The International Journal for
Translation & Interpreting
Research
trans-int.org

A task segment framework to study keylogged translation processes

Ricardo Muñoz Martín
Università di Bologna, Italy
ricardo.munoz@unibo.it

Matthias Apfelthaler
Universität Graz, Austria
matthias.apfelthaler@edu.uni-graz.at

DOI: 10.12807/ti.114202.2022.a02

Abstract: The Task Segment Framework (TSF) is a tool to analyze full typing flows of translation tasks as keylogged with timestamps recorded for keydown, keyup, mouse clicks and moves, and actions performed in other applications. The TSF assumes that intentional pauses flag stretches where subjects concentrate on unrecorded cognitive processes such as planning and assessment. The interspersed typing stretches are task segments, with or without text, where basic subtasks may be observed, mainly adding new text, changing existing text, and searching for information. Accumulated experience and planning allow translators to lump strategically similar activities together, in order to spare efforts and task switching costs while maximizing efficiency. Hence, task segments may contain activities of just one such subtask or many. Translation fluency is a key notion of the TSF, operationalized through many indicators such as typing speed, prior pause length, TS (task segment) length in events, text length as full words, number of typos and respites (=mid inter-keystroke intervals), subtask(s), and the like. The approach seems particularly sensitive to translation expertise levels and may be applied with variations to other multilectal mediated communication tasks. This article lays down the conceptual basis of the TSF and summarizes its basic notions and constructs.

Keywords: Translating; keylogging; task segment; subtask; fluency.

1. Introduction

This paper presents a framework to study full translation processes as recorded with keyloggers that timestamp keydown and keyup actions for single keystrokes. It is called the *Task Segment Framework* because its basic trait is that it uses time lapses to break down the flow of keylogged data into segments with registered behavior—namely, *task segments*—but not necessarily text. The Task Segment Framework (henceforth, TSF) draws from Cognitive Translatology (Muñoz 2010; Muñoz & González, 2021), where translating is seen as a form of restricted production (Risku, Milošević & Pein-Weber 2016, p. 64). As an interpersonal communicative activity, it basically entails constructing a text expected to be thought of as having a relationship of (mainly, meaning) correspondence with another text or texts.

Translating is often undertaken without having read the source text to completion. The whole performance is driven by top-down and bottom-up mental processes, where successive readings of text stretches foster the emergence of formulations in the target language that will then be assessed as candidate translations. From this perspective, translating amounts to successfully managing and steering one's own mental resources to meet expectations. Good results lead to entrenching successful working routines, including self-conventionalized solutions (i.e., *default translations,* Halverson, 2019). This is what learning how to translate and acquiring expertise amounts to: efficiently adapting behavior to yield socially acceptable products for multilectal mediated communication.

Professional translating often demands high quality, but also enhanced efficiency and productivity, to make ends meet. In many developed countries translators often use workstations. Such working environments demand quite complex behaviors and also come with special demands, regarding the way work routines are carried out. For instance, professionals tend not to use print information sources any longer. Thus, on top of performing and managing several situated, complex, parallel, nested and recursive mental operations, translators need to master adaptive behaviors to perform smoothly and efficiently in digital environments. Word processors are still the most usual digital environment for translators to carry out their tasks. Different tools, such as translation memories and post-editing tools, will prompt differences in behavior that cannot be the subject of this paper, although future work will strive to adapt the TSF to account for such differences. What follows illustrates the application of the TSF to translating with a word processor and an Internet connection.

Writing text with a pen is much slower than typing, but typing still is the slowest aspect of text production in a word processor. Typing speed rhythms and variations seem indeed excellent indicators in our research. Fine-grained keylogging analyses provide an unobtrusive, detailed and time-stamped record of the activities performed in several applications with a standard keyboard with at least 84 keys, and a standard mouse, with at least two buttons and a scrolling wheel. Besides, in spite of all the hype regarding MT (Läubli, Sennrich & Volk, 2018; Toral et al., 2018), postediting (e.g., Vieira, 2017), and dictated translation (cf. Gouadec 2007, p. 375 vs Ciobanu, 2016), keyboards will be around for a while, simply because not all markets, languages, and goals are identical, and the silent environments needed to use one's voice as input are rather rare. Even in high-tech professional environments working on major world languages, nobody would ever dream of removing the keyboard and only use alternative input methods now and in the near future.

Technological hype may have relegated keylogging to a secondary position in multi-method research projects where eyetracking is often center stage. In our view, only when a full, independent analysis of the process is developed for keylogging will it be possible to interface it correctly with eyetracking and other data collection tools for translating and related tasks. The advantages of eyetracking are obvious, but our research program consciously focuses on keylogging because it provides much more, and much more precise information, and also because data collection is ecologically valid, perhaps the best trade-off between unobtrusiveness and control we have now. Research keyloggers—such as Inputlog (Leijten & Van Waes, 2013)—are often free, so keylogging does not work as a financial gatekeeper to access research. Furthermore, initiatives such as the TPR-DB foster access and reuse of keylogged data and, in consequence, transparency (e.g., Carl, 2012).[1]

2. Keylogging

Keyloggers register what people type and do with their mouse, and when they do it. This takes the shape of a string (actually, a database) of codes and timestamps, which depict typing (§ 1.1) and pausing (§ 1.2). The codes can often be linked in behavioral sequences, many operating on language strings, so that the keylogged flow may be broken down into *task segments* (§1.3).

2.1 *Typing*

Priming and EEG evidence suggest that typing relies on stored mental structures controlling the timing and coordination of muscular activities, i.e., on *motor programs* (Crump & Logan, 2010; Logan, Miller & Strayer, 2011). Motor programs activate, process and sequence physical movements to execute several keystrokes through graded activation—in parallel and beforehand (Behmer & Crump, 2017). Motor programs are taken to operate only after typists have settled on the text stretch they want to enter. Thus, language processing is assumed to happen first in an "outer loop", which focuses on the encoding of symbol sequences and on monitoring them once they are typed. The outer loop is independent of an "inner loop" which produces keystroke sequences (Logan & Crump, 2011). In other words, typists settle on a certain language formulation; then, it is transformed into minimal sequences of (keyboard) symbols; and then, the sequences of (keyboard) symbols become sequences of letter targeting (inner loop): finger choice, movement direction, keypress, and haptic feedback (cf. Yamaguchi & Logan, 2013). These activities are easier to understand sequentially, but different sequencing processes may happen at once that make different phases for different string processes coincide in time.

Logan & Crump (2011) further assume that words lie at the interface between the outer and inner loops. Words might thus be the rail switchyard of two kinds of operations: (analytical) text chunking—breaking down text into manageable stretches—and (synthetic) motor chunking, adding up movement routines to yield reproducible behavioral sequences.[2] This, however, does not imply that language features do not have an influence on typing. Immonen (2006, pp. 327–328) finds that pause lengths increase with the size of the linguistic unit, both in translating and monolingual drafting. Pinet, Ziegler & Alario (2016) find that both highly frequent and sound-spelling consistent words lead to shorter keypress reaction times and higher typing accuracy.

---

[1] Bogusława Whyatt (personal communication) notes that keylogging data are easy to store (light), compared to 'heavy' or bulky eye-tracking or screen-capture data. Keylogging makes it easier to link to one's raw data in a repository and thus fosters openness and accountability.

[2] Cowan's (2005) theory of focused attention suggests that people chunk information at different levels and that chunking is organized hierarchically.

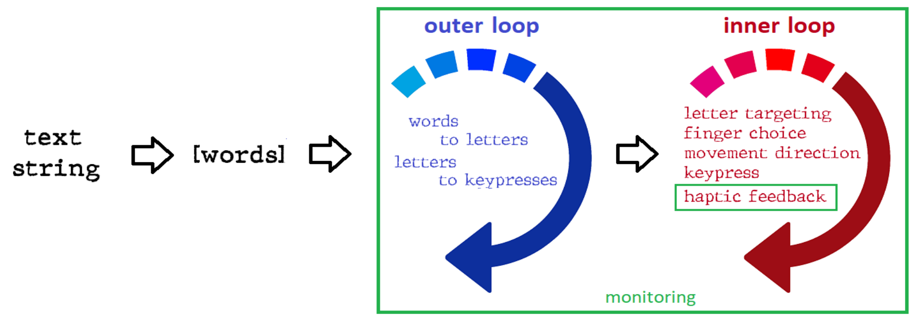

Figure 1. Outer and inner loops as phases in the transduction of text strings into action, after Logan & Crump (2011)

Words, however, might not always be the encapsulated chunking units that drive motor programming in typing: (a) strong language effects may be found within long words, such as between prefixes, syllables and other sub-lexical units (Scaltritti, Alario & Longcamp, 2018). *Inter-keystroke intervals* (IKIs) are also longer at syllable boundaries, and shorter within highly frequent bigrams (Pinet, Ziegler & Alario, 2016). And (b), some effects are above word level. Overlearned typing sequences of frequent bigrams, trigrams, but also common words and multiword expressions are associated with higher typing speeds. Immonen & Mäkisalo (2010, p. 60) find that pauses before noun phrases tend to be longer than those before verb phrases. In contrast, there are shorter pauses prior to noun phrases longer than four words, suggesting that very long noun phrases are not processed as a single chunk (Immonen & Mäkisalo, 2010, p. 55, note 13). In brief, language features seem to have an influence on typing, but typing motor strategies may disregard the integrity of language units (e.g., phrases, words) once they are being processed in the outer and inner loops.

Translators tend to mentally chunk their source text into small portions whose translations they often type uninterruptedly, dependent as they are on one and the same single sequence, one or several ad hoc motor programs devised and linked for the occasion right before action starts. When compared to monolingual drafting, translating has longer mean-pause lengths between smaller language units (letters, syllables, prefixes+roots, phrases, subordinate clauses) but shorter pauses between longer units—before main clauses and above (Immonen 2006, Immonen & Mäkisalo, 2010). Immonen (2006, p. 332) suggests that:

> The fact that a translation is not planned from scratch may shorten these higher level pauses. On the other hand, pauses at lower levels may be lengthened because it is at word and clause level that decisions concerning word choices and grammatical options are made.

Immonen & Mäkisalo (2010, p. 57) also suggest that translating involves *less* planning at sentence and paragraph levels. So, we assume that translators tend to work at word, phrase, and clause levels, depending on the specific local demands (cf. Dragsted, 2005). Strömqvist (1999) notes that IKIs within words are ubiquitous and very short, and that planning and monitoring only have a marginal impact on them. He further suggests that the median *within-word* IKI may be the most reliable indicator of typing skill. Before we consider the impact of translating on typing, we need to have some notion of typists' normal behavior.

Table 1. Indicators of typist profiles, from Dhakal et al. (2018)

		number of fingers	words per minute	keypress (ms)	interkey intervals (ms)	rollover ratio
all	**mean**	6.95	51.56	116.25	238.66	25
	s.d.	*2.95*	*20.2*	*23.88*	*11.6*	*17*
fast	**mean**	8.4	89.56	104.49	121.7	49.9
	s.d.	*2.2*	*9.53*	*17.38*	*11.96*	*14*
slow	**mean**	5.3	20.91	128.99	481.03	7.6
	s.d.	*3.2*	*4.05*	*28.85*	*123.36*	*6.4*
trained	**mean**	8	54.35	118.39	223.55	29
	s.d.	*2.46*	*20.8*	*23.81*	*107.78*	*17.7*
untrained	**mean**	6.5	49	115.29	245.34	24
	s.d.	*3*	*19.73*	*23.85*	*112.6*	*16.7*

Based on data provided by 168,000 volunteers transcribing English sentences online with modern computer keyboards, Dhakal et al. (2018) offer average values for all typists, as well as for fast vs slow, and trained vs untrained typists (Table 1). Fast typists are those with speeds higher than 90% of participants (>78 WPM); slow typists belong to the slowest 10% (<26 WPM). (Touch) typing training and the number of fingers used were self-reported. *Rollover* refers to starting to press one key before having released the previous one in full. An average of one out of every four keystrokes (25%) displays rollover effects, and in fast typists, half of their keystrokes (49.9%) overlap. The mean overlap span between consecutive keypresses is 30 ms, but sometimes it may reach 100 ms. This is close to a (fast) full keypress interval, and here it is taken to be a powerful indicator of motor programs being at work. Table 1 shows that differences in keypress length between fast and slow typists are minimal, while those between their IKIs are very important. This supports Strömqvist's (1999) choice of median within-word IKIs as typing skill indicator.

Conijn, Roeser & van Zaanen (2019) find that (a) overall IKI means are stable across tasks (copying, email writing, academic summarizing); (b) intervals between words and other (sub)sentence units only differ between some tasks; and (c) overall features other than IKI, such as the number of words, revisions, and total time, differ across tasks. Thus, IKI values for smaller units may safely be assumed to be similar enough in regular typing and in typewritten multilectal mediated communication tasks.[3] Dakhal et al.'s (2018) large number of informants, and the hypothesized relative independence of the workings of the inner loop—transforming planned-keystroke series into finger-movement sequences—supports, at least for now, that the part related strictly to typing in Dakhal et al.'s copying tasks and in translation tasks are the same. Hence, differences in these values while typing will be taken as indicative of influences potentially related to the task at hand, e.g., translating.

Dragsted (2005, p. 66) describes an analytic mode of translating, typically associated with translation trainees, and an integrated processing mode characteristic of experienced professionals:

[3] See also Dam-Jensen & Heine (2013) and Risku, Milošević & Pein-Weber (2016).

> The features characterising the analytic mode are: short average segment size, low production speed and long pauses, processing at word/phrase level, many single-word segments, and few exceptionally long segments, whereas the features characterising the integrated processing mode are: long average segment size, high production speed and short pauses, processing at clause/sentence level, few single-word segments, and many exceptionally long segments.

Fluent typing is customarily assumed to correlate positively with text quality, because motor programs free cognitive resources that typists can assign to other processes (Kellogg, 1999). Touch typing frees translators from looking at the keyboard (Johansson et al., 2010), thereby fostering an integrated processing mode where reading and typing may overlap. Shaffer (1975) suggests that a very good touch typist could copy texts while repeating another text they listened to verbatim, showing little dual-task interference. Touch typing results in fewer typos and dramatically higher production rates both in text drafting and translating.

In contrast, *hunt-and-peck* (two-finger) typists need to look for every key by sight. They make larger, less strategic hand movements, and they need to spot reading and insertion points back in the source and target texts once and again—"[...] if attention has to be divided between the execution of untrained motor programs and the maintenance of memory traces, trade-offs are likely to occur and performance deteriorates" (Alves et al., 2007, p. 56). We might thus be tempted to associate touch typing with expert translating and hunt-and-peck typing with lay translating—but that would be a mistake. Most typists seem to be somewhere on a continuum between two poles (Johansson et al., 2010, p. 836).

Experienced translators who are not touch typists usually know their keyboard layouts as well as touch typists do. They may use fewer fingers but do so quite consistently, because motor programs develop independently of the ways people type. Thus, experienced translators, whether touch typing or not, often type without looking at the keyboard (or looking less) and can often coordinate reading and typing. On average, trained typists in Dhakal et al. (2018) type a modest 5 WPM faster than untrained typists (Table 1); the difference between their mean keypress lengths is just ~3 ms, and between their mean IKIs, only 22 ms. These narrow gaps point to hybrid ways of skilled typing other than touch typing, as hinted at by their also small differences in rollover ratios (29% for trained, and 25% for untrained typists) and the higher number of fingers untrained typists use, compared to slow typists. Thus, typing skill might better be readdressed as *keyboard efficiency* (Aldrige & Fontaine, 2019, p. 286).

Dragsted's (2005) professionals tended to type longer stretches faster and with fewer pauses, but, crucially, they would fall back into novices' serial strategies with more difficult texts. Still, "the translation of the difficult text did not always involve a shift into the more analytic processing mode in the professional group because of domain expertise" (p. 67). According to Alves et al. (2007, p. 58), and contrary to conventional wisdom in our realm:

> It is probably in execution periods that storage and processing demands are higher, and where Flower and Hayes's portrayal of a writer as "a thinker on a full-time cognitive overload" (1980b, p. 33) is most accurate. While typing, writers must literally keep in mind the representation of what they intend to write, pay attention to the output being produced, maybe plan further segments or revise the already written ones, or even pay attention to finding the keys on the keyboard.

That is, rather than considering text-stretch length and typing speed fixed characteristics of translators that depend only on their level of expertise and their working memory capacities, we might want to link them to the dynamic, often strategic interaction with the source text and other elements of the environment. Mental fatigue, for instance, has an effect on top-down cognitive control and on typing speed (de Jong et al., 2018).

Zhang et al. (2019) analyze students' essay writing processes and find that the group with higher scores seemed *less* fluent than the lower-scoring group. They suggest that this might be due to higher task engagement and writing efforts, which might play an important role in generating better quality text. Indeed, typists pause to plan and revise, so that lack of pauses is associated with increases in error rates (cf. Harris & Coltheart, 1986, p. 214). Medimorec & Risko (2016) and Medimorec, Young & Risko (2017) conclude that, for some typists, decreasing typing speed may yield positive effects on cognitive processing and text quality.

Since (a) professionals need to make a living from translating; (b) touch typing is known to lead to increased productivity and accuracy; and (c) typing skills do not seem particularly hard nor do they take long to master (estimates range from 20 hours to three months), sticking to hybrid typing styles might be a matter of choice, i.e., some people might find a hidden advantage in clumsier typing. Typing interruptions might be used strategically (if perhaps unconsciously), in a sort of tradeoff to manage mental effort by spreading it over time, and also to refresh contents in working memory. In the best (fluent) cases, the motor program translators devised has been carried out to completion after a while, and they proceed to face the next excerpt. Thus, they cease typing and often start (re)reading what they just typed or further ahead in the source text. In other cases, however, translators stop typing more or less abruptly because they changed their minds, want to fix a typo (that they may detect immediately or later on, through their parafoveal vision), note a mistake, or they are simply drawn to focus their attention elsewhere. All these options will get keylogged as pauses, usually longer than those derived from regular typing activities (cf. Crump & Logan, 2013). It is thus time for us to turn to pauses.

2.2. *Pausing*

We stated above that keylogging yields a behavioral view of translating as an alternating progression of pausing and typing. An *event* is any recorded, timestamped minimal action on an input device, such as a letter keystroke, a *dead key* (i.e., unprinted keystrokes necessary to produce accents, umlauts, etc.), a keypress of a modifier (e.g., Ctrl, Alt, Shift), and a mouse scroll. For each keyboard event, keyloggers such as Inputlog will calculate *keypress* or *action time* as the time span between pressing and releasing one and the same key, and *pause time* as the time span between releasing one key and pressing the next one. There is necessarily some time lag; otherwise, the keypresses will be considered simultaneous. Popular operating systems let users adjust a *repeat delay* threshold for their keyboards—the timespan until the computer reacts as if you had pressed a key twice, while you are actually holding it down—with typical ranges between 32 and 400 ms, and the default value often set at 200 ms. Yet, Chukharev-Hudilainen (2014, p. 80) argues that "[...] we can make a cautious claim that the duration of a pause could predict the type of mental processing that may be going on while the execution of typing is suspended". Hence, we need to stipulate the length for an IKI to be considered a pause.

The word *pause* often hints at intentional halts, whether planned or not, as is the case of *long(er)* IKIs, the main concern thus far in the translation literature. *Pauses* are here defined as voluntary interruptions of the typing flow—interspersed empty timespans where translators often perform mentally demanding activities, such as evaluating, planning, and reading (Olive & Kellogg, 2002).[4] Pauses often point to the end of the motor program's execution for the current typewritten stretch—"when writers run out of content" (Galbraith & Baaijen, 2019, p. 311)—or the sudden reallocation of attentional resources, with or without a typing breakdown. Thus, pauses may technically be considered IKIs, but conceptually they are not, for they are not what happens between keypresses, but rather process periods in their own right. Let us, for now, note that pauses are not considered part of the typing flow, because translators are not even intending to type.

Restricting the use of the term *pause* to intentional time spans demands that they be quantified. Pause thresholds vary a lot in translation and in writing process research. Kaufer, Hayes & Flower (1986) argue that 2,000 ms IKIs flag "conceptual planning" episodes—when writers run out of content and need to create new one. Newell (1990, p. 129) argues that a person can provide an unplanned reaction to some environmental stimulus within ~1,000 ms, which he considers a relatively long pause. Limpo & Alves (2017, p. 308) suggest that IKIs between 30 ms and 2,000 ms reflect typing (transcription) processes, and pauses above 2,000 ms, planning, revising and the like. Chukharev-Hudilainen (2014, p. 80) argues that IKIs above 500 ms point to linguistic hesitations and those above 1,200 ms, to planning beyond the current unit. Thresholds are admittedly arbitrary and using one threshold value across samples will probably not be equally right for all informants, and even for the same informant at different times. In brief, there is no agreement about where to set cutoff points and an overwhelming proportion of IKIs is below 1,000 ms. Van Waes et al. (2016) showed that 75% of all IKIs were below 250 ms, out of which 77.1% were found within words and only 20.8% between them.

Pauses are, then, far from exhausting the occurrences of IKIs. Mentally demanding operations are also present as translators type, but they are not always so attentionally engaging as to lead typists to stop. Instead, typists try and often succeed to go on with the execution of their planned sequence. They may pay a toll for it, in the form of IKIs longer than average, but shorter than pauses—and possibly with more frequent typos as well. Lacruz & Shreve (2014) found that clusters of IKIs as short as 500 ms are good indicators of higher cognitive efforts in translators. Newell (1990, p. 129) finds that, when informants need no planning, their reaction can be as short as ~500 ms. Chukharev-Hudilainen (2014, p. 80) argues that IKIs around 500 ms are rather frequent, and that they are not always consciously detected by typists.

Thus, we may want to distinguish between pauses and *respites,* defined as typing disfluencies that (a) are not intentional; (b) do not stop the text

[4] Hansen (2002, p. 33) classifies pauses into *orientation* pauses, *control* pauses, *internal* pauses and *monitoring* pauses. It is doubtful, though, that pauses will be devoted to a single process. Changes in the text right after a pause may be agreed to hint at evaluation having occurred in that prior pause, but it is unlikely that they are the only cognitive process at work. For instance, before resuming keyboard activity, some planning needs to take place, so a pause for monitoring or evaluating may host planning as well. A more detailed level of granularity (e.g., with eyetracking) might reveal activities such as (gaze regressions for) skimming and (visual searches for) spotting the right place.

production flow; (c) may become conscious; and (d) may be associated with task-related aspects of cognitive performance. Many respites are linked to processing difficulties; for instance, they are often found between the last keystrokes before a pause, especially when a motor program was not fully executed, resulting in broken words, incomplete phrases, etc. We might further assume that respites do not usually involve planning, but we prefer to leave that as an open, empirical question. Baaijen, Galbraith & Glopper (2012, pp. 16–17) find, however, that the distribution curve of IKI length is heavily skewed—reflecting a mixture of components—and find three peaks, which roughly correspond to Muñoz & Martín's (2018) long, mid and short IKIs.[5]

The first peak in Baaijen, Galbraith & Glopper (2012) comprises 65% of IKIs, with a mean length of 330 ms, which they link to word retrieval processes. The second peak gathers 26% of the IKIs, lasting an average of 735 ms, and hints at phrase boundary processes; these might be roughly our respites, although respites crucially do not happen only at the boundaries of linguistic units. The third peak is more diffuse and fuzzier than the other two. It covers about 9% of the IKIs; their mean length is 2697 ms and the authors associate them with message planning or reflection. This would correspond, roughly, to our pauses (which are shorter in translation). Again, pauses do not necessarily fall between self-contained linguistic units.

Baaijen, Galbraith & Glopper's (2012) analysis reveals that the model with three distributions is the best fit for 58% of their participants, with 23% participants indeterminate between two or three distributions and the remaining 19% being fit best by a two-distribution model (p. 264). In brief, we have pauses, respites and very often a third cluster of numerous, shorter IKIs that we will call *delays*. Typing delays are shorter than respites, very often below 500 ms; they tend to be (a) within words and (b) both unintentional and unconscious. In our view, they may be caused by lexical and mechanical processes, but there may be other sources for these smaller disfluencies. Given the high intra- and inter-subject variability, suggesting a limit between task-related, "cognitive" IKIs and typing-related, mechanical IKIs is very risky. We know, however, that customarily "activities below 30 ms are all attributed to so-called slips" (Leijten, Van Horenbeeck & Van Waes 2019, p. 72). So, we may want to set a minimal length for IKIs to be considered relevant.

There are at least two powerful reasons to establish a minimal length for IKIs: accuracy and relevance. According to Pinet et al. (2017, p. 1164), IKIs are typically much shorter (down to tens of milliseconds) than the mean reaction times measured in psychology, and thus inaccurate time measurements may induce more important distortions when interpreting IKIs. They explain that typical keyboards use USB ports that are sampled every 8 ms (a rate of 125 Hz), so that small differences in milliseconds near that range might be distorted. Furthermore, keyboard settings may modify the ways operating systems interpret sequenced or combined keystrokes, such as multiple keypresses interpreted as one, combinations of keypresses triggering a particular event, etc.

[5] Incidentally, Campione & Véronis (2002) distinguish between brief (< 200 ms), medium (200-1000 ms) and long (> 1000 ms) intervals in spoken language, with peaks at 150, 500, and 1500 ms respectively. We assume that spoken language is faster and its intervals, in general, shorter than those in written language. Campione & Véronis (2002) add that reading aloud brings about fewer long pauses compared to spontaneous speech, due to less planning, and also that these values are remarkably stable across languages.

Reimers & Stewart (2015, p. 326) find that web-based measurements may overestimate reaction times by 30–100 ms. This would seriously affect online keylogging measurements from tools such as TransCenter, PET and CASMACAT (descriptions in Vieira, 2013) and also any research project that combines these online measurements with stand-alone keyloggers. In brief, with today's equipment, we might not be safe with accuracy demands below 100 ms.

As for relevance, Van Waes et al. (2016) describe keyboard text production as characterized by within-word pauses related to mechanical transitions between keys (p. 411) and argue that "[t]he main reason for defining a pause threshold is to avoid 'mechanical noise'" (p. 414). Song, Wagner & Tian (2001) tested informants typing digraphs and found variations of *at least* 200 ms in absolute terms due to mechanical matters. Typing one lowercase letter followed by typing one number with the other hand is the fastest combination (80% of trials under 100 ms), whereas consecutively typing two lowercase letters with the same finger or a letter and a number with the same hand are the slowest (Table 2). In these controlled, repetitive runs, planning was minimally involved. If we were to tack on the 500 ms Newell (1990) added for planning, mechanical IKIs might, in exceptional cases, reach 700-800 ms.

Table 2. Percentages of IKIs of different lengths, depending on digraph combinations, according to Song, Wagner & Tian (2001)

Digraph combo	1 letter +	number		letter		
	hand	same	other	same		other
	finger	–	–	same	other	–
Time span (ms)	<100		80		67	50
	100–150	10	20		27	50
	150–200	15		50	6	
	200–250	35		45		
	250–300	35		5		
	>300	5				

We argued that we need to distinguish pauses from respites and delays. In order to do so, we will need to establish a minimal cutoff point, a baseline for an IKI to qualify as a delay, and then two thresholds to separate delays from respites, and respites from pauses. We have also stated that pauses are the longest typing halts, but we may need to qualify that. In spoken language research, Mead (2000, p. 92) distinguishes between silent and filled pauses, the latter simply described as hesitation sounds such as *eh* and *um.* Muñoz & Martín (2018, p. 37) observed that in many action spans between two pauses no text was added, deleted or moved around. The keylogger only registered apparently ineffective clicks and cursor moves about either the source or the target texts.

Muñoz & Cardona (2019, p. 529) argue that such typing periods may host reflex or intentional movements of the mouse aimed at, e.g., removing the cursor from the screen for reading, or adjusting text height and also may "include performing repetitive actions with no consequences so as to maintain physical readiness, and strategic/preparatory activation and control". Thus, these "useless" keylogged, minimal activities might rather be considered accidental—or intentional but unrelated to text-production—fillers *within* a longer pause. Thus, the longest break between consecutive typing periods might not be a pause, but rather the addition of consecutive pauses plus the keypress durations of the interspersed filler events. Up to now we have considered typing only as the physical activity of entering text in a computer. We now turn to the goals and nature of action sequences in the task of translating.

2.3. *Task, not text*

Models in writing process research seem to miss important components, such as information search and attention management (Leijten et al., 2014). One way to try to avoid overinterpretation is building successive layers of behavioral description, starting from single events up to the whole keylogged contents between two pauses. Typing events combine into sequences, linked by common goals—often within a single motor program or a transitory combination of motor programs with various degrees of entrenchment. Typing one word is the obvious example, but let us decompose an action to better illustrate the minimal interpretation involved in identifying actions in the keylogged stream. In Figure 2, a translator searches for the expression *Trail of Tears* in a web browser while translating. One of the possible ways to carry out this action entails the following keylogged steps:

1 moving the cursor to the taskbar
2 clicking on the web search engine icon
3 placing the cursor in the web search engine window
4 clicking on the web search engine window
5 typing *Trail of Tears*—i.e., typing
[SHIFT]+t·r·a·i·l·[SPACEBAR]·o·f·[SPACEBAR]·[SHIFT]+t·e·a·r·s·[SPACEBAR]
6 pressing [ENTER]
7 [in the simplest case, a time span, probably a respite or a pause, where the subject may be reading]
8 moving the cursor to the taskbar
9 clicking on the target text icon

Figure 2. Sequence of recorded steps and events constituting the search *Trail of Tears* in a web search engine while typing

Some actions entail other unregistered actions, such as looking for the mouse on the table, seeking the target on the screen, and reading (step 7). Furthermore, in Figure 2 all typing events adding text have been lumped into just one step (5), so as to declutter the illustration. Important for us here is that we may assume that [*shift*]+*t·r·a·i·l*+[*spacebar*] and [*shift*]+*t·e·a·r·s*+[*spacebar*] are regular sequences with specific motor programs. Crucially, sequences 1–4 and 6–9 may often be found in searches. Furthermore, *o·f*·[*spacebar*] can be thought of as a potentially overlearned (i.e., totally routinized) sequence that will show up in many different configurations (e.g., *House of Cards, thereof*). These building blocks seem to have a recursive nature: For instance, the example shows that minimal searches may be composed of steps 3–6, but only step 5 would be really different in each case. Steps 1–2 are necessary here but will also be present if we want to revisit a webpage we kept open in the background. They may be seen as a focus-changing routine—a sequence to activate and access a different application/window—that may display variations, such as the one represented by steps 8+9, which we might consider a "search-exit" routine.

Many building blocks with no text (with only unprinted events) might have their own motor programs too, like a sequence of steps for the word-processing command *print*. Whether they have them or not, the overall action of searching for *Trail of Tears* would not be possible without carrying out the whole sequence (or an alternative one) to completion. Thus, typing this text stretch is nested within a more complex behavioral compound, and tasks and goals are understood as recursively nested categories.[6] The nesting behavioral sequences may shed light on what was going on in the minds of translators as they were entering their text stretches. All these building blocks, whether with text or not, are equally important in a record which is not about language, but about communication. It is about how somebody uses a workstation to build a text, mainly, but not only, with language—images and other means may also come into play. In other words, it is the recorded behavior to carry out the complex task of building a sophisticated communicative artifact. The basic unit of the translation process is a *task segment,* not a language string.

Building blocks in a translation keylogged flow may relate to one of these basic behaviors: (a) adding to the *text written so far* (TWSF); (b) changing the TWSF; (c) searching for information; or (d) interacting with the computer for purposes other than those in *a, b* and *c.* We will label these basic behaviors ADD, CHANGE, SEARCH and HCI, respectively. ADD also includes inserting elements other than text, such as images, diagrams, and tables; additions happen at the last point of new insertions—in western languages, usually the rightmost point in the text (within the last or current task segment). CHANGE includes all modifications done in the existing copy, such as deleting elements, adding new ones and reorganizing their order, before the last insertion point. SEARCH comprises the set of behaviors necessary for searching, finding, saving, copying, pasting information on the Internet, plus navigating through websites and pages.[7] ADD, CHANGE, and SEARCH may be considered subtasks within a translation task.

HUMAN-COMPUTER INTERACTION is a hodgepodge of all those behaviors not fitting into the previous categories. HCI behaviors may be task-internal, such as adjusting computer and application parameters; neutral, such as entering fillers; or task-external, such as checking email or managing the music a translator is listening to while at task. We chose *internal* vs *external* to avoid labels that would distinguish between “task-related” and “task-unrelated” behaviors. Even so, readers are advised to take “internal” and “external” with a generous dose of skepticism, since many apparently unrelated behaviors may be motivated by task goals as well. For instance, lowering the music volume may be used to better concentrate and solve a problem, and reading email or instant messages often may be used to cope with a boring task (Risku, Milošević & Pein-Weber 2016). In principle, any keylogged behavior is part of a

[6] Here *tasks* are “undertakings that require a person to mentally process new information (i.e., acquire and organize knowledge/learn) and allow them to recall, retrieve that information from memory and to use that information at a later time in the same or similar situation (i.e., transfer)” (Kester & Kirschner, 2012, p. 619). In general, however, we use *task* to mean an activity or course of action commissioned by somebody in authority, often expected to be accomplished within a period of time or by a deadline, usually as part of a larger project. Thus, *task* refers to translating, revising, postediting and the like. We will keep both usages, in the hope that the readers will be able to tell the difference in each case.

[7] There are actually circumstances where text is entered for a purpose other than searching for information, such as writing down a note for future reference. These other uses are marginal and we will need a larger pool of data to be able to suggest slicing away parts of the SEARCH task segments.

multitasked complex that includes translating and hence may have an impact on it. HCI is a subtask, but it is somewhat different in that it comprises several heterogeneous *behavioral sets* or *repertoires*.

A *behavioral set* (Holaday, 2015, p. 95) is "[...] a relatively stable and habitual behavioral pattern of responses to particular drives or stimuli. It is learned behavior and is influenced by knowledge, attitudes, and beliefs". This notion is close to *skillset,* "a collection of skills and abilities that can be applied to a professional or creative endeavor" *(Webster's Dictionary).* It is also close to *(behavioral) repertoire,* "[...] the sum total of potential behavior or responses" that a person can perform, and "usually refers to behavior that has been learned and is generally quantified through the study of past behavior" (VandenBos, 2015, p. 906). We do not intend to introduce these notions as part of the motivation of the TSF but use them to suggest the apparent psychological reality of TSF translation subtasks as they materialize in task segments, in terms of stored and efficient potential behavior (see Muñoz & Apfelthaler, 2021).

We claim that each of the three subtasks above is linked to different behavioral repertoires comprising reading, text production, etc. There might be a tendency for certain kinds of reading (Urquhart & Weir, 1998, p. 106–108) to happen more in some of the subtasks: ADD calls for *skimming* (fast, gist reading) in first-pass-reading—cf. Shreve et al.'s (1993) *reading in anticipation of translating*—and for *careful reading* (non-selective reading, trying to build a macrostructure) in second-pass reading.[8] CHANGE demands *scanning,* or reading selectively to locate a particular piece of information; SEARCH, usually online, requires *skimming* and *search reading* through large amounts of material and immediately evaluating its quality and reliability.

Reading is not keylogged, so it is technically not part of the TSF, but it is a component of writing and cannot be ignored, even if not measured but indirectly so. Translators' behavior seems to be different when reading for translating (Shreve et al., 1993; Castro, 2008; Jakobsen & Jensen, 2008). Dragsted & Hansen (2008, p. 4) argue that, when translating, comprehension and production cannot be easily separated into two distinct activities because translators read with the goal of producing an output in another language. Indeed, Whyatt (personal communication) notes that reading interacts with all identified keylogging subtasks—ADD and SEARCH often come after ST reading and CHANGE, after having read the TT.

Text production when translating also appears to work differently in these subtasks: In ADD task segments, we may find somewhat regular typing. Associated cognitive processes may include less planning than and the same monitoring as in monolingual text production, and also furthering the macrostructure translators build on the fly as *tertium comparationis* between the source and their target texts. CHANGE task segments often deal with shorter language stretches, down to letters, diacritics, and punctuation marks, which translators need to fit into the new copy usually trying not to create a domino effect that will necessitate further changes. Interestingly, corrections in the ADD task segment in progress seem to be often done with the backspace key, whereas

[8] Jakobsen & Jensen (2008, p. 16) suggest "that a fair amount of pre-translation probably enters into the reading of a text as soon as it is taken to be the source text for translation". It is our contention that translators let meaning spontaneously emerge in successive reading passes or rounds that they use to activate, stimulate, and refresh potential contents and restrictions. This will however not be further pursued here since it is clearly off topic.

deletions in CHANGE task segments may often consist of highlighting the excerpt and overtyping on it. These behavioral differences are also noticeable when entering text in places other than the target text. SEARCH task segments, for instance, often contain only lowercase full words, no diacritics, no flexed forms, and keywords in other languages as part of search strategies that hint at some lack of inhibition or at code-switching (e.g., *Trail of Tears español; Trail of Tears significado,* etc.).

These behavioral differences in reading, text production and other task features suggest alternative palettes of behaviors and rules that might be costly to simultaneously maintain throughout the task (for they often include mutually exclusive rules, such as different spelling conventions). Instead, translators spontaneously learn to improve their efficiency by trying to keep using the same *behavioral repertoire* throughout full task segments, i.e., they tend to stay within one single subtask for a whole task segment (or, alternatively, to pause and start a new task segment when they need to switch). This leads to ADD, CHANGE and SEARCH task segments, as opposed to MIX task segments (the latter combining at least two *behavioral repertoires*)—that is, they try to coherently steer their behavior within each task segment and in as many task segments as possible by separately focusing on *either* adding or changing text, searching for information, or otherwise interacting with their workstation. This leads us finally to sketch the basics of a new analytical approach to translation processes: Let us flesh out the Task Segment Framework.

3. The Task Segment Framework

As of today, the TSF is only a structured set of assumptions and constructs to study typing processes. It has been built to study mainly translating, but we aim to adjust it to other multilectal mediated communication tasks, and to writing and other typing tasks (e.g., coding). The TSF arbitrarily sets a minimal baseline for IKIs and two thresholds. Altogether, they separate IKIs into four kinds: *pauses, respites, delays* and *lags.* Task-segment categories are extended with two more categories, FILLER and MIX. These two new kinds of task segments do not necessarily contain work on the target text. They are neutral or open in this respect. *Fluency* was indirectly mentioned as a characteristic of keystroke sequences typed without interruptions. In the framework, it is center stage also through considering typos and segment-internal respites—or, rather, lack thereof—within task segments.

3.1. *A minimal baseline*

We will start by choosing a baseline, or cutoff point, to distinguish *lags* from all other kinds of IKIs (delays, respites and pauses). That baseline is set at 200 ms, for the following reasons:

1. Madl, Baars & Franklin (2011) suggest that human cognition consists of cascading cycles of recurring brain events. A cognitive cycle consists of (a) sensing the current situation (80–100 ms from stimulus onset under optimal conditions); (b) a conscious episode of interpreting the stimulus with reference to ongoing goals (200–280 ms after stimulus onset), and (c) selecting an action in response (planning, 60–110 ms from the start of the conscious phase). Translators would be conscious of a new stimulus or constellation of stimuli only after 200 ms.

2. Pulvermüller (2001) suggests that lexical and semantic brain processes occur near-simultaneously, but that lexico-semantic activation has two distinct steps: access or *ignition* (word recognition; latency 100–250 ms), followed by active memory for words or *reverberation,* >250 ms. Differences between semantic word categories can appear early in the neurophysiological brain reaction (100–200 ms after stimulus onset). Translators would then *fully* activate word meanings only after 200 ms. See also Hauk et al. (2012).
3. The level at which minimal cognitive, perceptual, and action operations interact is suggested to be around 300 ms (Ballard et al., 1997; Gray & Fu, 2004). So, a 300 ms baseline can capture embodiment. Setting the baseline at 200 ms ensures that we will not miss differences among subjects and that measured timespans are not unduly influenced by measurement error.
4. Disfluencies in neurotypical adult English speakers have latencies higher than 200 ms (Goldman-Eisler, 1968, p. 42). A baseline of 200 ms will thus capture all translators' disfluencies.
5. Switching tasks mentally (e.g., from reading to web searching, or to typing) entails an extra cost of at least 200 ms from the control processes involved in setting new task parameters, ending previous task parameters, and overcoming interferences from the latter (Monsell, 2003, p. 135). For instance, reactions to stimuli while talking over the phone take an average of 250 ms longer (Caird et al., 2008). A baseline of 200 ms is not likely to miss any switch costs translators may experience when moving from one task to the next.
6. Logan (1982) finds that motor programs may govern only sub-lexical units and suggests that keystrokes can be controlled individually. The average skilled typist (55 WPM) takes 200 ms to press a single key (Kieras, 2001). The point of no return is customarily set at 166 ms. Subjects take only 250–300 ms from stimulus onset to stop typing, and type about three letters and (longer) overlearned words such as *them.* A baseline of 200 ms will capture all typing goal breaks and changes.
7. Amano et al. (2006, p. 3990) write that the best guess of the time between visual stimulus onset and detection is 150–250 ms (and then another 150–200 ms before subjects react to a stimulus by pressing a key, i.e., 300–450 ms from onset). Writing entails reading, so a baseline of 200 ms ensures that translators' reactions to visual stimuli are covered (even if unregistered).
8. Smith & Levy (2010, p. 1313) state that, in practice, the average fixation length in reading is 200 ms, comparable to the time required to plan a motor saccade. Thus, one of the reasons to set the baseline at 200 ms is fixations and saccades. Eyetracking is not an integral part of the TSF, but if units common to the study of eye movements (fixations and saccades) cohere with the baseline in the TSF, then multi method studies may be able to use the same time scales.

Setting the baseline at 200 ms means that *lags* from 1 ms to 199 ms will be ignored, even when computing timespans and speeds, and will not be subjected to study. This does *not* mean that such lags are not relevant but simply that the shortest IKIs are not a priority given today's research equipment. In keyloggers such as Inputlog, it makes sense to set the minimal pause at 200 ms (it is the default value), so as to make sure that dwell times (IKIs) in these research projects are comparably quantified. The 200 ms baseline has two implications for the TSF: First, all computations of thresholds will not take into account lags. Second, 200 ms will be considered the minimum unit of relevance. The specification at precisely 200 ms is still somewhat arbitrary, but the scope of choice is limited in a principled and easy way, to make different research projects both reliable and comparable.

3.2. *Setting thresholds*

Partially inspired by and drawing on Rosenqvist (2015), we will set two thresholds. An upper threshold will help us distinguish pauses from respites. Pauses are time spans of non-recorded cognitive activities assumed to be mainly task-related, which tend to happen between words and higher language units. They are conscious, intentional and part of the keylogged task flow—which they break down into task segments—but not of typing. Respites are shorter than pauses. They may or may not be conscious, but they are not intentional. Respites are part of typing, and they hint at variations in attentional states. They are actually major indicators of disfluencies, especially when they happen within words or right before a pause or a punctuation mark, or when they come up in chains (i.e., more than one respite within a certain span, see below).

We will use median, instead of mean, values to set thresholds, because IKI distributions are skewed. We are looking for central tendencies and want our values to be somewhat robust against outliers. Not all pauses are planned. Typing breakdowns and sudden reallocations of attentional resources may lead to words and phrases broken by a pause, but these pauses are assumed to be of the same nature as those between complete language units, at word level and above. Some keyloggers ignore the difference between them. We will set the upper threshold at three times the median value for IKIs between words. This is an arbitrary criterion that so far has proved best to identify intended pauses (Muñoz & Martín, 2018; Muñoz & Cardona, 2019; see also Rosenqvist, 2015). Some information as to the graphic representation of log contents after applying the TSF is secondary but helpful. In the graphic representation of keylogged contents, pauses break down the flow into consecutive task segments, displayed as successive lines (Figure 3). Thus, each line starts with the prior pause value in milliseconds in blue, and the next column displays the contents of that task segment. Please note that the source text could have been aligned in an additional, left-most column; the ST would be fragmented so as to parallel the task segments in the translation(s).

As for the lower threshold, it separates respites from delays—the latter assumed to be mainly mechanical, unconscious IKIs, with scarce or no task-related cognitive relevance. We will consider the median of all IKIs within words the *normal transition time* (cf. Wengelin, 2006, p. 127, who averages only IKIs between lowercase letters). We set the lower threshold at two times the median within-word IKI value. This threshold is also arbitrary, and tries to spot IKIs that may flag conscious disturbances (cf. Chukharev-Hudilainen, 2014). We do not multiply per three here because we already removed probably more than 50% of all IKIs when we discarded lags (IKIs <200 ms). Besides, Muñoz & Martín (2018) tried 1.5 times above the median within word and had to face too much noise. This is how thresholds are calculated in the TSF for each subject in each separate keylogged session as of now. Further research will show us how to improve the calculations.

<table>
<tr><th></th><th>Spanish translation</th><th></th><th>Norwegian translation</th></tr>
<tr><td>1840</td><td>cl|E|cl|l_alcalde_se_levant|'|o•, _</td><td>1792</td><td>,_•6•s•a_</td></tr>
<tr><td>1536</td><td>dijo_adi|'|os_c•o⌫</td><td>4525</td><td>far•vel_</td></tr>
<tr><td>4760</td><td>⌫12se_despidi|'|•o•3•⌫11</td><td>4745</td><td>med_••et_•3•⌫2••·m·tb··□·gargle - GC: ordnett·m·casual ¶ casual - GC: ordnett••·m·</td></tr>
<tr><td>2368</td><td>desi⌫pidi|'|endose_•3•con_un_saludo_militar_</td><td>2688</td><td>·m·tb·□·{TT}•·m·‹uformelt</td></tr>
<tr><td>1360</td><td>informal</td><td>5888</td><td>⌫8en_uformell_militær</td></tr>
<tr><td></td><td></td><td>1690</td><td>·m·tb·•·□·casual - GC: ordnett·m·military_salute ¶ military salute - GC: ordnett•3•·⇳·••·□·⇳·•·m·</td></tr>
<tr><td></td><td></td><td>2250</td><td>·m·tb¦·□·{TT}·m·•_ho⌫3honnør_⌫</td></tr>
<tr><td>1887</td><td>,•_y_camin|'|o_haca⌫ia_la_puerta,_estir•ando_las_piernas,_sin_abrocharse_la_t|'|unic•a•.</td><td>2895</td><td>, ⌫_og_gikk_mot_•3•døre•3•n•3•_mens</td></tr>
<tr><td></td><td></td><td>4888</td><td>_hans_⌫2_strekte_på_ben••a••⌫2ina_</td></tr>
<tr><td></td><td></td><td>2082</td><td>⌫•,_uten_å_•kenpp⌫4neppe_opp_igjen_tunik•a•en_.⌫2._••·m·tb·□·military salute - GC: ordnett·⇳·□3·tunic ¶ tunic - GC: ordnett••·m·•3•·⇳·m·tb··□·{TT}·m·</td></tr>
</table>

Figure 3. Examples of the TSF's graphic representation of keylogged translations into Spanish and Norwegian

Delays are not represented graphically (Figures 3 and 4), but they are used in all calculations (e.g., typing speed and IKI length). Respites are represented graphically with a big blue dot • representing increments of 200 ms. A dot represents a respite that equals the lower threshold established for that session + 1 ms, and up to 200 ms above that threshold. Two dots represent a respite between 201 and 400 ms above the lower threshold, etc. Starting with the third level (respites between 401 and 600 ms above the lower threshold), rather than representing them by adding more dots, a number between two dots indicates respite length in 200 ms spans. Hence, •4• symbolizes a respite between 601 and 800 ms above the lower threshold, •5• a respite between 801 and 1000 ms above the lower threshold.

This approach to graphic representation aims to make respite length more meaningful for the researcher at first sight, and more intuitively comparable across informants, without sacrificing accuracy in calculations (which are made based on IKI values in milliseconds). The approach is based upon (a) the observation of the apparent tendency that the shorter the IKI, the lesser the intersubject variation; and (b) 200 ms time spans are meaningful for the reasons provided in § 3.1, and a higher granularity will not necessarily yield better results, especially during the first steps of the TSF's application.

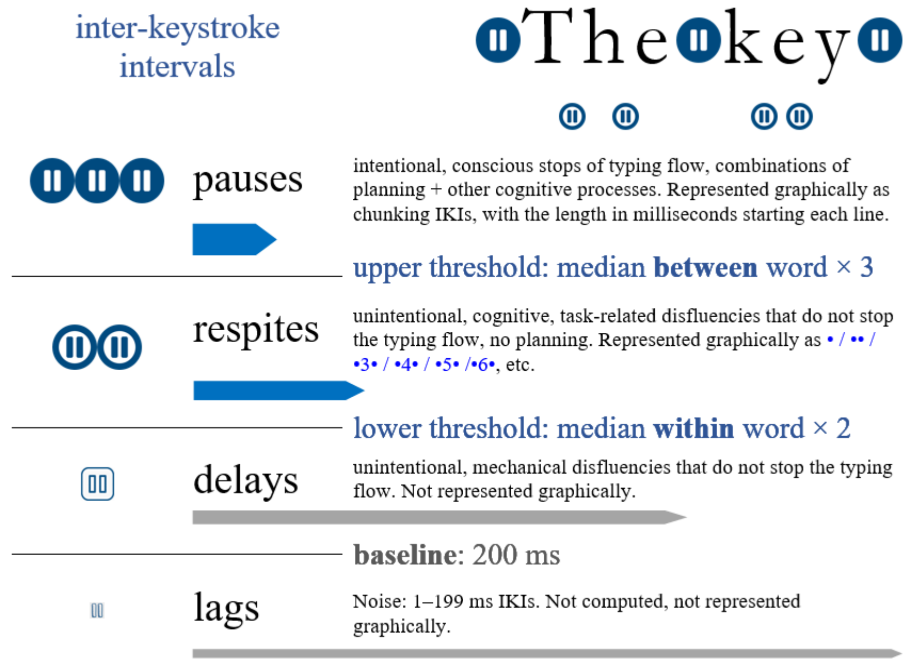

Figure 4. Summary of IKI categories, baseline and thresholds in the Task Segment Framework

3.3. *Task segment categories*

Task segments are excerpts of a keylogged task flow flanked by pauses. Many events come in ordered sequences—building blocks that belong to the behavioral subtasks ADD, CHANGE, SEARCH or HCI (see § 1.3). FILLERS are task segments reflecting minimal (often isolated) ineffective, apparently purposeless clicks, cursor moves and the like. MIX task segments contain combinations of different subtasks (e.g., ADD+CHANGE, ADD+SEARCH, SEARCH+HCI, etc.). Since subtasks are assumed to be potentially associated with particular behavioral repertoires, task segments that only contain events and building blocks of one subtask are thought to be more fluent, in that they are hypothesized not to entail a cognitive effort as high as task segments where more than one behavioral repertoire is active or where task switches between two or more behavioral repertoires take place (instead of switching tasks in pauses). Figure 5 summarizes the basic analytical units in the TSF so far.

Fluent ADD task segments are further assumed to be governed by one or several typing motor programs. Weingarten, Nottbusch & Will (2004, p. 536) found that keystroke timing may be affected by up to three previous keystrokes and the one following it. Besides, Rayner (1975) sets the perceptual span of skilled readers (in left-to-right languages) at about 3–4 characters *to the left.* We should, then, expect sequences of 3–4 keypresses to proceed smoothly and without interruption. Transitioning to the next motor program might result in delays (mechanical, slightly longer than average IKIs) that typing skills and task experience will iron out over time. In brief, there should be no respites in fluent ADD task segments, especially within words.

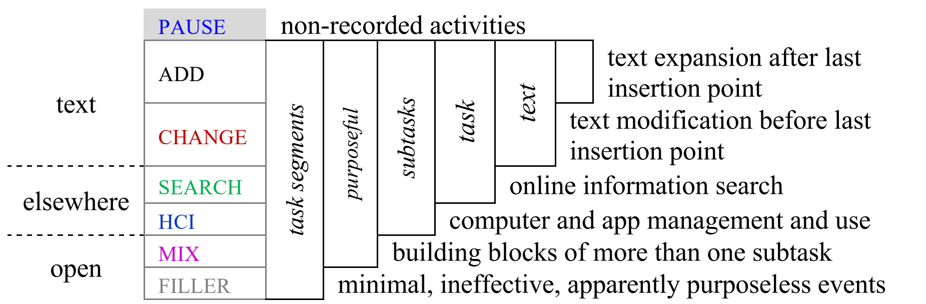

Figure 5. Basic units and color codes in the Task Segment Framework

A respite may be interpreted as indicative of attention and/or resources being drawn away from typing, and thus may hint at some other process going on, or a distraction. This interpretation, of course, stands on shaky ground when no other clue supports it. However, in view of the above, another longer IKI—especially, a second respite within the next 4 keypresses might be hypothesized to be *chained,* i.e., to respond to a single cause, and thus taken as an indicator of a slight struggle in attention and mental resource management. The same can be said about typos. This line is not going to be pursued here but is part of the TSF, and a research priority at that, because it offers a heuristic rule of thumb to further distinguish fluent from non-fluent typing, and the span may need to be fine-tuned.

4. Possibilities by way of conclusion

Based on the notion of fluency, and on the alternation of typing periods and intentional pauses, we have laid down an analytical framework for keylogged multilectal mediated communication tasks that allows for a principled, realistic chunking of the task flow based on observed subtasks. The TSF may allow more precise calculations of typing speed, justify the distinction between current and past WM contents, and accommodate first-pass readings (probably in *superpauses,* i.e., the result of adding the length of two pauses and the filler task segment between them). Muñoz & Martín (2018), Muñoz & Cardona (2019) and Muñoz & Apfelthaler (2021) have explored a wide palette of possible TSF applications to study translation processes. More importantly, applying the TSF may render different research projects on translation and other writing/typing tasks more comparable.

References

Aldridge, M., & Fontaine, L. (2019). Using keystroke logging to capture the impact of cognitive complexity and typing fluency on written language production. In E. Lindgren & K. P. H. Sullivan (Eds.), *Observing writing: Insights from keystroke logging and handwriting* (pp. 285–305). https://doi.org/10.1163/9789004392526_014

Alves, R. A., Castro, S. L., de Sousa, L., & Strömqvist, S. (2007). Influence of typing skill on 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. https://doi.org/10.1163/9781849508223_005

Amano, K., Goda, N., Nishida, S., Ejima, Y., Takeda, T., & Ohtani, Y. (2006). Estimation of the timing of human visual perception from magnetoencephalography. *The Journal of Neuroscience, 26*(15), 3981–3991. https://doi.org/10.1523/JNEUROSCI.4343-05.2006

Baaijen, V., 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

Ballard, D. H., Hayhoe, M. M., Pook, P. K. & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*(4), 723–767. https://doi.org/10.1017/S0140525X97001611

Behmer, L. P., Jr., & Crump, M. J. C. (2017). The dynamic range of response set activation during action sequencing. *Journal of Experimental Psychology: Human Perception and Performance, 43*(3), 537–554. https://doi.org/10.1037/xhp0000335

Caird, J., Willness, C., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention, 40*(4), 1282–1293. https://doi.org/10.1016/j.aap.2008.01.009

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. In *Proceedings of Speech Prosody 2002* (pp. 199–202). Laboratoire Parole et Langage. https://www.isca-speech.org/archive_open/sp2002/sp02_199.pdf

Carl, M. (2012). The CRITT TPR-DB 1.0: A database for empirical human translation process research. In S. O'Brien, M. Simard & L. Specia (Eds.), *Proceedings of the AMTA 2012 Workshop on Post-Editing Technology and Practice* (WPTP 2012) (pp. 9–18). AMTA. https://aclanthology.org/2012.amta-wptp.1.pdf

Castro Arce, Mª. (2008). Procesos de lectura y comprensión al traducir. In M. M. Fernández & R. Muñoz Martín (Eds.), *Aproximaciones cognitivas al estudio de la traducción y la interpretación* (pp. 31–54). Comares.

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/jowr-2014.06.01.3

Ciobanu, D. I. (2016). Automatic speech recognition in the professional translation process. *Translation Spaces, 5*(1), 124–144. https://doi.org/10.1075/ts.5.1.07cio

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

Cowan, N. (2005). *Working memory capacity*. Psychology Press. https://doi.org/10.4324/9780203342398

Crump, M. J. C., & Logan, G. D. (2010). Hierarchical control and skilled typing: Evidence for word-level control over the execution of individual keystrokes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1369–1380. https://doi.org/10.1037/a0020696

Crump, M. J. C., & Logan, G. D. (2013). Prevention and correction in post-error performance: An ounce of prevention, a pound of cure. *Journal of Experimental Psychology: General, 142*(3), 692–709. https://doi.org/10.1037/a0030014

Dam-Jensen, H., & Heine, C. (2013). Writing and translation process research: Bridging the gap. *Journal of Writing Research, 5*(1), 89–101. https://doi.org/10.17239/jowr-2013.05.01.4

de Jong, M., Jolij, J., Pimenta, A., & Lorist M. M. (2018). Age modulates the effects of mental fatigue on typewriting. *Frontiers in Psychology, 9*(1113), 1–15. https://doi.org/10.3389/fpsyg.2018.01113

Dhakal, V., Feit, A. M., Kristensson, P. O., & Oulasvirta, A. (2018). Observations on typing from 136 million keystrokes. In *Proceedings of the 2018 CHI Conference* *on Human Factors in Computing Systems (CHI '18)* (pp. 1–12). https://doi.org/10.1145/3173574.3174220

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., & Hansen, I. G. (2008). Comprehension and production in translation: A pilot study on segmentation and the coordination of reading and writing processes. *Copenhagen Studies in Language*, (36), 9-29.

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

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* Academic Press.

Gouadec, D. (2007). *Translation as a profession.* John Benjamins. https://doi.org/10.1075/btl.73

Gray, W. D. & Fu, W.T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in the world for imperfect knowledge in the head. *Cognitive Science, 28*(3), 359–382. https://doi.org/10.1207/s15516709cog2803_3

Halverson, S. L. (2019). 'Default' translation: A construct for cognitive translation and interpreting studies. *Translation, Cognition & Behavior, 2*(2), 187–210. https://doi.org/10.1075/tcb.00023.hal

Hansen, G. (2002). Zeit und Qualität im Übersetzungsprozess. In G. Hansen (Ed.), *Empirical translation studies: Process and product* (pp. 29–54). Samfundslitteratur.

Harris, M., & Coltheart, M. (1986). *Language processing in children and adults: An introduction*. Routledge.

Hauk, O., Coutout, C., Holden, A., & Chen, Y. (2012). The time-course of single-word reading: Evidence from fast behavioral and brain responses. *NeuroImage, 60*(2), 1462–1477. https://doi.org/10.1016/j.neuroimage.2012.01.061

Holaday, B. (2015). Dorothy Johnson's behavioral system model and its applications. In M. C. Smith & M. E. Parker (Eds.), *Nursing theories and nursing practice* (pp. 89–104). F. A. Davis.

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., & Mäkisalo, J. (2010). Pauses reflecting the processing of syntactic units in monolingual text production and translation. *Hermes, 44,* 45–61. https://doi.org/10.7146/hjlcb.v23i44.97266

Jakobsen, A. L., & Jensen, K. T. H. (2008). Eye movement behaviour across four different types of reading tasks. In S. Göpferich, A. L. Jakobsen & I. M. Mees (Eds.), *Looking at eyes: Eye-tracking studies of reading and translation processing* (pp. 103–124). Samfundslitteratur.

Johansson, R., Wengelin, Å., Johansson, V., & Holmqvist, K. (2010). Looking at the keyboard or the monitor: Relationship with text production processes. *Reading and Writing, 23*(7), 835–851. https://doi.org/10.1007/s11145-009-9189-3

Kaufer, D., Hayes, J. R., & Flower, L. S. (1986). Composing written sentences. *Research in the teaching of English, 20*(2), 121–140. https://www.jstor.org/stable/40171073.

Kellogg, R. T. (1999). Components of working memory in text production. In M. Torrance & G. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory effects in text production* (pp. 43–61). Amsterdam University Press.

Kester, L., & Kirschner, P. A. (2012). Cognitive tasks and learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 619–622). Springer. https://doi.org/10.1007/978-1-4419-1428-6_225

Kieras, D. (2001). *Using the keystroke-level model to estimate execution times.* PhD dissertation, University of Michigan. http://www.cs.loyola.edu/~lawrie/CS774/S06/homework/klm.pdf

Lacruz, I., & Shreve, G. M. (2014). Pauses and cognitive effort in post-editing. In S. O'Brien, L. Winther Balling, M. Carl, M. Simard & L. Specia (Eds.), *Post-editing of machine translation: Processes and applications* (pp. 246–272). Cambridge Scholars Publishing.

Läubli, S., Sennrich, R., & Volk, M. (2018). Has machine translation achieved human parity? A case for document-level evaluation. In *Proceedings of EMNLP* (pp. 4791–4796). https://doi.org/10.18653/v1/D18-1512

Leijten, M., Van Horenbeeck, E., & Van Waes, L. (2019). Analysing keystroke logging data from a linguistic perspective. In E. Lindgren & K. P. H. Sullivan (Eds.), *Observing writing: Insights from keystroke logging and handwriting* (pp. 71–95). Brill. https://doi.org/10.1163/9789004392526_005

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

Leijten, M., Van Waes, L., Schriver, K., & Hayes, J. R. (2014). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research, 5*(3), 285–337. https://doi.org/10.17239/jowr-2014.05.03.3

Limpo, T., & Alves, R. A. (2017). Written language bursts mediate the relationship between transcription and writing performance. *Written Communication, 34*, 306–332. https://doi.org/10.1177/0741088317714234

Logan, G. D. (1982). On the ability to inhibit complex movements: A stop-signal study of typewriting. *Journal of Experimental Psychology: Human Perception and Performance, 8*(6), 778–792. https://doi.org/10.1037/0096-1523.8.6.778

Logan, G. D., & Crump, M. J. C. (2011). Hierarchical control of cognitive processes: The case for skilled typewriting. In B. H. Ross (Ed.), *Advances in research and theory* (pp. 1–27). Academic Press. https://doi.org/10.1016/B978-0-12-385527-5.00001-2

Logan, G. D., Miller, A. E., & Strayer, D. L. (2011). Electrophysiological evidence for parallel response selection in skilled typists. *Psychological Science, 22*(1), 54–56. https://doi.org/10.1177/0956797610390382

Madl, T., Baars, B. J., & Franklin, S. (2011). The timing of the cognitive cycle. *PLoS ONE, 6*(4), e14803. https://doi.org/10.1371/journal.pone.0014803

Mead, P. (2000). Control of pauses by trainee interpreters in their A and B languages. *The Interpreters' Newsletter, 10,* 89–102. https://hdl.handle.net/10077/2451

Medimorec, S., & Risko, E. F. (2016). Effects of disfluency in writing. *British Journal of Psychology, 107*(4), 625–650. https://doi.org/10.1111/bjop.12177

Medimorec, S., Young, T. P., & Risko, E. F. (2017). Disfluency effects on lexical selection. *Cognition, 158,* 28–32. https://doi.org/10.1016/j.cognition.2016.10.008

Monsell S. (2003). Task switching. *Trends in Cognitive Science, 7*(3), 134–140. https://doi.org/10.1016/s1364-6613(03)00028-7

Muñoz Martín, R. (2010). On paradigms and cognitive translatology. In G. Shreve & E. Angelone (Eds.), *Translation and cognition* (pp. 169–187). John Benjamins. https://doi.org/10.1075/ata.xv.10mun

Muñoz Martín, R., & Apfelthaler, M. (2021). Spillover effects in task-segment switching: A study of translation subtasks as behavioral categories within the Task Segment Framework. In R. Muñoz, S. Sun & D. Li (Eds.), *Advances in cognitive translation studies* (pp. 19–45). Springer. https://doi.org/10.1007/978-981-16-2070-6_2

Muñoz Martín, R. & Cardona Guerra, J. Mª. (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., & González Fernández, C. A. (2021). Cognitive translatology: A primer, revisited. *语言、翻译与认知 (Studies in Language, Translation & Cognition), 1,* 131–165.

Muñoz Martín, R., & Martín de León, C. (2018). Fascinatin' rhythm – And pauses in translators' cognitive processes. *HERMES - Journal of Language and Communication in Business*, *57*, 29–47. https://doi.org/10.7146/hjlcb.v0i57.106192

Newell, A. (1990). *Unified theories of cognition.* Harvard University Press.

Olive, T., & Kellogg, R. T. (2002). Concurrent activation of high- and low-level production processes in written composition. *Memory & Cognition, 30*(4), 594–600. https://doi.org/10.3758/BF03194960

Pinet, S., Ziegler, J. C., & Alario, F. (2016). Typing is writing: Linguistic properties modulate typing execution. *Psychonomic Bulletin & Review, 23*(6), 1898–1906. https://doi.org/10.3758/s13423-016-1044-3

Pinet, S., Zielinski, C., Mathôt, S., Dufau, S., Alario, F. X., & Longcamp, M. (2017). Measuring sequences of keystrokes with jsPsych: Reliability of response times and interkeystroke intervals. *Behavior Research Methods, 49*(3), 1163–1176. https://doi.org/10.3758/s13428-016-0776-3

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences, 5*(12), 517–524. https://doi.org/10.1016/s1364-6613(00)01803-9

Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior Research Methods, 47*(2), 309–327. https://doi.org/10.3758/s13428-014-0471-1

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7*(1), 65–81. https://doi.org/10.1016/0010-0285(75)90005-5

Risku, H., Milošević, J., & Pein-Weber, C. (2016). Writing vs. translating: Dimensions of text production in comparison. In R. Muñoz Martín (Ed.), *Reembedding translation process research* (pp. 47–68). John Benjamins. https://doi.org/10.1075/btl.128.03ris

Rosenqvist, S. (2015). *Developing pause thresholds for keystroke logging analysis.* B. A. thesis, University of Umea, Sweden. http://www.diva-portal.org/smash/get/diva2:834468/FULLTEXT01.pdf

Scaltritti, M., Alario, F.-X., & Longcamp, M. (2018). The scope of planning serial actions during typing. *Journal of Cognitive Neuroscience, 30*(11), 1620–1629. https://doi.org/10.1162/jocn_a_01305

Shaffer, L. H. (1975). Multiple attention in continuous verbal tasks. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 157–167). Academic Press.

Shreve, G. M., Schäffner, C., Danks, J. H., & Griffin, J. (1993). Is there a special kind of 'reading' for translation? An empirical investigation of reading in the translation process. *Target, 5*(1), 21–41. https://doi.org/10.1075/target.5.1.03shr

Smith, N. J., & Levy, R. (2010). Fixation durations in first-pass reading reflect uncertainty about word identity. *Proceedings of the Annual Meeting of the Cognitive Science Society, 32,* 1313–1318. https://escholarship.org/uc/item/2d09z0n9

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.* https://www.usenix.org/legacy/publications/library/proceedings/sec01/song.html

Strömqvist, S. (1999). Production rate profiles. In S. Strömqvist, & E. Ahlsén (Eds.), *The process of writing: A progress report* (pp. 53–70).

Toral, A., Castilho, S., Hu, K., & Way, A. (2018). Attaining the unattainable? Reassessing claims of human parity in neural machine translation. In *Proceedings of the Third Conference on Machine Translation: Research papers* (pp. 113–123). https://doi.org/10.18653/v1/W18-6312

Urquhart, A. H., & Weir, C. J. (1998). *Reading in a second language: Process, product and practice.* Longman.

VandenBos, G. (Ed.), (2015). *APA dictionary of psychology* (2nd ed.). APA.
Van Waes, L., Leijten, M., Lindgren, E., & Wengelin, Å. (2016). Keystroke logging in writing research: Analyzing online writing processes. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 410–426). The Guilford Press.
Vieira, L. Nunes (2013). An evaluation of tools for post-editing research: The current picture and further needs. In S. O'Brien, M. Simard & L. Specia (Eds.), *Proceedings of the MT Summit XIV Workshop on Post-editing Technology and Practice (WPTP-2), Nice, September 2, 2013* (pp. 93–101). https://research-information.bristol.ac.uk/files/53392966/MTS_2013_W2_Vieira.pdf
Vieira, L. Nunes. (2017). From process to product: Links between post-editing effort and post-edited quality. In A. L. Jakobsen & B. Mesa Lao (Eds.), *Translation in transition: Between cognition, computing and technology* (pp. 162–186). John Benjamins. https://doi.org/doi:10.1075/btl.133.06vie
Weingarten, R., Nottbusch, G., & Will, U. (2004). Morphemes, syllables and graphemes in written word production. In T. Pechman & C. Habel (Eds.), *Multidisciplinary approaches to language production* (pp. 529–572). Mouton de Gruyter. https://doi.org/10.1515/9783110894028.529
Wengelin, Å. (2006). Examining pauses in writing: Theory, methods and empirical data. In K. Sullivan & E. Lindgren (Eds.), *Computer keystroke logging and writing: Methods and applications* (pp. 107–130). Elsevier.
Yamaguchi, M., & Logan, G. D. (2013). Pushing typists back on the learning curve: Revealing chunking in skilled typewriting. *Journal of Experimental Psychology: Human Perception & Performance, 40*(2), 592–612. https://doi.org/10.1037/a0033809
Zhang, M., Zhu, M., Deane, P., & Guo, H. (2019). Identifying and comparing writing process patterns using keystroke logs. In M. Wiberg, S. Culpepper, R. Janssen, J. González & D. Molenaar (Eds.), *Quantitative psychology.* Springer. https://doi.org/10.1007/978-3-030-01310-3_32