. This is not just caused by expanded immigration across the globe but also due to the rising collaboration in science, economy, politics, and other fields between different countries. Thus, more and more professional translators are in demand. Furthermore, there is an increasing number of multilinguals due to immigration (Byers-Heinlein et al., 2019; Luk, 2017). As a result, understanding how language is processed in the brain, as well as the difference between professional translators and untrained bilinguals, becomes progressively relevant in research.

Most of the research on translation studies and between-language activation focused on the processing speed of single-word features like the cognate status, the level of concreteness, the level of frequency, and 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). A well-known effect at the text level refers to differences in the translation process caused by direction. More specifically, forward translation (FT) from first (L1) to second language (L2) is typically found to have shorter response times than backward translation (BT) from L2 to L1 (Muñoz et al., 2019). Neuroimaging studies found that BT and FT are partly based on different mechanisms and neuronal activities (García & Ibáñez, 2016; Klein et al., 1995; Quaresima et al., 2002; Rinne et al., 2000). Moreover, FT is thought to require more cognitive effort than BT and thus place more demands on executive control (García et al., 2016)*.* Furthermore, an interesting aspect that has often been neglected in studies of multilingual language processing is the type of English used in input texts (e.g., Albl-Mikasa, 2013; Albl-Mikasa, Guggisberg, & Talirz, 2017). A consequence of globalization is the increasing number of English texts in a wide variety of fields, which are written by authors of different native languages and English as a second language for communication. Such a proliferation of English as a lingua franca (ELF) has the disadvantage that this factor could make translating texts more difficult as it might be ambiguous, incoherent, and imprecise on the textual level leading to additional cognitive demands like plausibility checks and compensation loops for translators, as shown in recent workplace studies (Ehrensberger-Dow et al., 2020). Therefore, the impaired bottom-up processing likely complicates basic cognitive processes of translating. Greater reliability in top-down processing and higher-order cognitive processes is needed (Albl-Mikasa, 2013), suggesting that ELF is increasing cognitive load, at least for professional translators.

Although the research on translating and bilingualism is relatively new and has faced challenges in the past, some studies found that professional translators show changes in neuronal pathways and connections as well as lexical processing with increasing language expertise, respectively, 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). Professional translators have been found to outperform non-translators or translation students in various linguistic-related tasks, for example, semantic error detection (Fabbro et al., 1991; Yudes et al., 2013), reading speed, the lexical decision for non-words, and categorization of atypical examples (Bajo et al., 2000). Regarding non-linguistic tasks, more skilled and experienced translators seem to be able to manage more global aspects of a task (e.g., Ehrensberger-Dow, 2014; Ehrensberger-Dow & Massey, 2013; Heeb, 2016). García et al. (2014) studied the impact of translation expertise during reading aloud in L1 and L2, as well as BT and FT translating noun pairs from a single word list, manipulating lexical variables such as concreteness and cognate status. Their second experiment tested 36 native Spanish speakers with high proficiency levels in English as L2. The participants were divided into three groups of 12 persons based on their expertise: First-year translation students, senior-year translation students, and professional translators with field experience. The results underline that, in general, effects of concreteness and cognate status on response times were visible, reading in L1 was faster than in L2, reading was faster than translating, but the predicted advantage of BT over FT was not visible. Moreover, irrespective of L2 proficiency, translation expertise affected lexical processing, as translators with field experience and senior-year students responded faster than first-year translation students. Overall, the influence of translation expertise was greater in word translation than in reading. In a study by Ibáñez et al. (2010), the effect of formal training in translation on lexical access and language switching costs was addressed in two reading experiments. Both experiments tested twelve translators with more than two years of translating experience and twelve bilinguals matched in age, language fluency, and working memory capacity. Both experiments manipulated L1 (Spanish) and L2 (English), the cognate status, and the switching condition. 84 cognate and non-cognate words were randomly placed in both English and Spanish sentences, but never as the first or last word. The sentences appeared word-by-word, and the participant’s self-paced reading speed was taken as an index of the processing time. Furthermore, in experiment 1, the participants had to repeat the presented sentence aloud after reading before continuing to read the next sentence, whereas, in experiment 2, there was no repeating task. The results revealed that the cognate status affected the response time of translators, whereas bilinguals did not show any significant difference (experiment 1). Interestingly, the reading times in experiment 1 were slower in translators suggesting a different strategy of translators possibly activating both languages that may have led to higher attentional demands and the slowdown. However, experiment 2 demonstrated that both groups showed an effect of the cognate status, but there was no difference in reading times. Concerning language switching, professionals in experiment 1 did not display asymmetrical language switching costs, while bilinguals showed an effect when switching from L2 to their dominant L1. However, in experiment 2, neither bilinguals nor translators showed overall switching costs. The authors conclude a remarkable ability of professional translators to separate two languages during comprehension and production and constantly switching between them. An interesting study by Carl and Kay (2012) examined the focus of attention during translating. They investigated the gazing and typing activities (UAD) of 12 professional and student translators, each with Danish as L1 and English as L2 while translating an English text of 160 words into Danish (BT). The authors applied the data-acquisition software Translog, often used for computer-based reading and writing experiments, and a remote eye-tracker with the GWM software to capture the reading behavior. During the experiment, the source text was presented in the upper part of the monitor, and the target text was visible in the lower part of the monitor. In this way, the typing and translation speed could be observed and whether the participants’ gaze was fixed on the source or target text. The results showed that professionals simultaneously focused their gaze on the source text while producing the target text during translating. In contrast, the students sequentially shifted their attention from one text to another. Although the typing speed in terms of characters per time within the production of the target text for both groups was similar, students were slower in producing the target text as their typing behavior was more fragmented and with more and longer pauses than professionals.

Language processing during reading and translation is highly complex and relies on numerous subprocesses distributed across the brain. Those processes can be divided into feature- (i.e., recognizing visual patterns, letters), sublexical- (onsets, rhymes, syllables, morphemes), lexical- (length, word frequency, familiarity, orthographic and phonological neighborhood), semantic- (lexicality, concreteness/imageability, meaningfulness), syntactic level (see Balota et al. (2006) for a review), as well as higher cognitive functions associated with storing and accessing information, creating associations, alertness, selective or vigilant attention, controlling, inhibition, error-correction, and executive functions such as planning, setting goals and decision making (Indefrey & Levelt, 2004; Jäncke, 2017).

Taken together, we expect the neurophysiological brain response during reading and translation to affect mental workload measurable by electroencephalographic metrics. Based on the cognitive load theory (Sweller, 2010), a close relation between working memory capacity and the cognitive effort required for a task is assumed (Anderson, 2011). Processing new information claims more working memory capacity and is associated with a higher cognitive load (Antonenko et al., 2010). Moreover, available working memory capacity and cognitive load are inversely related (Anderson et al., 2011). Thus, cognitive load indicates the occupied space of working memory (Mills et al., 2017). One aspect of cognitive load is the intrinsic cognitive load, which reflects the mental effort that arises with the difficulty of an underlying task (Anderson et al., 2011).

In contrast to subjective measurements, objective measurements targeting brain activity have been labeled as highly advantageous because they directly correlate with workload (Hogervorst et al., 2014). Among the psychophysiological measurements including electrodermal activity (Kohlisch & Schaefer, 1996; Reimer & Mehler, 2011), eye movements (Brookings et al., 1996; Veltman & Gaillard, 1998) or pupil size (Hampson et al., 2010; Porter et al., 2007), heart rate (Vogt et al., 2006) or heart rate variability (Aasman et al., 1987; Hancock et al., 1985), the EEG has been found to be the most promising technique as it is believed to reflect cognitive load most sensitively (Mills 2017) (Hogervorst et al., 2014). Furthermore, using a battery of different cognitive tests, Berka and colleagues (2007) reported that the obtained EEG indices of workload and engagement correlated with subjective and objective performance data and therefore seem to deliver a good image of online cognitive load. The EEG measures cognitive load directly (Swerdloff & Hargrove, 2020) and with high temporal precision (Hunter, 2020). Of particular interest are event-related desynchronization (ERD) and synchronization (ERS), which have been associated with cognitive demands (Hunter, 2020). In fact, numerous studies reported a synchronization in frontal theta power with increasing cognitive load using working memory tasks (i.e., Berka et al., 2007; Borghini et al., 2014; Cavanagh & Frank, 2014; Chuang, Huang, & Hung, 2013; Friese et al., 2013; Gevins & Smith, 2003; Gevins et al., 1998; Gevins, Smith, McEvoy, & Yu, 1997; Holm, Lukander, Korpela, Sallinen, & Müller, 2009; Ishii et al., 1999; Jensen & Tesche, 2002; Sammer et al., 2007; So, Wong, Mak, & Chan, 2017). In contrast, parietal alpha power has been shown to desynchronize with increasing cognitive load (i.e. Gevins et al., 1997; W. Klimesch, Doppelmayr, Pachinger, & Russegger, 1997; W. Klimesch, Doppelmayr, Russegger, Pachinger, & Schwaiger, 1998; Wolfgang Klimesch, 2012; Stipacek, Grabner, Neuper, Fink, & Neubauer, 2003; Vassileiou, Meyer, Beese, & Friederici, 2018)

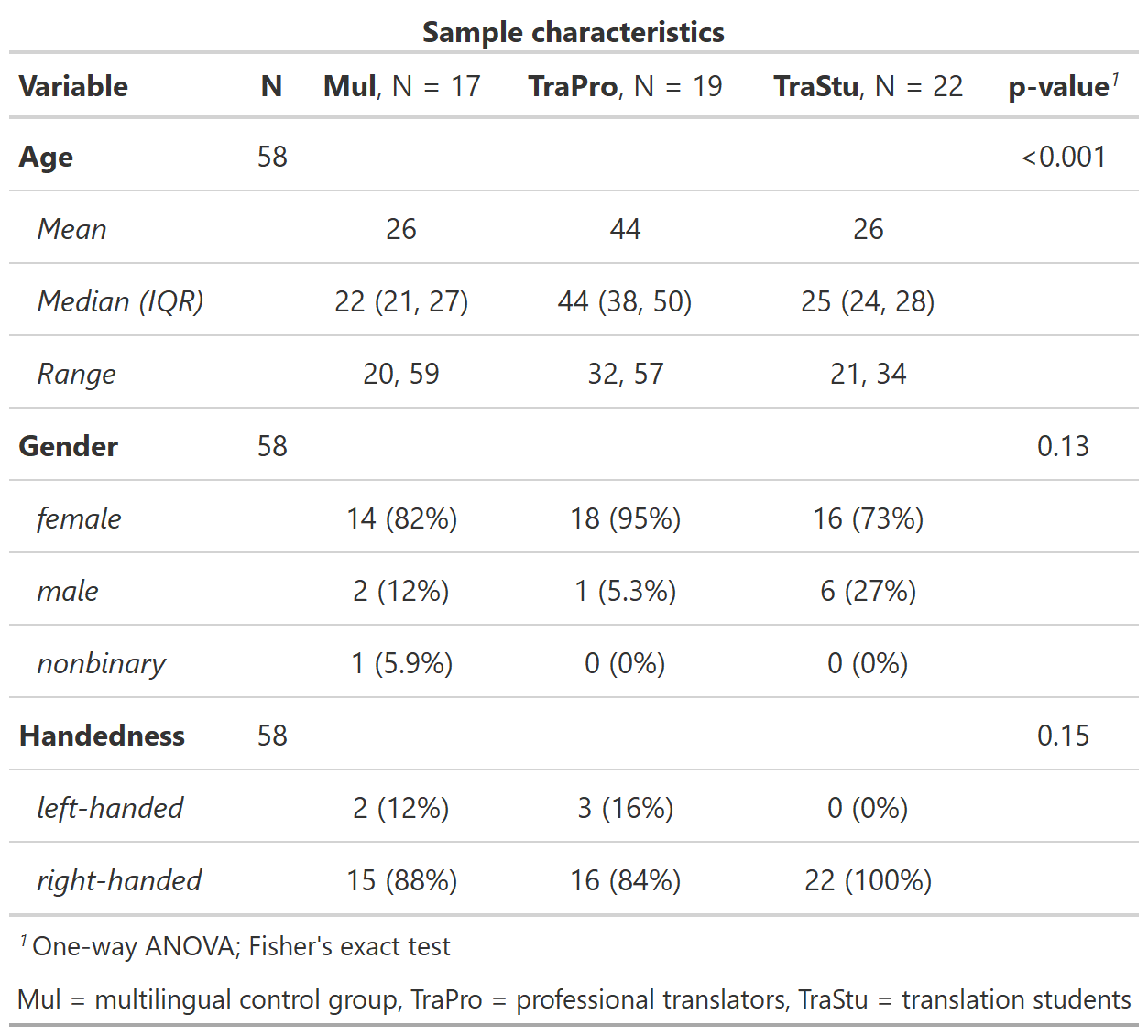
Recent studies focusing on brain oscillations during language-processing tasks found the theta and alpha bands to be responsive during semantic information processing, and the alpha band during cognitive memory performance (Bastiaansen & Hagoort, 2006; Klimesch, 1999). Other studies assessing the dynamics between L1 and L2 observed oscillatory responses in the alpha, beta, and theta bands (Bastiaansen et al., 2010; Birba et al., 2020; Vilas et al., 2019; Weiss et al., 2005). The study of Pérez et al. (2022) focussed on ERS and ERD while reading in L1, and L2, and performing FT and BT, measuring 27 early bilinguals with Spanish as L1 and high proficiency in English as L2. The reading tasks were included to test potential differences in single-language processes compared to translation. The stimulus set comprised 192 nouns for each language and was matched equally for cognate status and concreteness. The single nouns from a word list were presented for 200 ms on a computer screen, whereas the participants had to read them aloud as quickly as possible in the reading task and translate them in the translation task. The participants were instructed to press a key when they were ready to give their response. The results showed that relative to BT, FT was performed slower. Moreover, during the early time window (0-300 ms), FT produced a higher theta (4-7 Hz) power in frontal sites and lower frontoparietal connectivity in the lower-alpha (>10 Hz) band. Whereas in a later window (300-600 ms), reduced lower-beta (14-20 Hz) in centro-posterior sites and upper-beta (21-30 Hz) power in centro-frontal sites were observed. Moreover, Grabner et al. (2007) focussed on ERD and ERS patterns during language translation with different difficulty levels induced through low- and high-frequency English words in a simple word task. A pool of 13 advanced translation and interpreting students with German as L1 and English as L2 were recruited. The stimulus material consisted of 50 non-words, high-, and low-frequency words each, matched for concreteness, the number of phonemes, and the number of letters. The participants were asked to find the appropriate German translation of single words and typewrite the translation on the screen. The results showed that compared to high-frequency words, low-frequency words were translated slower, displayed a higher parietal theta (4-7 Hz) ERS at 300-600 ms, and frontal upper-alpha (10-13 Hz) ERD at 300-500 ms. Successfully translated low-frequency words showed stronger ERD in the lower-alpha (7-10 Hz) and upper-alpha (10-13 Hz) band than not translated words. Moreover, it was observed that beta (20-30 Hz) ERD around 400 ms was more pronounced for high- than low-frequency words.

This study aimed to shed more light on the influence of expertise in translation in the most ecologically valid setting possible. For this purpose, we invited a relatively large sample of professional translators, translation students, and a bilingual control group to have them read, transcribe, and translate texts. In doing so, we measured various psychometric and behavioral data and the EEG’s cognitive workload. The participants were native speakers of German who used English as their primary working language. In our study, we assessed both a measure for L2 proficiency in an English test and indicators for translation expertise through self-reports. As stimulus material, we used authentic texts written in English by non-native speakers for an international research conference as ELF version and their edited equivalents (EdE). We expected professional translators to generate shorter reading times and greater output during copying and translation compared to translation students and bilinguals without training in translation. We hypothesized that translating would involve a higher cognitive load than transcribing and reading the texts, expressed by stronger theta and weaker alpha oscillations during this task. Furthermore, we assumed that professional translators have a lower cognitive workload during all tasks than translation students and bilinguals. Regarding the differences in the text input, we expected the ELF version to be more difficult to work with than the EdE version. However, this difference might only be evident in professional, experienced translators. To the best of our knowledge, this is the first neurophysiological translation study at the text level that also attempts to assess the cognitive demands of ELF in language professionals and non-professionals.

**Methods**

**Participants**

We collected data from 72 native German (L1) participants (62 females, one nonbinary) with a professional English (L2) background. We recruited three different groups: professional *written* translators (TraPro, N = 19), trainee translators (TraStu, N = 22), and a multilingual control group (Mul, N = 17). Members of the multilingual control group were English language and literature studies students or teachers of English as a foreign language from high schools. All participants were required to use L2 in their daily routine, and the primary direction of translating for professional and student translators was from L2 to L1. Since we recruited participants with varying levels of professional experience, groups could not be matched regarding their age (Table X). However, they did not differ in respect of gender or handedness (Annett, 1970). 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). Thus, we analyzed data from 58 participants. The study was carried out according to the principles in the declaration of Helsinki and approved by the Swiss National Science Foundation ethics committee.



*Table 1: Sample characteristics*

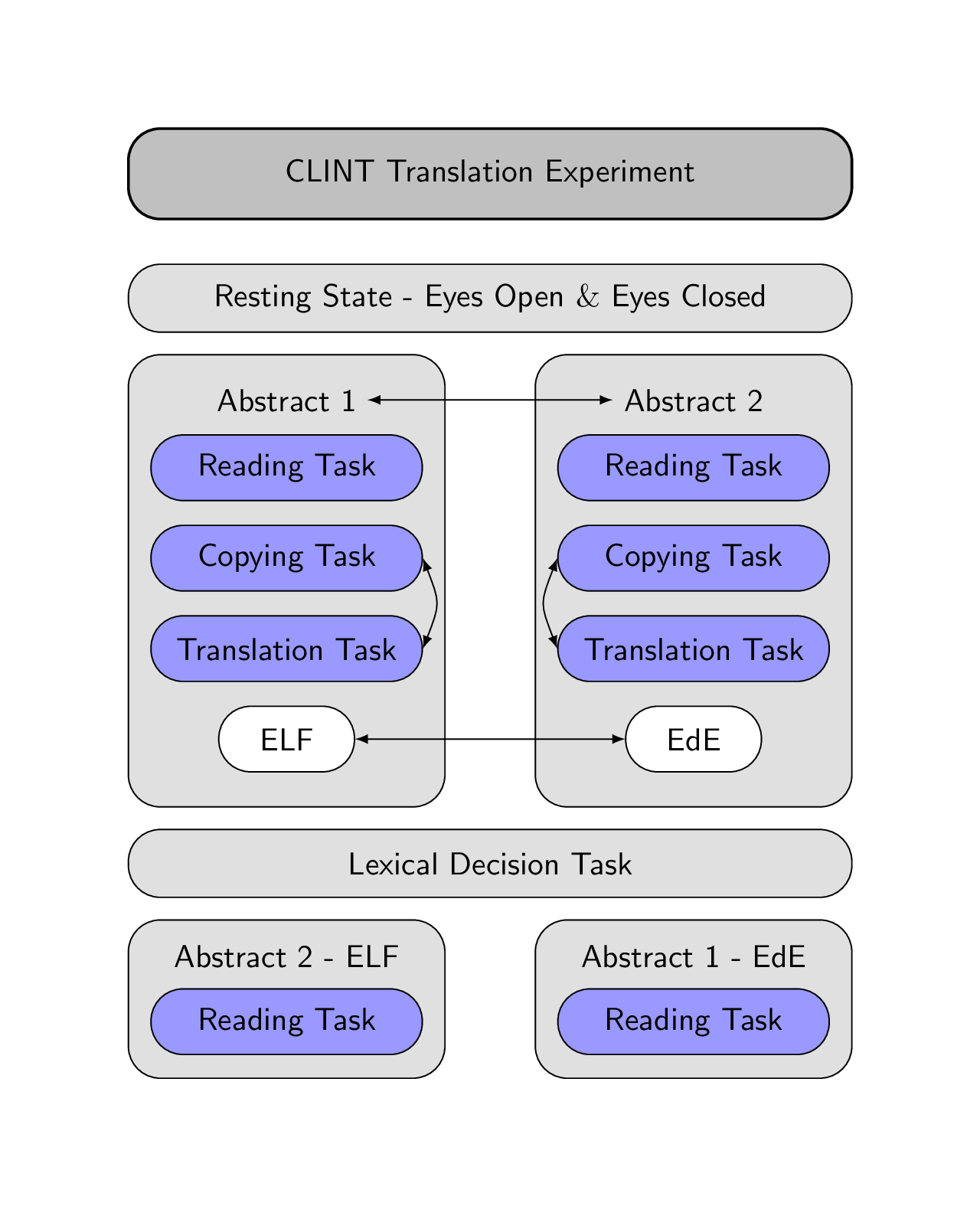
**Psychometrics and questionnaires**

Every participant completed a short English language test ([https://www.sprachtest.de/ einstufungstest-englisch](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 training hours and cumulative training hours per day since the age of 17) in a language background questionnaire. To assess working memory capacity, participants completed both a visual and an auditory 3-back task comprising 60 letter stimuli, of which 20 were target stimuli. The order of the tasks was pseudorandomized across the groups. We analyzed N-back data using d-primes (d’). D-primes were calculated as the difference between the z-transformed hit rate and false alarm rate (Hautus et al., 2021). Additionally, we evaluated the cognitive capabilities 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. Using the standardized T-values, the four subtests 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 the next step, the two abstracts were processed into an edited-to-standard English (EdE) version by professional translators of the Zurich University of Applied Sciences (ZHAW). As few changes as possible were made to keep the text as close 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**

****

*Figure 1: Experimental design. Arrows indicate randomizations in the experiment.*

First, participants completed an EEG resting-state eyes open, and eyes closed condition for three minutes each. Second, participants started with the reading task, followed by the copying and translation task for the first abstract. In the reading task, the text was presented sentence by sentence, and participants could read at a self-paced speed through a button press. Subsequently, we asked the participants how difficult they thought this task was, 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 if participants had read the text. Those questions were multiple-choice, and participants had to choose one of three answer possibilities by pressing a key on the keyboard. Third, in the copying task, participants were asked to copy the presented sentences, and therefore, the generated output was English (L2). After completing a sentence, participants could move on to the following sentence by pressing “Enter”. Fourth, in the translation task, the presented sentences were then translated to German (L1), reflecting a forward translation (FT). 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 thought this task was, while collecting the answers on a 10 cm horizontal bar.

In all tasks, the words of the presented sentence were separated with 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 duration of the reading task differed based on the self-paced reading of participants. However, the copying and translating task was limited to five minutes each. After working on the first abstract, participants continued using the same procedure with the second abstract.

In the experiment, we randomized the order of the abstracts (text 1, text 2), the version (ELF, EdE), and the copying and translation task across participants, indicated by the arrows in Figure X. Therefore. each participant processed an abstract only in one version but not in the other. If the first abstract was in ELF, the second was in EdE and vice versa. Since the copying and translation task duration was limited, participants did not process the whole text but always started from the beginning and worked through the text sentence by sentence. However, it was made sure 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, participants completed all psychometric measurements. Afterward, they were prepared for 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. The experiment was programmed in MATLAB 2016b using the Psychophysics Toolbox Version 3 extension (Kleiner et al., 2007) for behavioral data acquisition. 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 good contact on the scalp, and impedances were kept below 40 kOhm. 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.

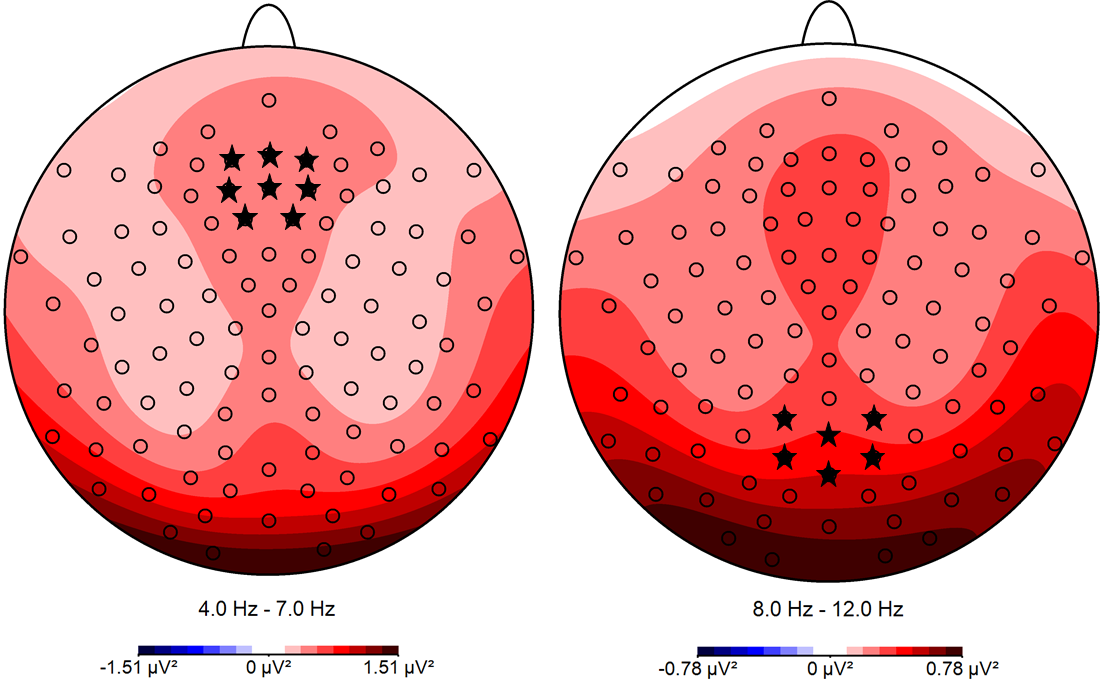
**Behavioral data processing**

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 of the mouse click from 0 (easy) for both the reading and translation tasks. Therefore, higher values indicate a more pronounced perceived difficulty of the task. Regarding the copying and translation tasks, we evaluated the total amount of chars typed during the 5 minutes, as well as the number of chars typed if deletions were subtracted (chars end version). Those variables can be regarded as a measure of the efficiency in the copying and translation task (Ehrensberger-Dow & Massey, 2014). Furthermore, we retrieved the percentage of deletions, which refers to 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 the rating of 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 provided by three German native translators with a master’s degree in translating and English as 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 a ICC(C,3) = 0.575 (95%-confidence interval = 0.421 – 0.694) for the fluency rating and a ICC(C,3) = 0.909 (95%-confidence interval = 0.875 – 0.934) for the accuracy rating. The ICC for the fluency rating likely reflects moderate reliability, whereas the ICC for the accuracy rating reflects excellent reliability (Koo & Li, 2016). Finally, we averaged the three raters to generate a mean rating score for fluency and accuracy further used in the statistical analyses.

**EEG data processing**

The data was 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. 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 frontal electrodes for the electrooculogram (EOG) regression. Fifth, power line noise 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 during pauses.

Further preprocessing of the EEG data was executed in the Brain Vision Analyser. First, we re-referenced the data to an average reference montage, and segmented the EEG into the different task segments. 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 were skipped. Forth, a fast Fourier transform (FFT) with a Hanning window (Length = 10%) was applied to all remaining segments. The resulting 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) band. Based on the voltage distribution of the grand average across the reading, copying, and translation tasks, we analyzed theta power at a frontal (E4, E5, E10, E11, E12, E16, E18, and E19) and alpha power at a parietal (E61, E62, E67, E72, E77, and E78) electrode pool (see Figure X for electrode positions). Finally, we averaged the power per pool and frequency band for statistical analysis.

**

*Figure X: Topographical voltage distribution maps for theta and alpha band across reading, copying and translation tasks, and all participants. The channels selected for analyses are marked with \*.*

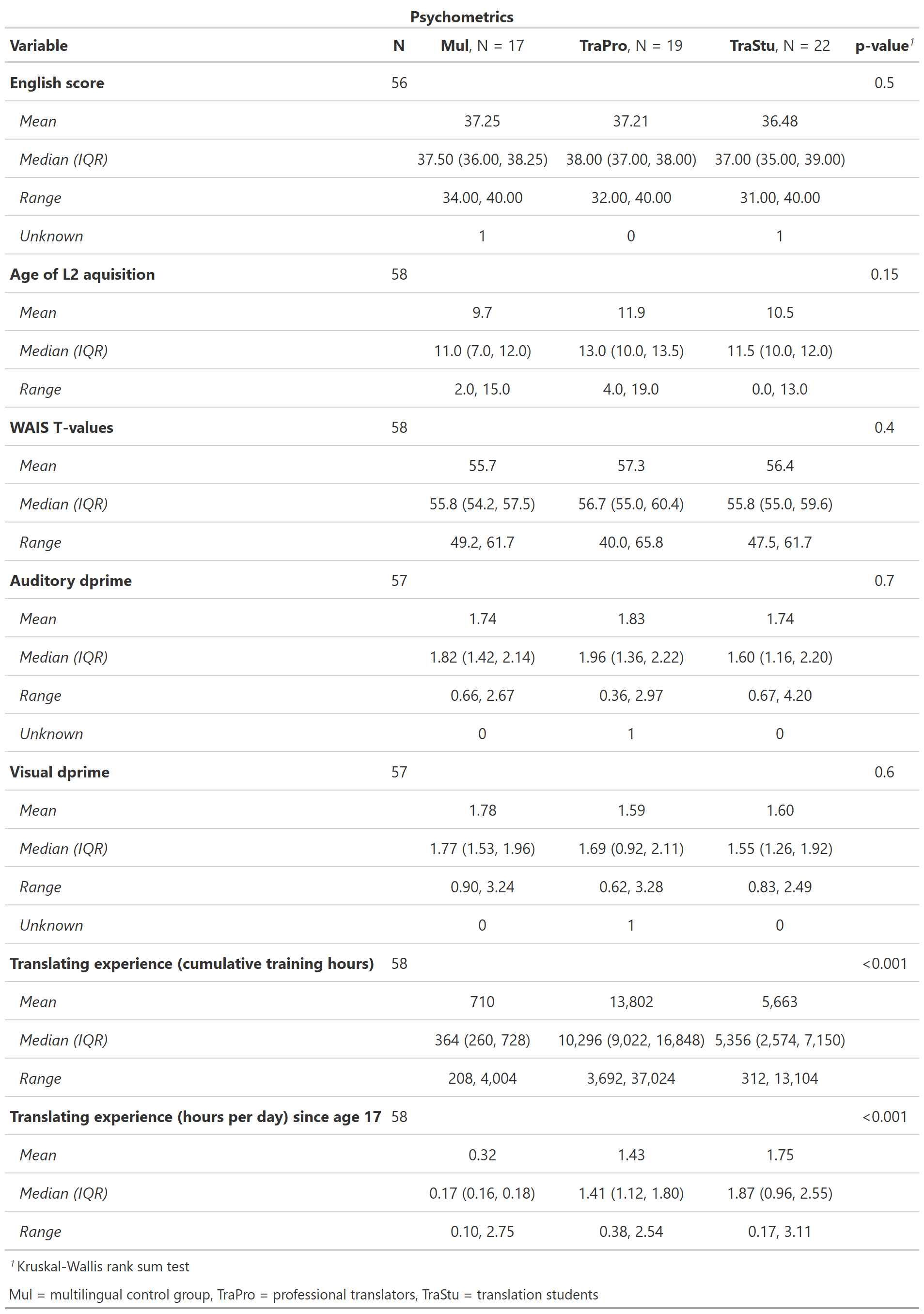
**Statistical analyses**

All statistical analyses were performed using Linear Mixed Models (LMM) implemented in the lme4 package (Version 1.1-23, <https://cran.r-project.org/web/packages/lme4/>) in R. For model-fitting, we used a bottom-up strategy starting with the null model and added fixed effects for our target variables. 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**

**Psychometrics and questionnaires.**

Our groups did not differ regarding English score, age of L2 acquisition, WAIS T-values, auditory d’, and visual d’ (Table X). However, as expected by our recruitment, our groups differed regarding cumulative training hours in translating and interpreting (TraPro: M = 13’802, TraStu: M = 5’663, Mul: M = 710, F(2,55) = 30.895, *p* < 0.001, η2G = 0.529), as well as in the cumulative training hours per day since the age of 17 (TraPro: M = 1.43, TraStu: M = 1.75, Mul: M = 0.32, F(2,55) = 18.858, *p* < 0.001, η2G = 0.407).

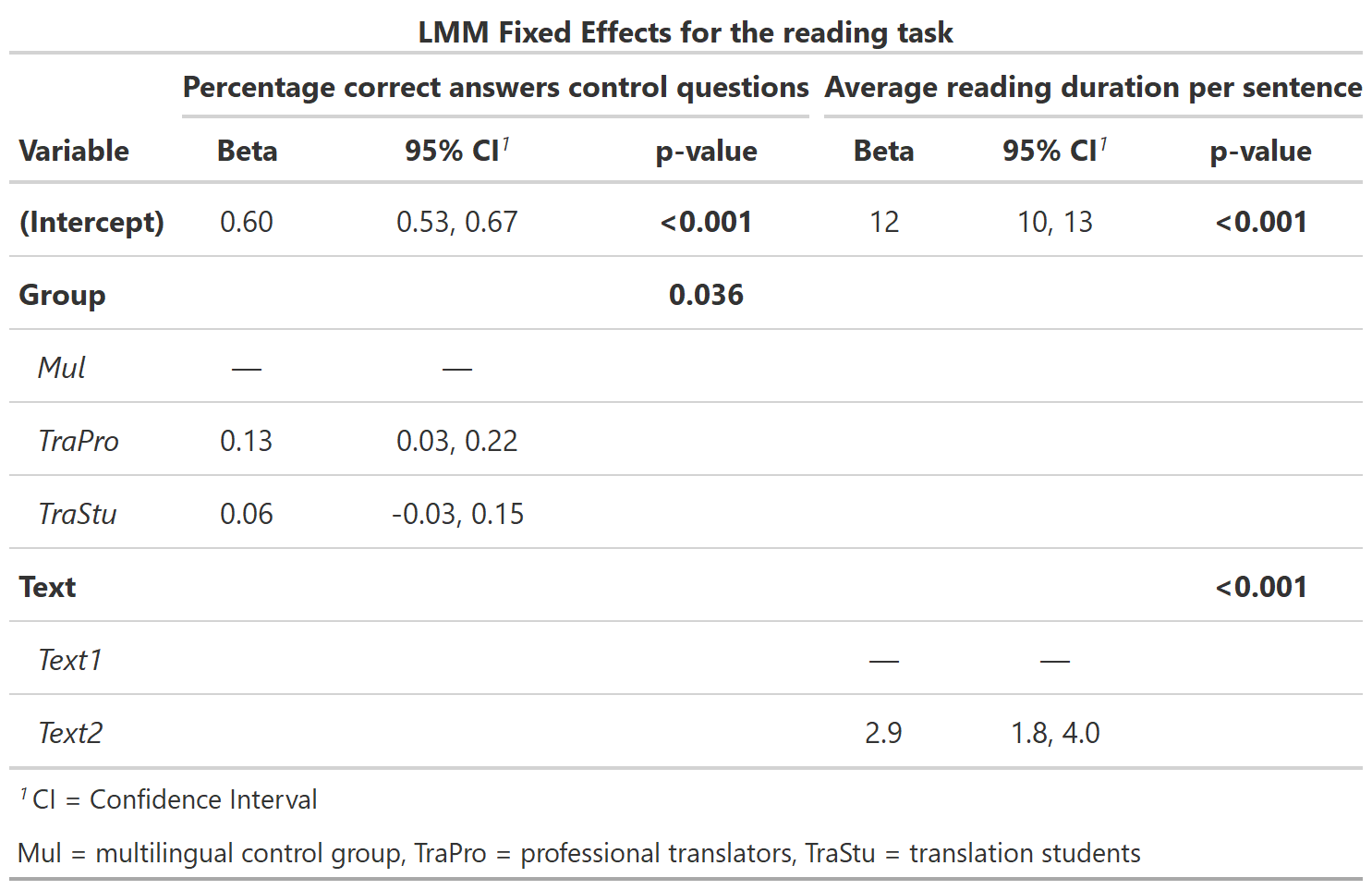
****

*Table X: Results of the psychometrics and questionnaires.*

**Behavioral results**

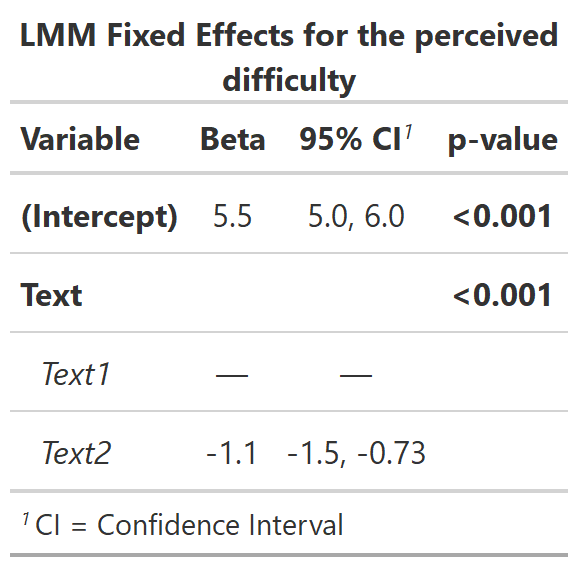
**Reading task**

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 text. Introducing a fixed effect for condition (∆Χ2 = 0.072, *p =* 0.789) as well as 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 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 group TraPro (*β* = 0.126) and TraStu (*β* = 0.059) compared to the multilingual control group. Thus, the percentage of correct answers to the control questions was best predicted by group. All LMM fixed effects of behavioral measurements in the reading task are summarized in Table X.

****

**Perceived difficulty of the reading and translation task**

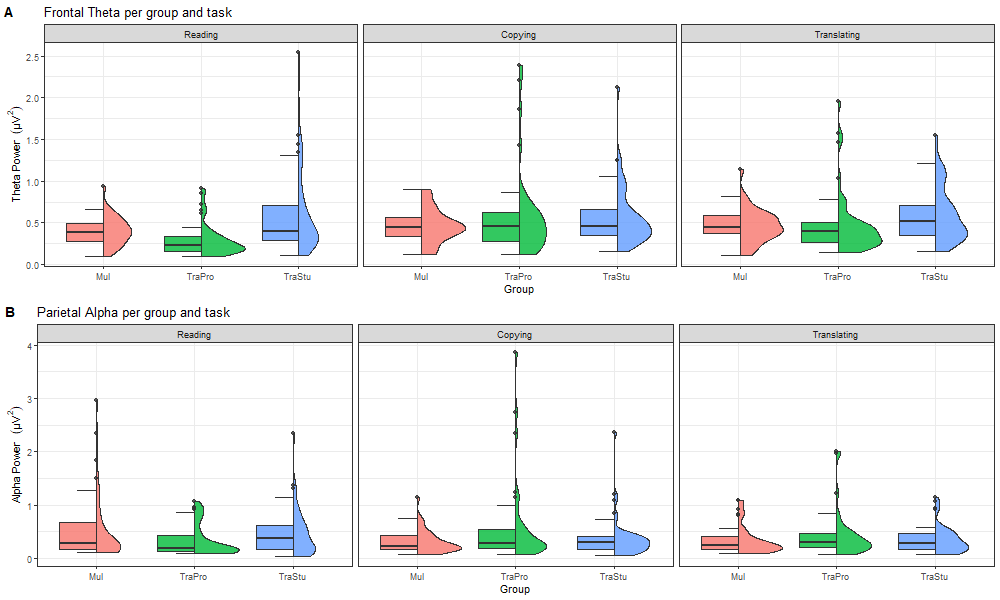
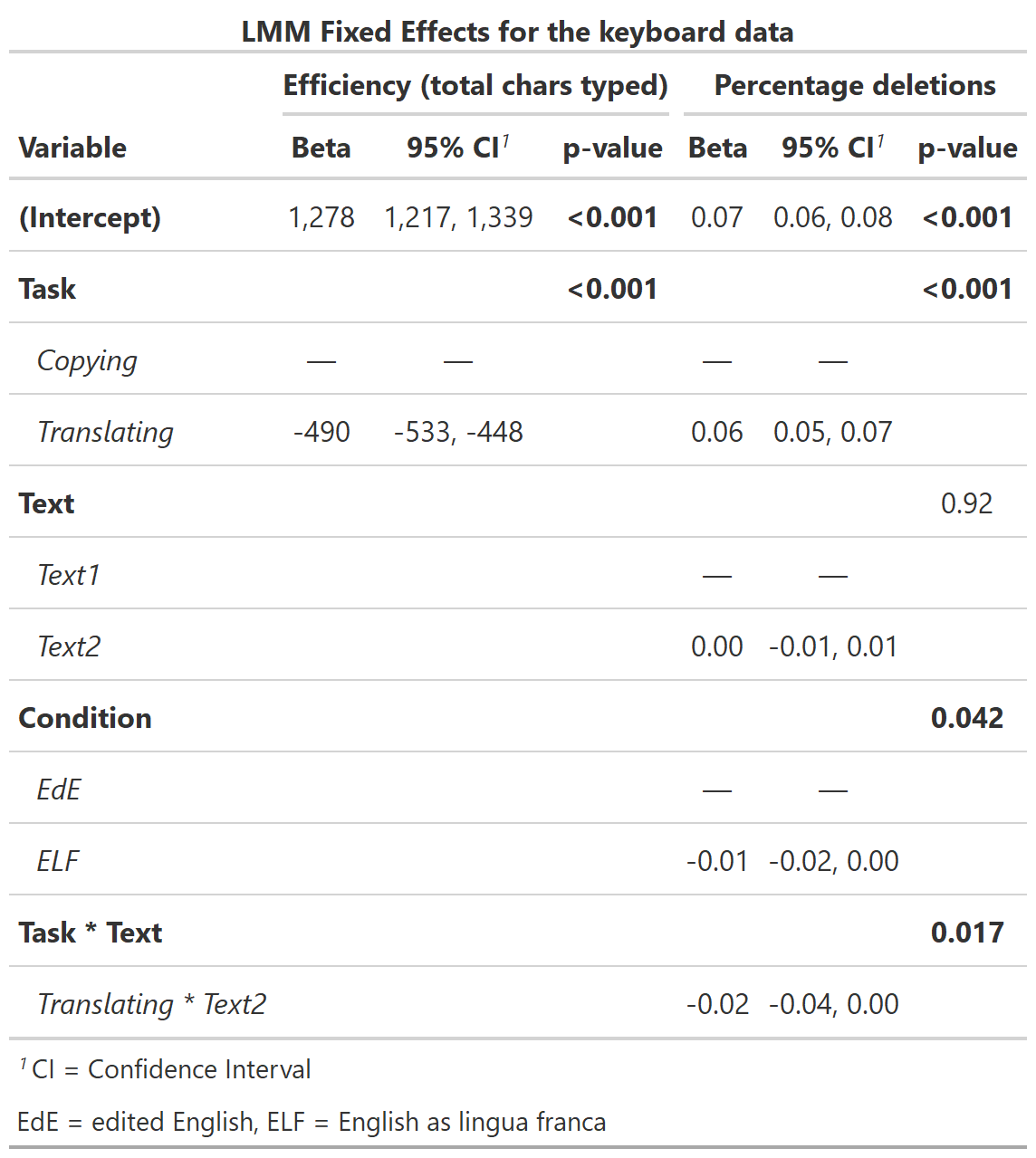
For the analysis of the perceived difficulty of the reading and translation task, including a fixed effect for task (∆Χ2 = 0.006, *p =* 0.941) did not significantly improve model fit. Introducing a fixed effect for text significantly improved model fit (∆Χ2 = 28.166, *p* < 0.001), indicating that the second abstract was perceived as less difficult (*β* = -1.128) compared to the first abstract. Adding a fixed effect for condition (∆Χ2 = 0.001, *p =* 0.993) as well as 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 X.

****

**Keyboard data**

For the analysis of the total number of chars typed in the copying and translation task, 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 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.

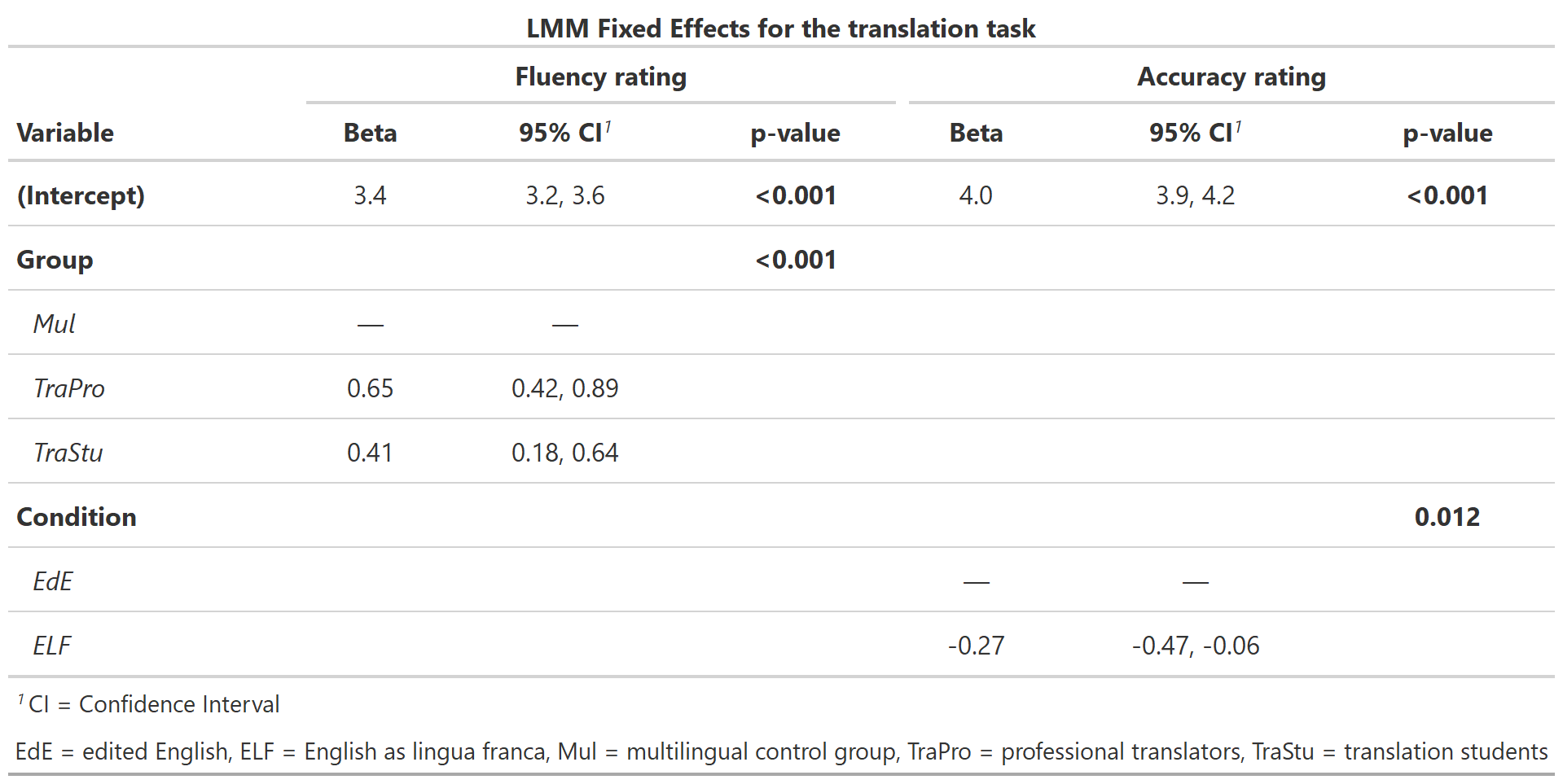
For the analysis of the percentage of deletions in the copying and translation task, including a fixed effect for task significantly improved model fit (∆Χ2 = 93.565, *p <* 0.001), indicating that the percentage of deletions was higher in the translating task (*β* = 0.052) compared to the copying task. 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 reflects a lower difference in the percentage of deletions between the copying and translation task 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 keyboard data are summarized in Table X.

****

**Translation task**

For the analysis of the fluency rating, including a fixed effect for text (∆Χ2 = 0.018, *p =* 0.893) as well as 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.

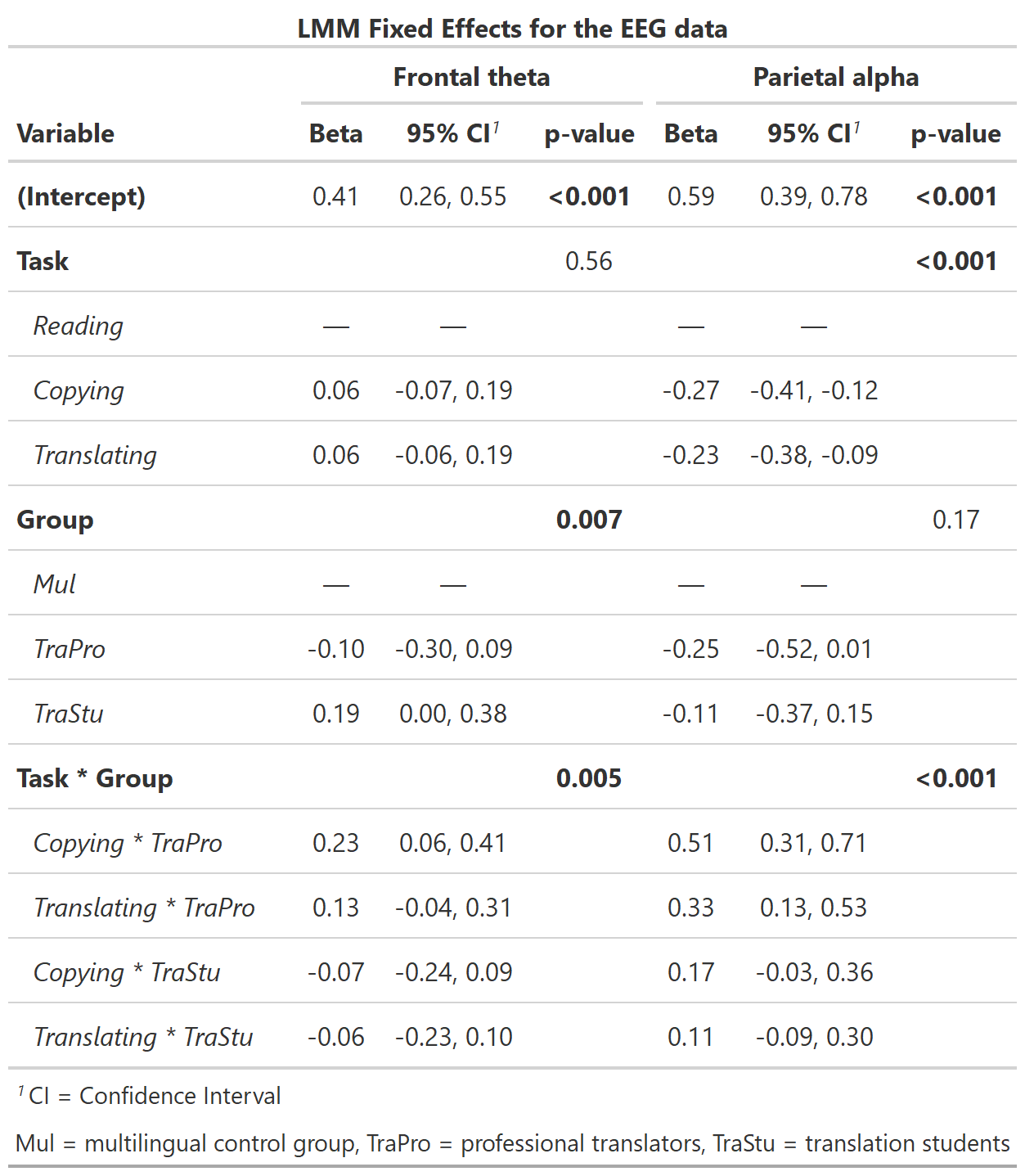
For the analysis of the accuracy rating, including a fixed effect for text (∆Χ2 = 0.171, *p =* 0.679) did not significantly improve model fit. Introducing a fixed effect for condition significantly improved model fit (∆Χ2 = 6.314, *p =* 0.012), indicating that the translations of the ELF version (*β* = -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 X.

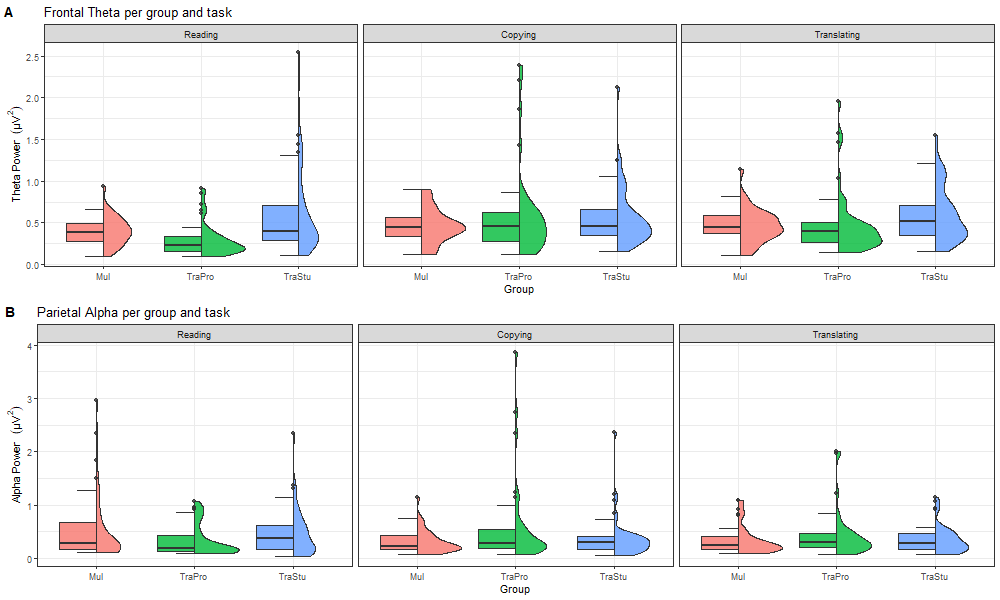
****

**EEG results**

For the analysis of the frontal theta, including a fixed effect for task significantly improved model fit (∆Χ2 = 9.646, *p =* 0.008), indicating higher theta activity in the copying task (*β* = 0.106) and translating task (*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 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 group of TraPro revealed lower theta band activity (*β* = -0.105) compared to the Mul group, whereas the group of TraStu revealed higher theta band activity (*β* = 0.186). In the copying task, the group of TraPro (*β* = 0.126) and TraStu (*β* = 0.111) were reflected by higher theta band activity compared to the Mul group. In the translating task, the group of TraPro (*β* = 0.030) and TraStu (0.123) showed higher theta band activity compared to the Mul group. Thus, the frontal theta activity was best predicted by task, group, and the interaction between group and task.

For the analysis of the parietal alpha, 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. However, adding an interaction between task and group significantly improved model fit (∆Χ2 = 26.611, *p <* 0.001) compared to a model with only main effects for task and group. In the reading task, the group of TraPro (*β* = -0.252) and TraStu (*β* = -0.107) revealed lower alpha band activity compared to the Mul group. In the copying task, the group of TraPro (*β* = 0.260) and TraStu (*β* = 0.058) were reflected by higher alpha band activity compared to the Mul group. In the translating task, the group of TraPro (*β* = 0.080) showed higher alpha band activity compared to the Mul group, whereas the group of TraStu revealed lower alpha band activity (*β* = -0.001). Thus, the parietal alpha activity was best predicted by task, group, and the interaction between group and task.



****

**Discussion**

General discussion:

* Workload in three different tasks related to translation, namely reading, copying and backward translation
* First neurophysiological assessment of translating at text level that also takes cognitive demands of non-native source text into account
* Behavioral metrics and EEG theta and alpha power
* 3 different groups, including professional translators, translator students and a multilingual control group
  + Differed in age, but not in age of L2 acquisition, L2 proficiency, general intelligence as indicated by the WAIS t-Value, d-primes in the visual and auditory 3-back task, gender, and handedness
  + Differed in cumulative training hours in translating and the translating training in hours per day since the age of 17.

Behavioral metrics:

* Reading task
  + For the average reading duration per sentence, we could only detect an effect of text, indicating that the second text was more complex to read than the first text.
  + Interestingly, we did not observe a difference in the reading speed between our groups as we would have expected professional translator to use less time for the same amount of text compared to less experienced student translators as well as the multilingual control group.
  + A possible explanation for the absence of group differences regarding the reading speed could be the independently selected speed during the task. Although instructions for the task were the same for all participants, they could have used different strategies for reading. Since the did an introduction for all tasks at the beginning of the measurement, they already knew they will have to translate at some point. Therefore, professional translators might have already used a strategy to read for translation, that is known to reduce reading speed (QUELLE?).
  + However, the examination of the control questions after the reading task revealed a better performance of the professional translator group compared to the student translators and the control group. Taken together, those results might indicate that professional translators were able to gather more information during the same time leading to a higher percentage of correct answers after reading the text.
* Copying / Translating task
  + Regarding the behavioral metrics in the copying and translating task we assessed the overall efficiency through the total amount of chars typed, the percentage of deletions, and the ratings of the fluency and accuracy of the text output our participants generated in the translating task.
  + As expected, the overall output was much less in the translating task compared to the copying task indicated by the reduced efficiency, as simply transcribing text from English to English is much less an effort than translating from English to German.
  + Interestingly, we did not observe an effect of group as we would have expected professional translators to be more efficient especially in the translating task compared to translation students and our control group. Since it is their daily work business, professional translators have been shown to outperform untrained bilinguals regarding the number of keystrokes in workplace studies (Ehrensberger, 2020). In contrast, the eye-tracking study of Carl and Klay (2012) had the surprising result, that professional translators did not differ from student translators regarding the number of characters per time, but they were able to focus their gaze on the source text while students shifted their attention more often from the source to the target text. However, we speculate that the absence of an expertise effect in the efficiency in our data might be due to the special setting of an EEG study. Even though we emphasized of applying a realistic, ecological valid, and for professional translators familiar setting in letting our participants translate sentence by sentence, the benefits of professional translators could be limited to the familiar environment of their own keyboard. In contrast, in our lab setting, they had to work with an unfamiliar keyboard, with an EEG cap on their head and producing text within a MATLAB script that does not exactly behave as other typewriting programs.
  + The percentage of deletions, which is the number of backslashes relative to the number of keystrokes can be used as an index of the complexity of a task. The more complex a task is, the more often a person will use the backslash. Therefore, it is no surprise, that the percentage of deletions was higher in the translating task compared to the copying task. Again, we did not observe an effect of group which speaks for a similarity in the performance of our groups. Interestingly, the type of the text input had an impact on the percentage of deletions as the ELF condition had less deletions compared to the EdE condition. This is somewhat surprising, as we would have expected non-native text input to be more challenging to transcribe or translate and therefore produce more backslashes by our participants. Possibly, our participants were more challenged by the ELF condition which might have led to the fact that the typos were noticed less and consequentially, the backslash was used less. Unfortunately, we were not able to check our data if there were more remaining typos in the ELF compared to the EdE version to support our claim, as this would be a very time-consuming effort.
  + Regarding the rating of the fluency of the text output, we could find an effect of group indicating a higher fluency of translated sentences in the professional translator group compared to the translator students as well as the multilingual control group. Even though we only achieved a moderate inter-rater reliability in our fluency rating, which is likely due to the challenge of rating the fluency on a sentence level, our results reflect the enhanced level of expertise in professional translators and student translator accurately.
  + For the rating of the accuracy, we did not observe an effect of group as we would have expected it, which could be explained by the overall higher ratings in the accuracy values and therefore, a better performance of the multilingual control group that was at the level of professional translators. Interestingly, our results indicate an overall lower accuracy when translating ELF input. Therefore, we provide some first evidence that translating non-native language input might be indeed more challenging and lead to a lower accuracy in the target text. This finding was true for all groups that were affected similarly by the ELF condition.
  + For the results of the perceived difficulty, for which participants had to indicate how challenging the reading and the translating task was on a 10 cm bar ranging from low to high, we could only find an effect of the text. Therefore, our participants rated the second abstract as being more challenging, regardless of if the ELF or EdE version was shown or if they had to read or translate the text. The assessment was designed to potentially measure the perceived workload as a self-assessment. Neither did our groups differ in the perceived workload of the tasks nor did the different versions of the text input. Our data show a extensive variability in this measure, as the 95% confidence interval ranged from 0.5 to

**Limitations:**

Research on bilingualism faces some major challenges and from it, the heterogeneity within the samples is the biggest (Calvo et al., 2016). For example, when comparing bilinguals, influencing variables such as socioeconomic status, migration experience, language switching habits, or frequency of language use are rarely considered or challenging to be consistently measured (Calvo et al., 2016; Ferreira et al., 2020). Moreover, major variables examined in experiments such as language proficiency, age of acquisition (AoA), and age of second language onset are rather determined by the subjective statements of the participants than objectively collected (Calvo et al., 2016; de Bruin, 2019). Another factor rarely considered is the context in which the second language was learned (de Bruin, 2019). Therefore, researchers are challenged to find objective measurements such as, for example, detailed assessments to assess language proficiency levels.

**Conclusions**

**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

Anderson, E. W., Potter, K. C., Matzen, L. E., Shepherd, J. F., Preston, G. A., & Silva, C. T. (2011). A user study of visualization effectiveness using EEG and cognitive load. *Computer Graphics Forum*, *30*(3), 791–800. https://doi.org/10.1111/j.1467-8659.2011.01928.x

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. *Psychology Review*, *22*(4), 425–438. https://doi.org/10.1007/s

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

Bastiaansen, M., & Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. *Progress in Brain Research*, *159*, 179–196. https://doi.org/10.1016/S0079-6123(06)59012-0

Bastiaansen, M., Magyari, L., & Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience*, *22*(7), 1333–1347. https://doi.org/10.1162/jocn.2009.21283

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., 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), 231–244.

Birba, A., Beltrán, D., Martorell Caro, M., Trevisan, P., Kogan, B., Sedeño, L., Ibáñez, A., & García, A. M. (2020). Motor-system dynamics during naturalistic reading of action narratives in first and second language. *NeuroImage*, *216*, 116820. https://doi.org/10.1016/J.NEUROIMAGE.2020.116820

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

Byers-Heinlein, K., Esposito, A. G., Winsler, A., Marian, V., Castro, D. C., & Luk, G. (2019). The case for measuring and reporting bilingualism in developmental research. *Collabra: Psychology*, *5*(1), 1–11. https://doi.org/10.1525/collabra.233

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*. 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

Chuang, L. Y., Huang, C. J., & Hung, T. M. (2013). The differences in frontal midline theta power between successful and unsuccessful basketball free throws of elite basketball players. *International Journal of Psychophysiology*, *90*(3), 321–328. https://doi.org/10.1016/j.ijpsycho.2013.10.002

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*, *25*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences*, *9*(3). https://doi.org/10.3390/bs9030033

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. (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. https://digitalcollection.zhaw.ch/handle/11475/2691

Ferreira, A., Schwieter, J. W., & Festman, J. (2020). Cognitive and neurocognitive effects from the unique bilingual experiences of interpreters. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.548755

Friese, U., Köster, M., Hassler, U., Martens, U., Trujillo-Barreto, N., & Gruber, T. (2013). Successful memory encoding is associated with increased cross-frequency coupling between frontal theta and posterior gamma oscillations in human scalp-recorded EEG. *NeuroImage*, *66*, 642–647. https://doi.org/10.1016/j.neuroimage.2012.11.002

García, A. M., & Ibáñez, A. (2016). Processes and verbs of doing, in the brain: Theoretical implications for systemic functional linguistics. *Functions of Language*, *23*(3), 305–335.

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*(NOV). https://doi.org/10.3389/fpsyg.2014.01302

García, A. M., Mikula, E., & Ibáñez, A. (2016). A neuroscientific toolkit for translation studies. In R. Muñoz Martín (Ed.), *Reembedding translation process research* (pp. 21–46). John Benjamins.

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., Leong, H., McEvoy, L., Whitfield, S., Du, R., & Rush, G. (1998). Monitoring Working Memory Load during Computer-Based Tasks with EEG Pattern Recognition Methods. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *40*(1), 79–91. https://doi.org/10.1518/001872098779480578

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. In *Aviation, Space, and Environmental Medicine* (Vol. 56, Issue 11, pp. 1110–1114). Aerospace Medical Assn.

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*(OCT), 1–14. 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

Hunter, C. R. (2020). Tracking Cognitive Spare Capacity during Speech Perception with EEG/ERP: Effects of Cognitive Load and Sentence Predictability. *Ear and Hearing*, 1144–1157. https://doi.org/10.1097/AUD.0000000000000856

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

Ishii, R., Shinosaki, K., Ukai, S., Inouye, T., Ishihara, T., Yoshimine, T., Hirabuki, N., Asada, H., Kihara, T., Robinson, S. E., & Takeda, M. (1999). Medial prefrontal cortex generates frontal midline theta rhythm. *NeuroReport*, *10*(4), 675–679. https://doi.org/10.1097/00001756-199903170-00003

Jäncke, L. (2017). *Lehrbuch kognitive Neurowissenschaften* (2nd ed.). Hogrefe.

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *Eur J Neurosci*, *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*, *92*(7), 2899–2903.

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. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, *29*(2–3), 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3

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

Klimesch, W., Doppelmayr, M., Pachinger, T., & Russegger, H. (1997). Event-related desynchronization in the alpha band and the processing of semantic information. *Cognitive Brain Research*, *6*(2), 83–94. https://doi.org/10.1016/S0926-6410(97)00018-9

Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., & Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neuroscience Letters*, *244*(2), 73–76. https://doi.org/10.1016/S0304-3940(98)00122-0

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

Luk, G. (2017). Bilingualism. In B. Hopkins, E. Geangu, & S. Lingenauger (Eds.), *The Cambridge encyclopedia of child development* (2nd ed., pp. 385–391). Cambridge University Press. https://doi.org/10.1017/9781316216491.062

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. *ACM International Conference Proceeding Series*, 80–89. https://doi.org/10.1145/3027385.3027431

Muñoz, E., Calvo, N., & García, A. M. (2019). Grounding translation and interpreting in the brain: What has been, can be, and must be done. *Perspectives: Studies in Translation Theory and Practice*, *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

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

Swerdloff, M. M., & Hargrove, L. J. (2020). Quantifying Cognitive Load using EEG during Ambulation and Postural Tasks. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2020*-*July*, 2849–2852. 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*(September 2017), 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

Vilas, M. G., Santilli, M., Mikulan, E., Adolfi, F., Martorell Caro, M., Manes, F., Herrera, E., Sedeño, L., Ibáñez, A., & García, A. M. (2019). Reading Shakespearean tropes in a foreign tongue: Age of L2 acquisition modulates neural responses to functional shifts. *Neuropsychologia*, *124*, 79–86. https://doi.org/10.1016/j.neuropsychologia.2019.01.007

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

Weiss, S., Mueller, H. M., Schack, B., King, J. W., Kutas, M., & Rappelsberger, P. (2005). Increased neuronal communication accompanying sentence comprehension. *International Journal of Psychophysiology*, *57*(2), 129–141. https://doi.org/10.1016/J.IJPSYCHO.2005.03.013