Keylogging

Published Version:
Munoz Martin, R., Sun, S., Du, Z., Puerini, S. (2025). Keylogging. Amsterdam : John Benjamins [10.1075/rmal.10.07mar].

Availability:
This version is available at: https://hdl.handle.net/11585/1014085 since: 2025-04-08

Published:
DOI: http://doi.org/10.1075/rmal.10.07mar

(Article begins on next page)

Chapter 7

Keylogging

Ricardo Muñoz Martín,[1] Sanjun Sun,[2]
Zhiqiang Du[1] & Sara Puerini[1]

[1] University of Bologna | [2] Beijing Foreign Studies University

Abstract: Keystroke logging, or *keylogging*, is a powerful research method for collecting unobtrusive, fine-grained data on text production processes. In cognitive translation and interpreting studies (CTIS), keyloggers like Translog-II and Inputlog capture the dynamics of translation behavior, including pauses, revisions, information searching, and typing patterns, which enrich language data and timestamp it, down to the millisecond. Keylogging data enables researchers to infer underlying cognitive processes, compare translator expertise levels, and assess task difficulty. This chapter provides an overview of keylogging research in CTIS, covering its conceptual basis, key variables, ethical considerations, analytical methods, and limitations. It emphasizes the need for ecological validity, standardized metrics, and thorough reporting in keylogging studies. Future applications may expand to

multimodal translation, collaborative workflows, and integration with sensor technologies.

Keystroke logging, or *keylogging,* is a data-collection method that systematically monitors and registers single and combined keyboard presses, whether sequenced or chorded, and often mouse actions (see Figure 1). Such minimal, self-contained events are the basic behavioral units in keylogging. Consider that there are more than 25,000 meaningful combinations of just three consecutive keys, alphanumeric or not. Keyloggers also timestamp events and the latency between them, or *interkeystroke intervals* (ikis), with millisecond precision. Features such as the number of words typed and edits vary across tasks (Conijn et al., 2019) and can be hypothesized to hint at cognitive processes such as problem solving and performance self-monitoring and assessment (Baaijen et al., 2012; Galbraith & Baaijen, 2019). Qualitative analyses provide valuable insights (e.g., Sala-Bubaré et al., 2021) but most keylogging research is quantitative, with time typically playing a central role.

Keyloggers were initially developed in the 1970s and 1980s primarily for system maintenance and debugging purposes. Yet many of them have been and are used as malware to illegally and surreptitiously register and transmit information to malicious actors, such as hackers. Notorious keyloggers include *Magic Lantern* (later, CIPAV), developed in the early 2000s by the FBI for criminal investigations; *Zeus*, a banking trojan with keylogging capabilities, *DarkTequila* and *Pegasus*. This malevolent use has tarnished the reputation of keylogging. Many antivirus programs detect research keyloggers as spyware, and often refuse to install them.

With the commercialization of keyloggers in the 1990s, their applications broadened to include gathering user feedback for digital products, diagnosing IT user problems, lifelogging, and aiding law enforcement investigations. They are widely used in the field of *behavioral dynamics*—mainly in cybersecurity and user behavior analysis—where they provide insights into typing patterns, human-computer interaction, and security vulnerabilities. Often, *commercial* keyloggers, such as SpyAgent and Spyrix, do not only register keyboard and mouse events.[1] They also record all kinds of activities such as system and microphone audio, webcam footage, and screen recordings. The granularity and precision of their logging, however, may not be extremely high for research purposes. Their deployment for monitoring internet usage, employee productivity, and even personal surveillance is subject of ethical debates.

Broadly, keyloggers are classified as either local or remote. *Remote* keyloggers work within web browsers to track users' navigation and click patterns or are integrated into particular websites, primarily for profiling users, rather than to collecting what they type. *Local* keyloggers can be hardware or software. *Hardware* keyloggers are physical devices (such as dongles, flash drives and plug-ins) inserted between the keyboard connector and its port on the computer. *Software* keyloggers, on the other hand, employ various techniques for data collection, typically tapping into the data sent from the keyboard to the CPU. Other types of local software keyloggers include wireless keyboard sniffers—which are capable of capturing the radio frequency signals such as those from Wi-Fi or Bluetooth, transmitted between a wireless keyboard and its corresponding receiver—and acoustic keyloggers, which analyze keyboard sounds to discern pressed keys.

Research keyloggers are typically local software applications. They generate logfiles that contain complete records of a session and provide a detailed chronological

---

[1] See https://www.spytech-web.com/ and https://www.spyrix.com/

framework for reconstructing the sequences of events in text production and related activities. The analysis of alternating keystrokes and ikis, their durations, distributions, rhythms, and interactions with each other and with linguistic and environmental aspects, along with sources of information and applications, yields a rich dataset. This data is crucial for formulating and testing hypotheses about users' behaviors and their mental processes. Researchers often interpret and recursively redescribe sequences of events and ikis as higher-order actions—for instance, a series of deletions followed by typing new letters can be redescribed as "replacing a word")—which are associated with specific text production subtasks, such as revision and online information search.

In cognitive translation and interpreting studies (CTIS), keylogging has emerged as a powerful tool for investigating the cognitive processes underlying translation and interpreting tasks. The following sections will address its application in CTIS research, and will discuss key concepts, variables, and other methodological considerations. The chapter will first provide an overview of the state of the art in keylogging research within CTIS, highlighting the insights gained from studying the differences between informant groups, the cognitive demands of various tasks, and the impact of emerging technologies. It will then explore the conceptual aspects of keylogging, including the selection and operationalization of variables, measurement techniques, and the interpretation of pause and revision data. Next, aspects of implementing a CTIS research project using keylogging will cover research design choices, data analysis methods, and the features of popular keylogging tools like Translog-II and Inputlog. Finally, the chapter will discuss advantages, limitations, and challenges of keylogging in CTIS, as well as ethical considerations and the need for ecological validity in research. By providing a comprehensive overview of keylogging in CTIS, this chapter aims to guide researchers in leveraging this powerful method to advance our understanding of the complex cognitive processes underlying multilectal mediated communication tasks.

1. The method, and key concepts

Keylogging is widely used in writing process studies (e.g., Allen et al., 2016; Chukharev-Hudilainen, 2014; Leijten & Van Waes, 2013; Manchón & Roca de Larios, 2023). It was the next logical step in a trend that departed from the study of the cognitive aspects of writing and moved to compare manual writing with computer typing (e.g., Van Waes, 1992). Duin and Bridwell (1985) and Bridwell, Sirc and Brooke (1985) pioneered keylogging in research when they used the program *Recording Wordstar* to study revision processes of novice writers. The advocated need of dedicated applications soon led to creating TraceIt (Severinson Eklundh & Kollberg, 1992), this first research keylogger, which run on Macintosh computers. Then there came Scriptlog (Ahlsén & Strömqvist, 1999; Strömqvist & Karlsson, 2002) and Inputlog (Leijten & Van Waes, 2006), to be followed by GGXLog (Usoof et al., 2020). Scriptlog and GGXLog incorporate a basic text editor and only register what happens within it, whereas Inputlog captures all actions performed on a Windows computer, with informants using MS Word for text processing.

Research typically focuses on the temporal aspects of text production, such as pause patterns and typing speed, to infer cognitive processes such as problem-solving, decision-making, and the mental effort when dealing with particular language structures, such as metaphors, or that involved in specific writing tasks. Researchers also study the role of language skills and users' expertise in different settings. For instance, Leijten et al. (2015) compared text production dynamics (pausing patterns) in L1 Dutch and L2 English, and found L2 writing to be less *fluent* (Chenoweth & Hayes, 2001) in terms of characters and words per minute, uninterrupted writing stretches, and final text length.

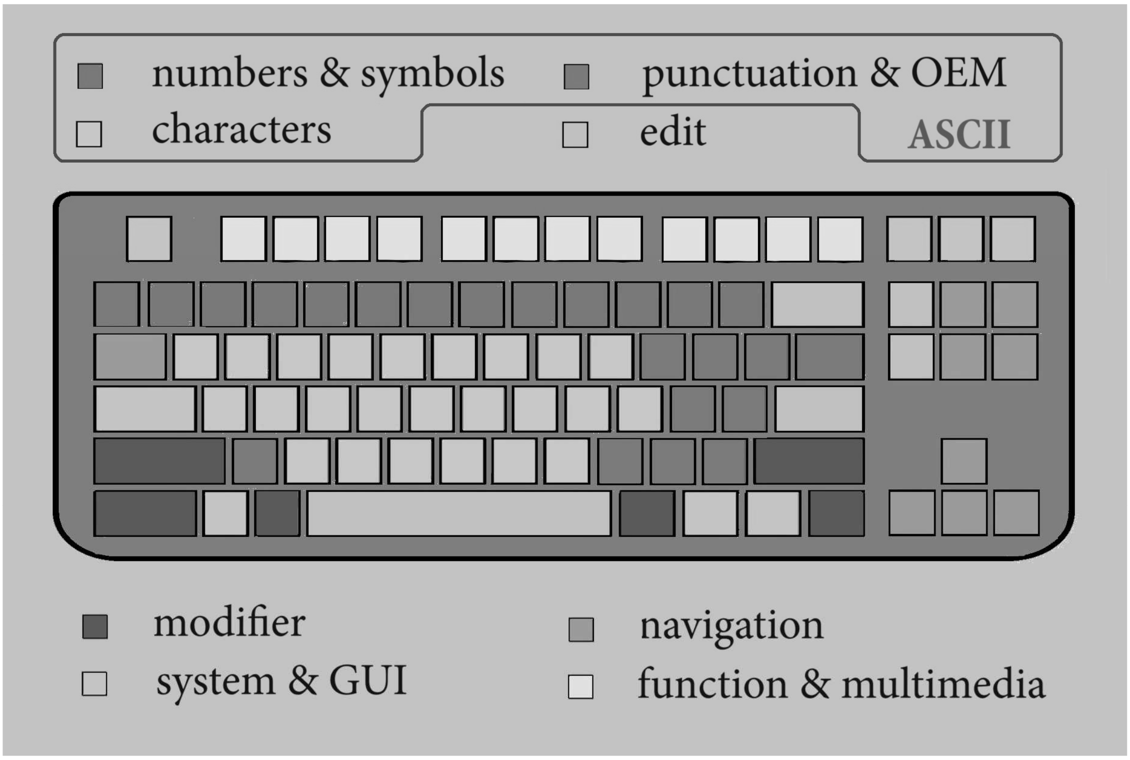

Figure 1. Types of keys in standard computer 104-105-key keyboard. They typically follow the ANSI or the ISO standards, with their keys mostly arranged in langue or country variations of the QWERTY layout. Keyboards often include keys to access functions of the operating system through a graphic user interface (GUI). The output of a few keys varies between keyboards and layouts because they are *originally encoded by the manufacturer* (OEM).

Typing speed, keystroke patterns, and ikis are also under the spotlight in *keystroke* or *keyboard dynamics*, a research domain within cybersecurity. In their search for robust methods for user authentication that will be less invasive than voice, iris, and fingerprint recognition, typing peculiarities are approached as one of the biometric parameters against cyberthreats. In keystroke dynamics, researchers study how people type short text stretches, like their names or their passwords. Working on such small samples, they apply fine-grained statistical analyses, advanced algorithms and deep learning techniques to assess additional indicators such as rhythm, keystroke duration and physical pressure on the keys. These studies aim to develop models that pinpoint unique characteristics of individuals and their emotional states, and can also infer *soft biometrics*, such as age and gender (Katerina & Nicolaos, 2018; Shadman et al., 2023; Shi et al., 2023; Thomas & Preetha Mathew, 2023; Tsimperidis & Arampatzis, 2020). This research domain is quite disconnected from writing process research and CTIS, but it is quite developed and sophisticated and the high granularity they adopt should be of great help in the interpretation of data in our realms. Let us now focus on CTIS.

1.1. *The state of the art in keylogging in CTIS*

Keylogging is also a popular data-collection tool in CTIS (Dragsted, 2004, 2005; Dragsted & Carl, 2013; Immonen, 2011; Muñoz & Martín, 2018; Puerini, 2023). As in the case of writing, the study of the mental processes while translating started with think-aloud techniques (Ericsson & Simon, 1993; see also chapter 5, this volume) but soon their validity raised concerns (e.g., Jakobsen, 2003). The first translation-oriented research keylogger was Translog (Jakobsen & Schou, 1999), whose first Windows version was released in 1999. Ten years later, it had featured in over 70 research publications (Schou et al., 2009). Inputlog and Translog II have emerged as the most prominent research keyloggers in CTIS and we will thus focus on them.

In CTIS, researchers have focused on aspects such as the differences between informant groups in linguistic processing, typing behavior, etc. Dragsted (2004), for instance, observed that translators tend to work on segments comprising two to four words, with three-word segments being most frequent. She noted that professionals usually work with fewer, longer text segments, often establishing equivalence at the clause level, whereas less experienced translators work more at the phrase level. Dragsted and Carl (2013) compared professionals and students translating and found that—contrary to Jakobsen’s (2003) findings—students generally engage in more initial planning than professionals. Professionals, on the other hand, were found to allocate more time to end revisions and to be more locally-oriented, possibly due to their proficiency in swiftly generating target text without referring much beyond the immediate co-text, unless they are faced with a structural challenge.

Beyond translating, keyloggers have been employed to scrutinize other multilectal mediated communication tasks. As early as 2006, O'Brien used keylogging to study post-editing and her data did not yield any significant correlation between the duration or location of pauses with text edits. She also noted no substantial difference in effort when post-editing sentences that were either more or less suitable for machine translation (MT). Nitzke and Gros (2021) investigated the tendency to edit beyond post-editing guidelines—referred to as *over-editing*—in both revising and post-editing tasks, and discovered that over-editing was much more common during revision.

Keylogging has also been applied to investigate potential correlations between cognitive demand and translation or source-text difficulty, as indicated by ikis duration and frequency. These measurements serve as proxies for cognitive effort: increased information processing times, more frequent extended ikis, and reduced speed in target-text production suggest heightened cognitive demands during translation (Kruger, 2016). An emerging area of interest is comparing cognitive efforts in translating vs. post-editing machine translations, to shed light on the interface between translators and the constantly evolving machine translation (MT) technologies. For instance, Jia and Sun (2023) comprehensively studied the cognitive demands in both, taking into account source-text complexity and the quality of MT outputs, and found both factors to have a considerable influence in the difficulty associated with post-editing tasks. Specifically, post-editing was deemed less challenging than translating in scenarios where high-quality MT outputs were paired with complex source texts.

Recent CTIS studies have used keylogging to shed light onto cognitive processes involved in tasks like subtitling, audio description, and simultaneous interpreting. Orrego, Dutka and Szarkowska (2018) applied keylogging to assess subtitling performance differences between professionals and trainees, while Jankowska (2021) found that the cognitive operations and phases involved in audio description are akin to those in writing and translation. These latter two studies reported on the use of multiple research methods (discussed below). Keylogging data collection keeps expanding. The latest addition is remote simultaneous interpreting (RSI). Du (2024) explores the use of keystroke logging, screen recording, and audio recording to investigate information-seeking behaviors during glossary compilation and term retrieval before and during RSI. Using constructs such as *ear-key span* (E2K) and *eye-voice span* (I2V), he sheds some light on cognitive strategies linking RSI subtasks over time and suggests improvements in computer-assisted interpreting (CAI) tool design.

Comparative research has also been conducted between translating and other text production tasks. Immonen and Mäkisalo (2010) studied pauses before subordinate clauses and found them to be shorter than those before main clauses in monolingual typing, whereas when translating all pauses before clauses are almost identical in length. This suggests that, irrespective of their syntactic dependency, when translating, clauses are processed independently. Immonen (2011) observed that, when translating, the mean pause lengths between smaller linguistic units (words, phrases) are longer than in monolingual typing, whereas pauses at the sentence and paragraph level are shorter when translating. This, she argued, points to differences in task goals and planning needs. Phrases, for instance, seem to play a more prominent role when translating. Puerini (2023) compared retyping, monolingual writing, translating, revising, and a multimodal task involving monolingual text production based on an infographic leaflet. She corroborated Immonen's findings but also counter-intuitively discovered that durations of infographic-based writing were longer than in both free writing and translation.

Keylogging is often used alongside many other data-collection methods (Wengelin et al., 2019). In writing process studies, Leijten et al. (2013) integrated keylogging and interviews to analyze the chronological progression of tasks, search activities, and the introduction of new content in the professional writing processes of a communication designer. Their study revealed that the professional engaged in thorough searches across various sources, generated both visual and verbal content, and effectively directed attention and maintained motivation throughout the task. Within CTIS, Schrijver, Van Vaerenbergh and Van Waes (2012) combined keylogging with think-aloud techniques to study transediting, observing variations in the distribution of transediting operations across production phases, which appeared to be influenced by participants' individual working styles. Alves and Couto Vale (2017) developed LITTERAE, an application designed to annotate keylogged translation process data to analyze the prototypical characteristics of drafting and revision phases in translation. Enríquez and Cai (2023) used a web screen recorder and a keylogger to scrutinize intra-subject and inter-task level changes in web search behaviors (query time, complexity, language) in translation trainees over a full semester.

The most typical combination, however, has been keylogging and screen eyetracking (see chapter 9, this volume). Schaeffer et al. (2019) explored when and how translators deal with errors in self end-revision, noting that a significant amount of concurrent reading and writing with minimal deletions during the drafting phase, which seems to lead to more streamlined revision processes, as cognitive effort linked to error detection and correction decreases in later stages. Hvelplund (2011) measured the duration of attention units and pupil size to discover that professionals tend to allocate cognitive resources more flexibly and exhibit higher cognitive efficiency than translation trainees.

In the last decade, engineering and business-led research into behavioral dynamics has been drifting towards hand-held devices but, in closing, we should acknowledge the pool of untapped knowledge on keyboard and mouse use that they did and will provide to CTIS. Pinet, Dubarry and Alario (2016) show that typing imposes specific motor constraints on language production processes and Leppich et al. (2023) compiled a dataset aimed at identifying emotions through the keystroke, mouse, and touchscreen dynamics. Keystroke dynamics has been successfully applied to detect (mild) cognitive impairments (Leijten et al., 2015) and in digital health diagnostics, e.g., of multiple sclerosis (Lam et al., 2021) and bipolar disorders (Zulueta et al., 2021). These applications underscore the importance of ethics and privacy in keylogging.

1.2. *Ethical issues in keylogging studies*

Keyloggers that capture and timestamp all activities across programs on a computer—as opposed to those only logging what is typed within an internal text processor—can collect extensive user data. The data may include sensitive details, like passwords, birthdates, phone numbers, bank account information, and personal instant messages. Beyond customary precautions, researchers need to be plain and clear when informing participants regarding what data is collected with their keylogger and the potential privacy risks involved. While commercial keyloggers may redact or remove such sensitive data, research keyloggers usually cannot. Nevertheless, the responsibility of handling this data falls on the researchers, who must commit to removing sensitive information from the dataset. Additionally, minimizing noise data is essential. This means deactivating any applications unrelated to the study that could generate superfluous mouse or keyboard events. This helps prevent capturing irrelevant information that might inadvertently expose personal details.

Furthermore, keylogging research reports must address ethical aspects and provide specifics on how participants were informed about keylogging and how data were redacted and kept confidential. Bias also demands close attention. Most research keyloggers will ask informants to identify themselves, even request some socio-demographic information (e.g., gender, age). Often the informants are a small number of translation and interpreting students, and sometimes their identities can be figured out. Researchers' analyses can also potentially link findings back to individual participants, especially during recruitment, so anonymization is vital. Effective anonymization techniques—like using pseudonyms or self-generated codes based on participant-provided answers—create unique identifiers, safeguarding anonymity while enabling accurate analysis (Calatrava et al., 2022; Yurek et al., 2008).

2. Conceptual aspects

2.1. *Variables in keylogging*

The independent variables in CTIS research projects using keylogging typically pertain to tasks, stimuli, and informant characteristics. The variables we can manipulate depend on study conditions or methodological choices, such as the envisioned *research setting*—e.g., classroom analysis, laboratory experiments under standardized conditions, and naturalistic data collection in workplaces or homes—and the *data collection apparatus,* whether single or multi-method, ensuring compatibility of each source with the project's objectives and other tools. Examples include choosing a keylogger and deciding on the use of translation memories or MT suggestions. Typical independent variables are:

- *Language(s):* Monolingual tasks, language pairs, and translation directionality (e.g., L2→L1, L1→L2).
- *Focus:* The aspect of interest within the task(s), such as planning, information searching, reformulating, revising, and time-constrained performance.
- *Stimuli:* Often texts and speeches, defined by genre, length, and complexity. Text profiling includes factors like lexical density, syntactic complexity, and

conceptual difficulty. Readability formulas should be avoided, due to their atheoretical nature and limited parameter focus.

- *Informant profiles:* Usually, aspects related to language acquisition, personal cognitive features (e.g., working memory), and task experience and expertise, impacting their strategies and efficiency.
- *Task type:* Monolectal, but mostly multilectal mediated communication tasks such as paraphrasing, translating, revising, machine translation post-editing, sight interpreting/translation, audiovisual translating, accessibility tasks, and interpreting (simultaneous, consecutive, dialogic, remote, etc). Often contrasted or combined.

CTIS researchers should meticulously select and manage variables and study conditions to guarantee accurate measurements and consistent results. Scientific rigor is crucial for maintaining credibility of CTIS research. A standard practice is to review related research, understand their rationales and, to enhance comparability, adapt only what is compatible with the new objectives. At the same time, researchers need to make sure that their project measures what it claims to measure. The independent variables need to accurately represent the phenomena or the aspects under study.

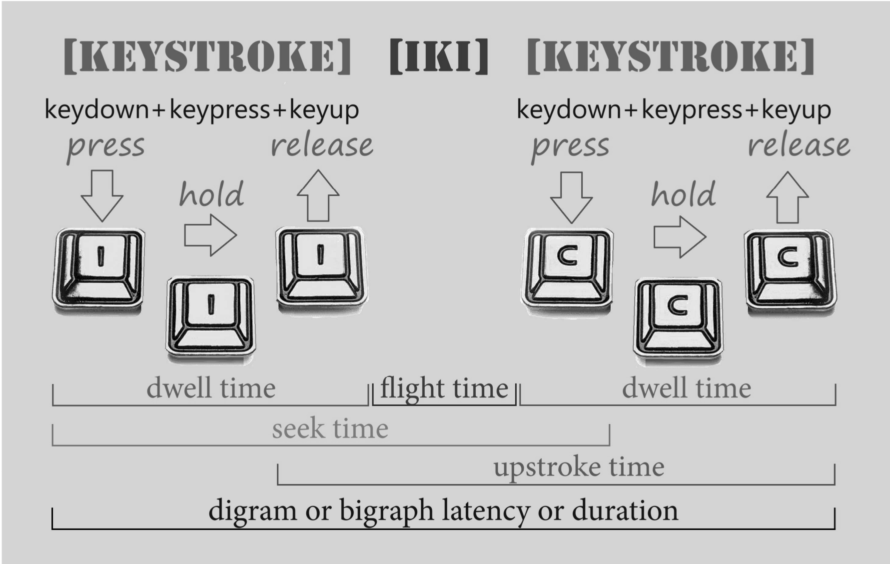

Figure 2. Main time-based metrics for keystrokes and intervals between them.

Dependent variables in keylogging usually involve measurable behavioral aspects of the task, their hypothesized cognitive correlates (such as cognitive effort) and communicative, linguistic and quality aspects of the outcomes that researchers seek to explain or predict. Common metrics include time metrics (iki length, location and frequency, typing speed, burst length), text modifications (additions, edits), text and computer interaction (reading vs. producing), information search, and product quality—often in terms of accuracy, fluency, and conformity to translation norms. In Jia and Sun's (2023) study, the dependent variables were cognitive effort indicators (pause duration, frequency, iki length), translation speed, revision frequency, and product quality. In contrast, Enríquez and Cai (2023) examined web search query time, complexity, language, and search result evaluation time as dependent variables.

Confounding variables may skew the results and lead to incorrect conclusions. In CTIS, common confounders include environmental influences, task-related factors not controlled for (such as source text complexity and MT output quality in Jia and Sun's 2023 study) and individual differences among participants. Such differences may include translator expertise, familiarity with the subject matter, typing proficiency, computer-assisted translation (CAT) tool experience (Jia & Sun 2023), and digital literacy, prior web searching habits, and language proficiency (Enríquez & Cai 2023). Variations in software type and settings can also introduce confounders, so all participants should use uniform tool versions and configurations. Other influences are the physical workspace, noise levels, interruptions, and ergonomic factors. While conducting experiments in controlled laboratory settings helps manage environmental confounders, it may come at the expense of ecological validity (see chapter 2, this volume).

Task-related factors, such as the nature of tasks (e.g., subtitling, revising), length, and characteristics of the stimuli (e.g., source text complexity), and the use of specific CAT tool features, can be confounders as well. Addressing these variables involves standardizing tasks or incorporating task complexity as a covariate in analyses. Using standardized tasks, instructions, and training sessions helps participants focus and enhances control. It also promotes comparability across studies. For this reason, reporting needs to be as complete and thorough as possible. Time constraints influence cognitive processes and strategies and they can be managed by either allocating consistent time or incorporating time pressure as an independent variable in study designs.

Randomization or matching techniques help distribute these confounders evenly across study conditions. Fatigue-related confounders requires careful scheduling of breaks and counterbalancing task order to control for sequence effects. Yet intentionally avoiding fatigue may inadvertently distort the behavior researchers aim to scrutinize, since motivation, freshness, sharpness, attentiveness, responsiveness, and other attributes of performance typically decline over the course of a work shift. Controlling variance in resource use, like online dictionaries and ChatGPT and other generative AI services, involves recording their use as potential covariates. Researchers should consider that providing consistent resources to all participants may again distort behavior, as preferences may vary.

2.2. *Measurement and operationalization in keylogging*

An exhaustive review of indicators and constructs is beyond our scope, so we will focus on ikis and some language and text drafting indicators as examples, and refer readers to the suggested further readings.

Gaps in the task flow (ikis) vary in length and have customarily been interpreted as related to the actions flanking them (Schilperoord, 2001). They have been studied from both linguistic and behavioral perspectives (Conijn et al., 2022; Van Waes et al., 2021). Basically, linguistic approaches study ikis of any length between language units of different sizes—e.g., syllables, roots and prefixes, words, phrases, clauses, sentences, and paragraphs—or before units of different types (e.g., metaphors, Sjørup, 2013). Cognitive processes are attributed to pauses according to their locations; for example, pauses before paragraphs or sentences are possibly related to planning (Immonen, 2006, 2011; Schilperoord, 1996), whereas those between words are possibly related to local processes like lexical search and syntactic construction (Immonen, 2011). Another linguistic perspective focuses on pauses linked to certain grammatical categories (e.g., verbs, noun phrases) and their behavioral implications during editing or translation into another language (e.g., Alves et al., 2010; Heilmann & Neumann, 2016; Serbina et al., 2015, 2017).

Behavioral approaches, conversely, establish arbitrary timespan thresholds to determine where pauses are located within the text, regardless of whether they occur between language units (e.g., Muñoz & Cardona, 2019). There is consensus that the shortest ones are (or need to be considered) noise whereas the longer ones are observable evidence that cognitive resources have been subtracted from typing and possibly reallocated to other activities (Rosenqvist, 2015). As in writing process studies, arbitrary thresholds for extended ikis tend to be used to chunk the translation flow into production bursts that in translation are often known as *translation units*, defined as "segments of the source text, independent of specific size or form, to which the translator's focus of attention is directed" (Alves & Goncalves, 2003, pp. 10–11). That is, longer ikis in the target text are interpreted as boundaries for source-text segments that are processed simultaneously. This conceptualization is different from, but not contradictory with, Muñoz and Apfelthaler's (2022) behavioral notion of *task segments* as performance building blocks flanked by longer ikis that may or may not contain text, but keyboard or mouse activities such as adjusting the music volume or scrolling. To put it simply, not all task segments contain translation units.

Ikis between one to five seconds (same across informants) were often assumed to hint at higher cognitive efforts and, by extension, to translation problems or intrinsic ST difficulty: "Major disruptions generally stem from hiccups either in ST comprehension or in target text reformulation" (Jakobsen, 2019, p. 72). Nonetheless, longer pauses may flag all kinds of processes and shorter pauses may also hint at difficulties (Lacruz et al., 2012; see also Vieira, 2017), suggesting that "a pause may signify both problem-free and problematic processing" Kumpulainen (2015, p. 48). Muñoz and Apfelthaler (2022) distinguish willful, longer ikis *(pauses)* from shorter ones, *respites*, or unintentional typing disfluencies potentially associated with task-related aspects that do not interrupt the text production flow and may become conscious. In their view, chains of respites and typos may signal a minor struggle with attention and mental resource allocation (see figure 3). Typing motor strategies may prioritize efficiency over the integrity of entire language units, indicating that both linguistic and behavioral factors are influential (cf. Heilmann & Neumann, 2016).

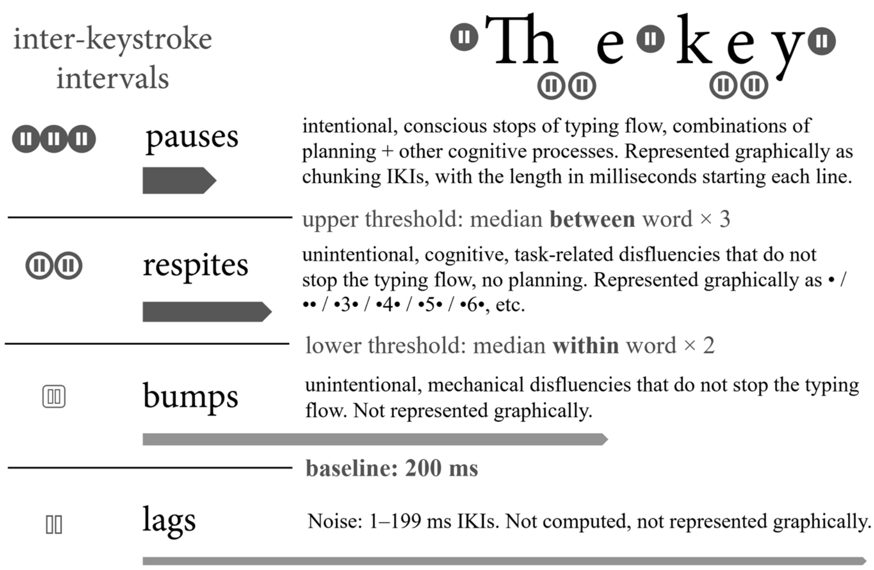

Figure 3. Categorization of interkeystroke intervals in the Task Segment Framework (Muñoz & Apfelthaler, 2022) into *pauses, respites, bumps* and *lags.* Adapted from table published in *The International Journal of Translation & Interpreting Research.*

The frequency and nature of revisions during and after drafting—such as lexical changes, syntactic adjustments, stylistic improvements—can shed light on decision-making, problem-solving, and self-monitoring strategies (Oliveira et al., 2020; Van Waes & Leijten, 2015), as well as on planning, assessment, and other mental processes and constructs (e.g., uncertainty). Temporal data, like changes in translating speed, bursts, and fluctuations in production, can thus be cross-referenced with text and language categories and indicators (e.g., number of words, characters, sentences, etc.) to identify factors impacting performance (Conijn et al., 2019; Leijten & Van Waes, 2013). Variations in typing speed across different task sections and patterns of acceleration or deceleration can reveal how cognitive processes unfold during text production. Combined with keylogging measures such as pause analysis, typing speed variations can yield a deeper understanding of the cognitive dimensions of the translation process (Alves et al., 2007; Dragsted, 2005; Grabowski, 2008).

3. Implementation

3.1. *Research designs in keylogging*

As always, the choice of a research design is contingent on the research question, the nature of the variables of interest, and the practical constraints of working with professional participants. Observational, experimental, quasi-experimental, and mixed methods designs are frequent in CTIS research. *Observational* designs, like case studies and ethnographic projects, are particularly common in didactics and training to investigate task processes in natural settings. Keyloggers are often employed to collect granular data on human-computer interaction without task manipulation, exemplified in studies by Whyatt (2018) and Olalla (2023).

Experimental designs effectively isolate tool features or tasks' impact on metrics like quality, time efficiency, and cognitive effort, as in Jiménez and Casillas (2021) and Rojo et al. (2021). *Quasi-experimental* designs are common when random assignment is impractical or unethical. These designs often compare groups based on inherent characteristics (e.g., expertise) to understand how these factors influence translation processes, as in Schrijver et al. (2016) and Zhang and Torres (2022). *Mixed-methods* research combines quantitative and qualitative approaches, recognized for yielding a deeper and wider understanding of translation phenomena. By integrating keylogger data with outputs from interviews, introspective techniques (see chapter 5), or textual analyses, researchers gain a multi-dimensional perspective on task processes, as in Ehrensberger-Dow and Massey (2014) and Sun et al. (2020).

Another important distinction for keylogging research is that between *cross–sectional* and *longitudinal* studies. Cross-sectional studies collect data from one or more population samples at one specific point in time. Longitudinal designs track task behavior or inferred cognitive processes over time, often in just one sample. They are apt for topics such as monitoring skill development in trainees and technology adoption by professionals, and they provide insights into task practice evolution and long-term tool impacts, as in Lehka-Paul (2020) and Enríquez and Cai (2023). Closely related to the above distinction is another one, based on participant assignment. Research projects may adopt *between-subject* or *within-subject* designs. Between-subjects design involves comparing groups of participants, where such groups are different, as in Dragsted (2005), or where each group receives a different treatment or condition, as in Wang and Sun (2023). Within-subjects designs, in contrast, involve a single group of participants experiencing more than one condition or treatment—e.g., performing more than one task, as in Jia and Sun (2022)—allowing for direct comparisons within individuals. Within-subject designs are gaining popularity because they significantly reduce confounders and extraneous factors linked to individual differences, leading to more precise and reliable results (e.g., Puerini, 2023).

Given the still relatively modest sample sizes in keylogging studies (see §4.2), striving for greater comparability is imperative to achieve more comprehensive and robust findings. While research methods standards are improving, there is room for optimization. A key step in this direction involves using consistent metrics for variables such as typing speed, pause duration, revision frequency, and textual output measures. Meanwhile, reporting keylogging research needs to be particularly meticulous and align with standard empirical reports, usually in the IMRaD format (e.g., Cuschierin et al., 2018). The methods section needs to be exhaustive, clear, and detailed, covering both data collection and analysis to support replicability (enabling other researchers to conduct similar studies) and reproducibility (allowing scrutiny at each step). All software versions used in the project need to be listed. Researchers also need to explain the rationales behind their methodological choices, from qualitative analysis techniques to the choice of statistical tests. Using a pre-registered report template from the Open Science Framework as a checklist can be very helpful.[2]

3.2. *Current popular CTIS research keyloggers*

Commercial keyloggers often collect audio and video data, which can greatly simplify research data collection and alignment. Open access keyloggers, some also open source, are available on online platforms like SourceForge and GitHub. They may offer fewer features and be less sophisticated than commercial options, but they generally record keystrokes effectively and are more easily detectable—which is advantageous. However, these keyloggers typically focus on *keydown* events (pressing the key) and do not register *keyup* (key release) events, potentially compromising their precision, especially concerning millisecond-level accuracy. When no keyup is registered, events are computed keydown to keydown, which means that prior IKI and keypress durations are lumped together. However, IKIs seem to display more variation and be more sensitive to cognitive changes than keypresses do. Furthermore, in most cases, open access keyloggers are distributed as source code (without a graphic user interface), which requires some programming ability from the user. Additionally, they often do not guarantee long-term support or maintenance, which can affect their version updates and overall dependability.

Translog-II, a substantial advancement from earlier versions, was introduced by Carl (2012a). Translog-II has two components. A *User* component records task sessions within a simple internal text processor, and a *Supervisor* component replays the xml log files and analyzes them (it also creates project files). Translog-II logs all keystrokes and mouse operations within a rudimentary internal text processor, categorizing them as

[2] https://osf.io/

insertions, deletions (including delete and backspace actions), navigation (cursor movements), copy/cut-and-paste operations, return key presses, and mouse operations.[3] It also records spoken language. When integrated with an eye-tracker, it records an even wider array of *user activity data* (such as gaze-sample points), computes fixations (clusters of gaze-samples), and maps them to the nearest character on the screen. The simultaneous recording of gaze and keystroke information, which both reference textual positions, allows researchers to investigate aspects such as attention allocation patterns, information retrieval behaviors, text segmentation, problem-solving strategies, and tool interactions. Data is stored in xml format for subsequent analysis within Translog-II or with external tools. Translog-II supports alphabetic languages as well as Chinese and Japanese, although the latter requires additional offline mapping steps.

Carl (2012b) also established the Translation Process Research Database (TPR-DB) to support empirical research in translation and text processing. TPR-DB initially hosted Translog datasets and now also includes contributions from projects like CASMACAT, focusing on machine translation post-editing.[4] It recently added conversion of Trados Studio keylogging data (Qualitivity) into Translog-II format (Yamada et al., 2022). With over 500 hours of text production data from 3000+ sessions and a substantial corpus of translated words in multiple languages, the public section of the TPR-DB is a rich repository that fosters collaboration and knowledge sharing. In a private section, researchers can securely manage data and upload, process, analyze, generate tables, and download Translog-II-compatible datasets.

Inputlog (Leijten & Van Waes, 2013) is also available for free after registration.[5] It comprises five modules: *record, pre-process, analyze, post-process,* and *play* (see Figure 4). The *record* module logs and timestamps keyboard and mouse events in MS Word, capturing details such as character position, document length, and

---

[3] https://sites.google.com/site/centretranslationinnovation/translog-ii
[4] https://www.casmacat.eu/
[5] https://www.inputlog.net/

copy/paste/move actions. It also monitors what happens in other Windows-based programs, and online activities as "focus" events. It also registers speech data. The *pre-process* module enables data processing from various angles, such as event-based (keyboard, mouse, and speech), time-based, or based on active window changes, and allows integration of log files from other tools, like Morae, Dragon Naturally Speaking, and Tobii eyetracking data. The *analyze* module—the core of the program—will be addressed below. The *post-process* module integrates log files for further statistical analysis, and the *play* module replays recorded sessions based on data, not on video, but it is prone to malfunction.

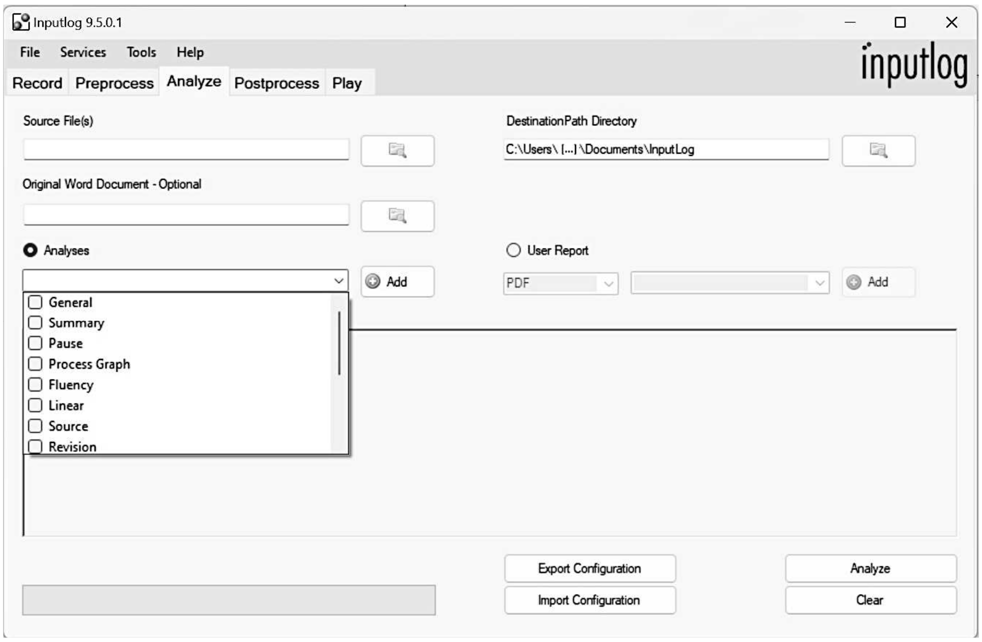

Figure 4. Interface of the Analyze tab in Inputlog 9.5 (December, 2023).

Inputlog's *analysis module* offers a range of ready-made scrutinies. *General analysis* includes basic log file details, event contents, position, time stamp, duration, and prior iki time. *Summary analysis* breaks down into five sections: process information (e.g., number of words produced), product information (e.g., number of words in the final product), product/process information (e.g., ratios), process time, and writing mode.

Linear analysis presents the evolving text step by step, including pauses exceeding a given threshold. *Pause analysis* provides insights into pausing behavior, cataloging their number, duration, location within and between words, production bursts, and variation per interval or period (see Figure 3). The *process graph* offers visual representations of text production, movements, product status, and source interaction. *Revision analysis* identifies and summarizes all instances of deletions and insertions. Inputlog also provides flexible, customized reports combining several analyses.

Inputlog also allows researchers to measure their informants' typing skills through an *Inputlog copy task* available in at least twelve languages, which provides several measures for a fine-grained analysis of low-level typing and motor skills (Van Waes et al., 2019). This brief supplementary task permits researchers to account for typing proficiency, a significant factor in online digital text production (cf. Dhakal et al., 2018; Song et al., 2001). The copy task can be initiated via the record module or directly through the Inputlog website.

4. Closing remarks

Research keyloggers like Inputlog and Translog-II are user-friendly and available at no cost. Not everything are advantages, of course, for keylogging does come with its own set of challenges.

4.1. *Challenges: How many participants are enough?*

Since the mid-2000s, there is a trend towards more participants in CTIS studies for increased generalizability (Balling & Hvelplund, 2015). To identify broader patterns, quantitative studies often include 20 participants or more. Balancing sample size is essential in experimental design. Over-recruitment wastes resources and sometimes detects trivial differences, while insufficient participants can obscure meaningful individual differences. Qualitative research may involve a small cohort for an in-depth exploration of individual translation processes. In quantitative projects, despite existing guidelines, calculating an appropriate sample size involves various considerations, like the study's nature, design, the number of conditions, statistical confidence, sensitivity, measurement variability, and meaningful differences (Gergle & Tan, 2014).

In factorial designs, dividing participants into groups exacerbates their scarcity. Structuring participant groups and determining their sizes must balance the depth of inquiry with practical feasibility. Choosing between homogeneity and heterogeneity in participant characteristics is crucial, and homogeneity is usually preferred in factors like language proficiency. A statistical power analysis helps ascertain the requisite sample size for detecting anticipated effects with sufficient power and significance levels. For studies targeting strong associations, a smaller sample might be preferable, reducing data coding efforts and concentrating on robust associations. As suggested, intra-subject designs mitigate the influence of confounders. In any case, to refine the experimental design and understand participant responses, conducting a pilot study is recommended. This step is vital for finalizing participant count and group configurations.

For data analysis, researchers using keylogging must carry out a thorough preliminary review to identify any data anomalies or inconsistencies. This usually involves *data cleaning* to refine the dataset for analysis. Researchers need to be aware that keylogging yields a substantial amount of data that demands considerable processing time. Checks for inter-rater reliability in qualitative analyses or sensitivity analyses for quantitative methods to assess the robustness of the findings are a good idea. Descriptive statistics offer an overview, while inferential statistics, using techniques such as regression analysis, ANOVA, or mixed-effects modeling, facilitate hypothesis testing and the examination of relationships between variables, tailored to the research questions and data structure (see, e.g., Mellinger & Hanson, 2016).

Correlating results from diverse methods or tools, such as aligning screen recording data with keystroke logs, can enhance the validity of findings by providing a multifaceted view of the translation process. Reusing or pooling data from previous studies is an option, but necessitates careful consideration for data integrity.

4.2. *Advantages and disadvantages of keylogging*

Keylogging stands out in CTIS as a nearly unparalleled tool for non-obtrusive quantitative data collection capabilities. That is, keyloggers typically cause minimal interference with task performance, operate discreetly, and often fade away from participants’ conscious awareness in a matter of minutes. Particularly effective for grounding and substantiating hypothetical explanations for complex communicative behaviors, keyloggers record all computer input and provide copious quantities of precise information up to millisecond accuracy. They also work well for priming and response time tests, and easily complement other data-collection tools such as audio and screen recording.

The ecological validity of lab-based keylogging is a concern, as it may not fully represent real-world translation behaviors (see chapter 2, this volume), particularly given the increasing reliance on technologies like MT, translation memories, translation memory clouds and online collaborative platforms. With rapidly evolving AI systems, like ChatGPT, researchers must adapt to new interfaces to comprehend their impact on cognitive processes. Controlling task conditions becomes harder, and necessitates alternative approaches like ethnographic keylogging for studying translation in its natural habitat (Xiao & Muñoz, 2020). Relying solely on laboratory studies risks overlooking insights from professional translation contexts (Risku, 2017). Static datasets also fail to consider how translators’ individual styles co-evolve with emerging technologies throughout their careers—underscoring the need for longitudinal studies.

Keyboard-based interaction has become only part of the story, as translation incorporates more multimodal elements (Teixeira et al., 2019) and performance is re-embedded in its interaction with the environment (Mellinger, 2023; Muñoz, 2021; Risku & Rogl, 2021). While eyetrackers and synchronizing keystroke data with screen recordings capture aspects of these interactions, fully understanding multimodal translation behavior may require new tools. Capturing and understanding the situated dimensions of cognition may also require reimagining the unit of analysis above and below the individual translation session. This may involve new annotation schemes to integrate multidimensional data like movements, gestures, glances to real-world objects and interactions between collaborating translators, whether co-located or distributed across networks. Sensor technologies may enable capturing aspects of translation currently opaque. However, this *big data keylogging* also risks further compromising participant privacy if not handled judiciously.

Another source of concern is that keyloggers were initially developed for alphabetical writing systems, mainly the Latin alphabet. Translog-II has broadened its reach to include UTF-8 character encoding, accommodating a diverse array of scripts beyond Latin and Cyrillic, such as Arabic, Chinese, Greek, Hebrew, Japanese, Korean, and many others. In order to do so, it records changes in the display of emerging texts, rather than tapping the keyboard output, so technically it is not a keylogger any longer (Carl et al., 2016). This may be crucial for some research questions but totally irrelevant for others (e.g., are non-alphanumeric keystrokes relevant in your project?). Inputlog currently supports the Latin alphabet and Chinese, with plans to extend to Korean, but other languages remain unsupported. Platform dependency is another significant limitation, with many keyloggers functioning solely on specific operating systems like Windows or macOS.

In conclusion, keylogging has proven to be a powerful tool in CTIS research, offering valuable insights into the cognitive processes underlying translation and interpreting tasks. However, researchers must be aware of its limitations and the challenges it presents, particularly in terms of ecological validity, the need for new tools to capture multimodal translation behavior, and the potential risks to participant privacy.

Further readings on keylogging

Feit, A. M., Weir, D., & Oulasvirta, A. (2016). How we type: movement strategies and performance in everyday typing. *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 4262–4273). ACM. https://doi.org/10.1145/2858036.2858233

Lindgren, E., & Sullivan, K. (Eds.). (2019). *Observing writing: Insights from keystroke logging and handwriting.* Brill. https://doi.org/ 10.1163/978900439252

Manchón Ruiz, R. M., & Roca de Larios, J. (2023). *Research methods in the study of L2 writing processes.* John Benjamins. https://doi.org/10.1075/rmal.5

Shadman, R., Wahab, A. A., Manno, M., Lukaszewski, M., Hou, D., & Hussain, F. (2023). *Keystroke dynamics: Concepts, techniques, and applications.* arXiv. https://doi.org/10.48550/arXiv.2303.04605

Zhu, M., Zhang, M. and Deane, P. (2019). Analysis of keystroke sequences in writing logs. *ETS research report series, 2019:* 1-16. https://doi.org/10.1002/ets2.12247

References

Ahlsén, E., & Strömqvist, S. (1999). ScriptLog: A tool for logging the writing process and its possible diagnostic use. In F. Loncke, J. Clibbens, H. Arvidson & L. Lloyd (Eds.). *Augmentative and alternative communication. New directions in research and practice* (pp. 144–149). Wiley.

Allen, L. K., Jacovina, M. E., Dascalu, M., Roscoe, R. D., Kent, K. M., Likens, A. D., & McNamara, D. S. (2016). {ENTER} ing the Time Series {SPACE}: Uncovering the writing process through keystroke analyses. In T. Barnes, M. Chi, and M. Feng (Eds.) *Proceedings of the 9th international conference on educational data mining* (pp. 22–29). Arizona State University.

Alves, F., & Couto Vale, D. (2017). On drafting and revision in translation: a corpus linguistics oriented analysis of translation process data. In S. Hansen-Schirra, S. Neumann & O. Čulo (Eds.) *Annotation, exploitation and evaluation of parallel corpora. TC3 I* (pp. 81–101). Language Science Press.

Alves, F. & Gonçalves, J. L. (2003). A relevance theory approach to the investigation of inferential processes in translation. In F. Alves (Ed.) *Triangulating translation: Perspectives in process oriented research* (pp. 3–24). John Benjamins.

Alves, F., A. S. Pagano, S. Neumann, E. Steiner & S. Hansen-Schirra. (2010). Translation units and grammatical shifts: Towards an integration of product- and process-based translation research. In G. M. Shreve & E. Angelone (Eds.) *Translation and cognition* (pp. 109–142). John Benjamins. http://dx.doi.org/10.1075/ata.xv.07alv

Alves, R. A., Castro, S. L., Sousa, L., & Strömqvist, S. (2007). Typing skill and pause–execution cycles in written composition. In M. Torrance, L. Van Waes & D. Galbraith (Eds.). *Writing and cognition: Research and applications* (pp. 55–65). Elsevier.

Baaijen, V. M., Galbraith, D., & de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication, 29*(3), 246-277. https://doi.org/10.1177/0741088312451108

Balling, L. W., & Hvelplund, K. T. (2015). Design and statistics in quantitative translation (process) research. *Translation Spaces, 4*(1), 169-186.

Bridwell, L., Sirc, G., Brooke, R. (1985). Revising and computing: Case studies of student writers. In S. Freedman (Ed.) *The acquisition of written language* (pp. 172–195). Ablex.

Calatrava Martínez, M., De Irala Estévez, J., Osorio de Rebellón, A., Benítez Sastoque, E. R., & Lopez del Burgo, C. (2022). Matched and fully private? A new self-generated identification code for school-based cohort studies to increase perceived anonymity. *Educational and Psychological Measurement*, *82*(3), 465–481. https://doi.org/10.1177/00131644211035436

Carl, M. (2012a). *Translog-II: A program for recording user activity data for empirical translation process research.* Paper presented at the 8th international conference on language resources and evaluation, LREC 2012, Istanbul, Turkey.

Carl, M. (2012b). The CRITT TPR-DB 1.0: A database for empirical human translation process research. In *Workshop on post-editing technology and practice*, San Diego, California, USA. AMTA.

Carl, M., Bangalore, S. & Schaeffer, M. (2016). Introduction and overview. In M. Carl, S. Bangalore & M. Schaeffer (Eds.) *New directions in empirical translation process research* (pp. 3–12). Springer.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing. *Written Communication, 18*(1), 80–98. https://doi.org/10.1177/0741088301018001004

Chukharev-Hudilainen, E. (2014). Pauses in spontaneous written communication: A keystroke logging study. *Journal of Writing Research, 6*(1), 61–84. https://doi.org/10.17239/jow

Conijn, R., Cook, C., Van Zaanen, M., & Van Waes, L. (2022). Early prediction of writing quality using keystroke logging. *International Journal of Artificial Intelligence in Education, 32*(4), 835–866. https://doi.org/10.1007/s40593-021-00268-w

Conijn, R., Roeser, J., & van Zaanen, M. (2019). Understanding the keystroke log: The effect of writing task on keystroke features. *Reading and Writing, 32*(9), 2353–2374. https://doi.org/10.1007/s11145-019-09953-8

Cuschieri, S., Grech, V., Savona Ventura, C. (2018). WASP (Write a Scientific Paper): How to write a scientific thesis. *Early Human Development*, *127*, 101–105. https://doi.org/10.1016/j.earlhumdev.2018.07.012

Dhakal, V., Feit, A. M., Kristensson, P. O., & Oulasvirta, A. (2018). Observations on typing from 136 million keystrokes. *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–12). ACM. https://doi.org/10.1145/3173574.3174220

Du, Z. (2024). *Bridging the gap. Exploring the cognitive impact of InterpretBank on Chinese interpreting trainees.* [Unpublished doctoral dissertation].University of Bologna.

Dragsted, B. (2004). *Segmentation in translation and translation memory systems. An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process.* Samfundslitteratur.

Dragsted, B. (2005). Segmentation in translation: Differences across levels of expertise and difficulty. *Target, 17*(1), 49–70. https://doi.org/10.1075/target.17.1.04dra

Dragsted, B., & Carl, M. (2013). Towards a classification of translation styles based on eye-tracking and key-logging data. *Journal of Writing Research, 5*(1), 133-158. https://doi.org/10.17239/jowr-2013.05.01.6

Duin, A. H., & Bridwell, L. S. (1985). Looking in-depth at writers: Computers as writing medium and research tool. In J. L. Collins, & E. A. Sommers (Eds.), *Writing on-line: Using computers in the teaching of writing* (pp. 115-121). Boynton/Cook.

Ehrensberger-Dow, M. & Massey, G. (2014). Cognitive ergonomic issues in professional translation. In J. W. Schwieter & A. Ferreira (Eds.). *The development* *of translation competence: Theories and methodologies from psycholinguistics and cognitive science* (pp. 58-86). Cambridge Scholars.

Enríquez Raído, V. & Cai, Y. (2023). Changes in web search query behavior of English-to-Chinese translation trainees. *Ampersand, 11:* 100137. https://doi.org/10.1016/j.amper.2023.100137

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). The MIT Press.

Galbraith, D., & Baaijen, V. M. (2019). Aligning keystrokes with cognitive processes in writing. In E. Lindgren & K. Sullivan (Eds.) *Observing writing*. *Insights from keystroke logging and handwriting* (pp. 306–325). Brill. https://doi.org/10.1163/9789004392526_015

Gergle, D., & Tan, D. S. (2014). Experimental research in HCI. In J. S. Olson & W. A. Kellogg (Eds.), *Ways of knowing in HCI* (pp. 191–227). Springer. https://doi.org/10.1007/978-1-4939-0378-8_9

Grabowski, J. (2008). The internal structure of university students' keyboard skills. *Journal of Writing Research, 1*(1), 27–52. 10.17239/jowr-2008.01.01.2

Heilmann, A. & Neumann, S. (2016). Dynamic pause assessment of keystroke logged data for the detection of complexity in translation and monolingual text production. In D. Brunato, F. Dell'Orletta, G. Venturi, T. François & Ph. Blache (Eds.). *Proceedings of the workshop on computational linguistics for linguistic complexity* (pp. 98–103). ACL. https://aclanthology.org/W16-4111

Hvelplund, K. T. (2011). *Allocation of cognitive resources in translation: An eye-tracking and key-logging study*. [Unpublished doctoral dissertation]. Copenhagen Business School.

Immonen, S. (2006). Translation as a writing process: Pauses in translation versus monolingual text production. *Target, 18*(2), 313-336. https://doi.org/10.1075/target.18.2.06imm

Immonen, S. (2011). Unravelling the processing units of translation. *Across Languages and Cultures, 12*(2), 235–257. https://doi.org/10.1556/Acr.12.2011.2.6

Immonen, S., & Mäkisalo, J. (2010). Pauses reflecting the processing of syntactic units in monolingual text production and translation. *Hermes, 23*(44), 45-61. 10.7146/hjlcb.v23i44.97266

Jakobsen, A. L. (2003). Effects of think aloud on translation speed, revision, and segmentation. In *Triangulating translation* (pp. 69–95). John Benjamins. https://doi.org/10.1075/btl.45.08jak

Jakobsen, A. L. (2019). Segmentation in translation: A look at expert behaviour. In D. Li, V. L. C. Lei & Y. He (Eds.) *Researching cognitive processes of translation* (pp. 71–108). Springer.

Jakobsen, A. L. & Schou, L. 1999. Translog documentation, Version 1.0. In G. Hansen (Ed.) *Probing the process in translation: Methods and results* (pp. 151-186). Samfundslitteratur.

Jankowska, A. (2021). Audio describing films: A first look into the description process. *JoSTrans, 36*, 26–52. https://jostrans.soap2.ch/issue36/art_jankowska.php

Jia, Y., & Sun, S. (2023). Man or machine? Comparing the difficulty of human translation versus neural machine translation post-editing. *Perspectives, 31*(5), 950-968. https://doi.org/10.1080/0907676X.2022.2129028

Katerina, T., & Nicolaos, P. (2018). Mouse behavioral patterns and keystroke dynamics in End-User Development: What can they tell us about users' behavioral attributes? *Computers in Human Behavior, 83,* 288–305. https://doi.org/10.1016/j.chb.2018.02.012

Kruger, H. (2016). What's happening when nothing's happening? Combining eyetracking and keylogging to explore cognitive processing during pauses in translation production. *Across Languages and Cultures, 17*(1), 25–52. https://doi.org/10.1556/084.2016.17.1.2

Kumpulainen, M. (2015). On the operationalisation of ‘pauses’ in translation process research. *Translation & Interpreting, 7*(1), 47–48. http://trans-int.org/index.php/transint/article/view/367/183

Lacruz, I., Shreve, G. M. & Angelone, E. 2012. Average pause ratio as an indicator of cognitive effort in post-editing: A case study. In S, O’Brien, M. Simard & L. Specia (Eds.) *Proceedings of the workshop on post-editing technology and practice* [s.p.]. https://aclanthology.org/2012.amta-wptp.3

Lam, K., Meijer, K., Loonstra, F., Coerver, E., Twose, J., Redeman, E., Moraal, B., Barkhof, F., de Groot, V., Uitdehaag, B., & Killestein, J. (2021). Real-world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis. *Multiple Sclerosis Journal*, *27*(9), 1421–1431. https://doi.org/10.1177/1352458520968797

Lehka-Paul, O. (2020). Behavioural indicators of translators' decisional styles in a translation task: A longitudinal study. *Yearbook of the Poznań Linguistic Meeting 6*(1), 81–111. https://doi.org/10.2478/yplm-2020-0007

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication, 30*(3), 358–392. https://doi.org/10.1177/0741088313491692

Leijten, M., Van Waes, L., & Van Horenbeeck, E. (2015). Analyzing writing process data: A linguistic perspective. In G. Cislaru (Ed.), *Writing(s) at the crossroads* (pp. 277–302). John Benjamins. https://doi.org/10.1075/z.194.14lei

Leppich, D., Bieber, C., Proschek, K., Harms, P., & Schubert, U. (2023). DUX: A dataset of user interactions and user emotions. *I-Com, 22*(2), 101–123. https://doi.org/10.1515/icom-2023-0014

Manchón Ruiz, R. M. & Roca de Larios, J. (2023). The study of L2 writing processes: Lines and methods of inquiry. In R. M. Manchón Ruiz & J. Roca de Larios (Eds.). *Research methods in the study of L2 writing processes* (pp. 6-31). John Benjamins.

Mellinger, C. D. (2023). Embedding, extending, and distributing interpreter cognition with technology. In G. Corpas Pastor & B. Defrancq (Eds.) *Interpreting technologies—Current and future trends* (pp. 195–216). John Benjamins. https://doi.org/10.1075/ivitra.37.08mel

Mellinger, C., & Hanson, T. (2016). *Quantitative research methods in translation and interpreting studies*. Routledge.

Muñoz Martín, Ricardo. (2021). Situated cognition. In Y. Gambier & L. van Doorslaer (Eds.) *Handbook of translation studies*, vol. 5 (pp. 207–212). John Benjamins. https://doi.org/10.1075/hts.5.sit1

Muñoz Martín, R., & Apfelthaler, M. (2022). A task segment framework to study keylogged translation processes. *Translation & Interpreting, 14*(2), 8-31. https://doi.org/10.12807/ti.114202.2022.a02

Muñoz Martín, R. & Cardona Guerra, J. M. (2019). Translating in fits and starts: pause thresholds and roles in the research of translation processes. *Perspectives, 27*(4), 525–551. https://doi.org/10.1080/0907676X.2018.1531897

Muñoz Martín, R., & Martín de León, C. (2018). Fascinatin' rhythm—And pauses in translators' cognitive processes. *Hermes, 57,* 29–47. https://doi.org/10.7146/hjlcb.v0i57.106192

Nitzke, J. & Gros, A. K. (2021) Preferential changes in revision and post-editing. In M. Koponen, B. Mossop, I. Robert & G.a Scocchera (Eds.) *Translation revision and post-editing*: *Industry practices and cognitive processes* (pp. 21–34). Routledge.

O'Brien, S. (2006). Pauses as indicators of cognitive effort in post-editing machine translation output. *Across Languages and Cultures, 7*(1), 1–21. https://doi.org/10.1556/ACROSS.7.2006.1

Olalla Soler, C. (2023). Literal vs. default translation. Challenging the constructs with middle Egyptian translation as an extreme case in point. *Sendebar, 34,* 65–92. https://doi.org/10.30827/sendebar.v34i0.27090

Oliveira, E. A., Conijn, R., De Barba, P. & Tresize, K. (2020). Writing analytics across essay tasks with different cognitive load demands. Paper presented at ASCILITE 2020.

Orrego Carmona, D., Dutka, L., & Szarkowska, A. (2018). Using translation process research to explore the creation of subtitles: An eye-tracking study comparing professional and trainee subtitlers. *JoSTrans, 30,* 150-180. https://jostrans.org/issue30/art_orrego-carmona_et_al.php

Pinet, S., Dubarry, A.-S., & Alario, F.-X. (2016). Response retrieval and motor planning during typing. *Brain and Language, 159*, 74–83. https://doi.org/10.1016/j.bandl.2016.05.012

Puerini, S. (2023). Text-production tasks at the keyboard. Linguistic and behavioral contrasts. *Translation, Cognition & Behavior*, 6(1), 29-59. https://doi.org/10.1075/tcb.00075.pue

Risku, H. (2017). Ethnographies of translation and situated cognition. In J. W. Schwieter & A. Ferreira (Eds.). *The handbook of translation and cognition* (pp. 290-310). New York: Wiley.

Risku, H., & Rogl, R. (2021). Translation and situated, embodied, distributed, embedded and extended cognition. In F. Alves & A. L. Jakobsen (eds.) *The Routledge handbook of translation and cognition* (pp. 478–499). Routledge.

Rojo López, A., Cifuentes Férez, P. & Ramos Caro, M. (2021). The role of creativity on the translation of motion verbs: Data on the translation product and process. *Onomázein, 52,* 99–126.

Rosenqvist, S. (2015). *Developing pause thresholds for keystroke logging analysis.* [Unpublished BA thesis]. University of Umea.

Sala-Bubaré, A., Castelló, M., & Rijlaarsdam, G. (2021). Writing processes as situated regulation processes: A context-based approach to doctoral writing. *Journal of Writing Research 13*(1): 299–328. https://doi.org/10.17239/jowr-2021.13.01.01

Schaeffer, M., Tardel, A., Hofmann, S., & Hansen-Schirra, S. (2019). Cognitive effort and efficiency in translation revision. In E. Huertas Barros, S. Vandepitte & E. Iglesias Fernández (Eds.) *Quality assurance and assessment practices in translation and interpreting* (pp. 226-243). IGI Global. https://doi.org/10.4018/978-1-5225-5225-3.ch010

Schilperoord, J. (1996). The distribution of pause time in written text production. In G. Rijlaarsdam, H. van den Bergh & M. Couzjin (Eds.) *Theories, models and methodology in writing research.* Amsterdam University Press. 21-35.

Schilperoord, J. (2001). On the cognitive status of pauses in discourse production. In T. Olive & C. M. Levy (Eds.). *Contemporary tools and techniques for studying writing* (pp. 61–90). Kluwer.

Schou, L, Dragsted, B. & Carl, M. 2009. Ten years of Translog. In I. Mees, F. Alves and S. Göpferich (eds). *Methodology, technology and innovation in translation process research* (pp. 37–48). Samfundslitteratur.

Schrijver, I., Van Vaerenbergh, L., Leijten, M. & Van Waes, L. (2016). The impact of writing training on transediting in translation, analyzed from a product and process perspective. *Perspectives, 24*(2), 218–234. https://doi.org/10.1080/0907676X.2015.1040034

Schrijver, I., Van Vaerenbergh, L., & Van Waes, L. (2012). An exploratory study of transediting in students' translation processes. *Hermes, 25*(49), 99–117. https://doi.org/10.7146/hjlcb.v25i49.97740

Serbina, T., Hintzen, S., Niemietz, P. & Neumann, S. (2017). Changes of word class during translation—Insights from a combined analysis of corpus, keystroke logging and eye-tracking data. In S. Hansen-Schirra, O. Czulo & S. Hofmann (Eds.) *Empirical modelling of translation and interpreting* (pp. 177–208). Language Science Press.

Serbina, T., Niemietz, P., Fricke, M., Meisen, P. & Neumann. S. (2015). Part of speech annotation of intermediate versions in the keystroke logged translation corpus. In A. Meyers, I. Rehbein & H.e Zinsmeister (Eds.) *Proceedings of the 9th linguistic annotation workshop* (pp. 102–111). ACL. https://aclanthology.org/W15-1611

Severinson Eklundh K. S., Kollberg P. (1992). Translating keystroke records into a general notation for the writing process. Department of Numerical Analysis and Computing Science, Royal Institute of Technology.

Shadman, R., Anu Wahab, A., Manno, M., Lukaszewski, M., Hou, D., & Hussain, F. (2023). Keystroke dynamics: Concepts, techniques, and applications. *arXiv e-prints.* https://doi.org/10.48550/arXiv.2303.04605.

Shi, Y., Wang, X., Zheng, K., & Cao, S. (2023). User authentication method based on keystroke dynamics and mouse dynamics using HDA. *Multimedia Systems, 29*(2), 653–668. https://doi.org/10.1007/s00530-022-00997-5

Sjørup, A. C. (2013). *Cognitive effort in metaphor translation: An eye-tracking and key-logging study.* [Unpublished doctoral duissertation]. Copenhagen Business School. http://hdl.handle.net/10419/208853

Song, D. X., Wagner, D., & Tian, X. (2001). Timing analysis of keystrokes and timing attacks on SSH. In *Proceedings of the 10th USENIX security symposium.*

Strömqvist, S., & Karlsson, H. (2002). *ScriptLog for Windows: User's manual* (technical report). Lund University and University College of Stavanger.

Sun, S., Li, T., & Zhou, X. (2020). Effects of thinking aloud on cognitive effort in translation. *Linguistica Antverpiensia, New Series–Themes in Translation Studies, 19*. https://doi.org/10.52034/lanstts.v19i0.556

Teixeira, C. S. C., Moorkens, J., Turner, D., Vreeke, J., & Way, A. (2019). Creating a multimodal translation tool and testing machine translation integration using touch and voice. *Human Brain Mapping, 45*(2), Article e26617. https://doi.org/10.1002/hbm.26617

Thomas, P. A., & Preetha Mathew, K. (2023). A broad review on non-intrusive active user authentication in biometrics. *Journal of Ambient Intelligence and Humanized Computing, 14*(1), 339–360. https://doi.org/10.1007/s12652-021-03301-x

Tsimperidis, I., & and Arampatzis, A. (2020). The keyboard knows about you: revealing user characteristics via keystroke dynamics. *International Journal of Technoethics 11*(2), 34–51. https://EconPapers.repec.org/RePEc:igg:jt0000:v:11:y:2020:i:2:p:34-51.

Usoof, H., Leblay, C. & Caporossi, G. (2020). GenoGraphiX-Log version 2.0 user guide. Technical report. *Les Cahiers du GERAD* G–2020–68, GERAD, HEC Montreal, Canada. https://www.gerad.ca/en/papers/G-2020-68/

Van Waes, L. & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition, 38,* 79–95.

Van Waes, L., Leijten, M., Pauwaert, T., & Van Horenbeeck, E. (2019). A multilingual copy task: Measuring typing and motor skills in writing with Inputlog. *Journal of Open Research Software, 7*(1:30), 1-8. https://doi.org/10.5334/jors.234

Van Waes, L., Leijten, M., Roeser, J., Olive, T., & Grabowski, J. (2021). Measuring and assessing typing skills in writing research. *Journal of Writing Research, 13*(1), 107–153. https://doi.org/10.17239/jowr-2021.13.01.04

Vieira, L Nunes. (2017). How do measures of cognitive effort relate to each other? A multivariate analysis of post-editing process data. *Machine Translation, 30*(1), 41-62. https://doi.org/10.1007/s10590-016-9188-5

Wang, L. & Sun, S. (2023). Dictating translations with automatic speech recognition: Effects on translators' performance. *Frontiers in Psychology 14,* 1108898. https://doi.org/10.3389/fpsyg.2023.1108898

Wengelin, Å., Frid, J., Johansson, R., & Johansson, V. (2019). Combining keystroke logging with other methods: Towards an experimental environment for writing process research. In E. Lindgren & K. Sullivan (Eds.) *Observing writing* (pp. 30–49). Brill. https://doi.org/10.1163/9789004392526_003

Whyatt, B. (2018). Testing indicators of translation expertise in an intralingual task. *Hermes, 57,* 63–78. https://doi.org/10.7146/hjlcb.v0i57.106194

Xiao, K. & Muñoz Martín, R. (2020). Cognitive Translation Studies: Models and methods at the cutting edge. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19,* 1–24.

Yamada, M., T. Mizowaki, L. Zou & M. Carl. (2022). Trados-to-Translog-II: Adding gaze and qualitivity data to the CRITT TPR-DB. In *Proceedings of the 23rd annual conference of the EMTA* (pp. 295–296). EAMT. https://aclanthology.org/2022.eamt-1.37.pdf

Yurek, L. A., Vasey, J., & Sullivan Havens, D. (2008). The use of self-generated identification codes in longitudinal research. *Evaluation Review*, *32*(5), 435–452. https://doi.org/10.1177/0193841X08316676

Zhang, H., & Torres Hostench, O. (2022). Training in machine translation post-editing for foreign language students. *Language Learning & Technology, 26*(1), 1–17. http://hdl.handle.net/10125/73466

Zulueta, J., Demos, A. P., Vesel, C., Ross, M., Piscitello, A., Hussain, F., Langenecker, S. A., McInnis, M., Nelson, P., Ryan, K., Leow, A., & Ajilore, O. (2021). The effects of bipolar disorder risk on a mobile phone keystroke dynamics based biomarker of brain age. *Frontiers in Psychiatry, 12*. https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2021.739022