Gilquin, Gaëtanelle. 2025. Keylogging and screencasting to help investigate L2 writing processes. In Karim Sadeghi (ed.) *Routledge Handbook of Technological Advances in Researching Language Learning* (pp. 285-296). London: Routledge.

Keylogging and screencasting to help investigate L2 writing processes

Gaëtanelle Gilquin

Abstract

L2 writing is often investigated on the basis of the product, that is, the final text as made available to the reader. This chapter shows how the technologies of keylogging and screencasting make it possible to examine writing processes, that is, anything that happens while the text is being composed, and how such information can be fruitfully used to investigate L2 writing. Three possible applications are provided as examples: the study of writing fluency, the cognitive processing of multiword units, and the use of external sources in writing. Some concrete results derived from the Process Corpus of English in Education (PROCEED) are presented. These examples serve to illustrate the affordances and possible challenges of keylogging and screencasting as techniques of investigation in L2 writing studies. Future directions for keylogging- and screencasting-based research are also suggested.

Keywords: L2 writing, writing processes, keylogging, screencasting, Process Corpus of English in Education

1. Background

"The ability to express one's ideas in writing in a second or foreign language and to do so with reasonable coherence and accuracy is a major achievement". This quote by Celce-Murcia (2001, p. 206) highlights the challenge that writing represents for second/foreign language (L2) learners. It is therefore not surprising that scholars have sought to gain a better understanding of L2 writing (how it is learned, how it develops, how it can be taught, etc.) and that L2 writing has become a well-established field of research (Manchón, 2018).

While different methods have been applied to investigate L2 writing, they mostly involve examining the written product, that is, the text as it stands at the end of the composition process. In learner corpus research, for example, which has greatly contributed to our understanding of L2 writing (see Granger et al., 2015), the main sources of information are databases of texts written by L2 learners which are only visible in their final forms. Writing processes, that is, all of the steps that have led to these final texts, are not accessible in typical written learner corpora.

Attention to writing processes garnered prominence in the 1970s through the so-called 'process approach' in teaching. The process approach views writing as being made up of several stages (such as prewriting, drafting, revising, and editing) and advocates the teaching of specific writing skills associated with each of these stages (see Seow, 2002). This pedagogical interest gave impetus to research on L2 writing processes, as "it was believed that for teaching procedures to be effective, they should be based on a theory that accounted for what writers actually do in the process of composing a text" (Roca De Larios et al., 2002, pp. 11-12).

Research on L2 writing processes sometimes relies on several drafts of the same text, as in Tuzi (2004). Such an approach offers valuable information on writing processes (and, in the case of a study like Tuzi (2004), the impact of feedback). However, it is limited to a few snapshots of the composition process, corresponding to the number of drafts available. What happens in between these drafts is unknown to the researcher. In this respect, the common use of the computer to write texts and the development of technologies have made it possible to go one step further by giving access to writing processes as they unfold. Among these technologies are keylogging and screencasting, which are the focus of this chapter.[1]

Section 2 introduces these two technologies, while Section 3 reviews some of their possible applications. Section 4 addresses the affordances and challenges of keylogging and screencasting. Section 5, finally, concludes the chapter and suggests some future directions for research on L2 writing processes based on these technologies.

2. Introducing keylogging and screencasting

Keylogging (also called keystroke logging) involves recording all keys struck on the keyboard. Although originally developed as a tool for surveillance (sometimes used illegally), the keylogger started to be used for research purposes around the turn of the millennium. Examples of keyloggers designed for research are ScriptLog (Strömqvist & Malmsten, 1997) and Inputlog (Leijten & Van Waes, 2013).

Keyloggers for writing research record detailed information about writing processes, including the keys that are struck, the positioning of the cursor, pause time, and sometimes also mouse movements and clicks. In addition, most of them allow for different types of quantitative analysis. Inputlog performs, among others, pause analysis (with statistics such as number of pauses or distribution of pause location) and bigram analysis (lists of two consecutive elements, including sequences of categories, e.g. character + space, and sequences of letters, e.g. ts). It also provides some visualizations, such as a fluency graph and a process graph (including information about the consultation of external sources).

Some keyloggers have a function to replay the writing session. The replay is based on the recorded keystrokes, which make it possible to reconstruct the writing process in a video-like manner. Despite the intrinsic value of this replay function, its data-based (rather than video-based) origin means that it is not always fully reliable (see Leijten & Van Waes, 2020 on Inputlog). For research that requires a replay, another option is available, namely screencasting (also known as screen recording or video screen capture). Screencasting records anything that is visible on the screen during the writing session. Its output is a video that shows writing processes in real time. Many programs allow for screencasting, including basic tools such as QuickTime Player. More specialized tools (usually not designed for research) include Camtasia and OBS Studio.

Although progress has been made with the automatic analysis of video content (see Siahaan & Chowanda, 2022), most of the time screencast videos still have to be processed manually, by actually watching the videos and coding them for relevant information, for example by means of annotation tools like ELAN (Brugman & Russel, 2004) or EXMARaLDA (Schmidt &

[1] Other technologies include the logging of handwriting (e.g. Alves et al., 2019) and eye-tracking, which makes it possible to record eye movement during a writing session (e.g. Hacker et al., 2017).

Wörner, 2014). After manual coding has been carried out, however, data analysis can usually be automated.

3. Possible applications

Keylogging and screencasting, by giving access to writing processes as they unfold, have opened up new possibilities for exploring different aspects of L2 writing. In what follows, some of these aspects are reviewed, namely writing fluency (Section 3.1), cognitive processing of multiword units (Section 3.2), and use of external sources (Section 3.3). It will be shown how the technologies of keylogging and screencasting have enhanced our understanding of these aspects. Examples from the literature are provided, as well as illustrations based on the Process Corpus of English in Education (PROCEED; Gilquin, 2022).

PROCEED is a corpus of argumentative essays produced by university students who are learners of English as a foreign language. Learners write a text in the target language (English), and some also write a text in their mother tongue (L1) – mostly French, so far. Each text is accompanied by the corresponding keylog file (recorded by means of Inputlog) and screencast video (recorded by means of OBS Studio), which make learners' writing processes accessible. The corpus also includes metadata that provide information about learners' sociolinguistic profiles (mother tongue, type and amount of exposure to English, etc.) and some of their cognitive and motor abilities (working memory, typing speed, etc.).

3.1 Writing fluency

Although fluency is a concept that is mostly investigated in speech, it can be applied to writing as well. Writing fluency can be defined as the ease with which writing is produced. Writing is fluent if the composition process is smooth and rapid and the writer does not pause or revise excessively. Traditionally, writing fluency has been established by means of product-based measures, such as the total number of words in the finished text and the average number of words produced per minute, calculated based on the number of words in the finished text and the time taken to write it (see, e.g., Sasaki, 2000). Occasionally, these have been complemented by process-based information. In Ellis and Yuan (2004), revisions are measured as the number of words crossed out and changed in handwritten manuscripts. Chenoweth and Hayes (2001) rely on voice-recorded think-aloud protocols, whereby writers explain what they are doing and thinking while writing. The transcriptions of these protocols serve as a basis to identify bursts of writing, that is, uninterrupted sequences of words terminated by a pause (P-bursts) or by a revision (R-bursts). It should be noted, however, that the degree of fluency thus established applies to the verbal protocols (i.e. speech) and may not be an accurate reflection of the writing process (e.g. writers could pause their verbalization but continue writing).

Observation of writing processes thanks to the technology of keylogging has allowed for a much more nuanced and accurate picture of writing fluency. Xu and Ding (2014) calculate writing fluency as the number of words produced per minute to compare their results with those of previous studies. However, they also compute the number and duration of pauses using keylogging. In Spelman Miller et al. (2008), the number of characters produced per minute is combined with the number of characters in bursts. Van Waes and Leijten (2015) examine a wide range of fluency-related variables and show how this offers a multidimensional perspective on writing fluency. Their analysis starts with over 200 variables, such as the

number of pauses, mean pause length, total pause time, length of P-bursts and R-bursts, and the number of insertions and deletions in the writing process. A principled selection of ten variables is then submitted to Principal Component Analysis, leading to the identification of four writing fluency components, namely production, process variation, revision, and pause behaviour.

Such process-based investigations have helped describe L2 writing fluency in a more precise way and in different types of writing contexts. Spelman Miller (2000), for example, compares fluency in a descriptive writing task and an evaluative one. Xu and Ding (2014) examine pause frequency and duration in different stages of the writing session. Xu and Qi (2017) compare pause behaviour among skilled and less skilled writers, while Spelman Miller et al. (2008) consider the evolution of writing fluency over a period of three years. Potential correlations between writing fluency and aspects of the finished texts have also been investigated, such as text quality in Spelman Miller et al. (2008).

The study of a sample from PROCEED has made it possible to compare L1 and L2 writing fluency based on keylogging data. The sample comprises texts produced by 38 students writing in their L1 (French) and L2 (English) for about 40-45 minutes. Both the so-called pause analysis and revision matrix of Inputlog have been used in the analysis. While the former provides a summary of statistics related to pause behaviour, the latter shows the number and type of revisions (insertions and deletions). The examined variables concern speed (number of keystrokes per minute), pausing (proportion of pause time, number of pauses, length of P-bursts, location of pauses), and revision (number of deletions and insertions, length of R-bursts, and product/process ratio, i.e. proportion of typed characters kept in the final text).

Table 1 summarizes the main results of the comparison, providing averages and standard deviations (SD) for the L1 and L2 data. It shows that for most variables, L1 production is more fluent than L2 production. The only exception is the total number of deletions, pointing to a slightly higher degree of fluency in L2 than in L1. Pause location does not seem to distinguish reliably between L1 and L2, as pause time within and between words is almost identical in the two sets of data (pause time between sentences indicates a higher degree of fluency in L1, however). Finally, most SD values underline the great variation among students.

Such quantitative results are only a first step in analyzing L1 and L2 writing fluency. The number of keystrokes per minute, for example, may partly be due to students' average typing speed and should therefore be assessed in relation to the results of a copy task (Van Waes et al., 2021) that all writers participating in the PROCEED project had to carry out. Revisions should be examined in context to determine whether they are true signs of disfluency (e.g. the correction of a typo) or whether they aim at stylistic improvement (e.g. replacing a word with a more formal synonym). As for pauses, they may reflect different processes (Schilperoord, 2002) and do not necessarily indicate a lack of fluency. Other methods are often necessary to identify the functions of pauses in context (see Section 4).

Variable	L1	L2	Outcome
Keystrokes per minute	75.01 (SD=16.61)	54.72 (SD=12.24)	L1 + fluent than L2
Proportion of pause time	41.88% (SD=12.29)	43.87% (SD=9.30)	L1 + fluent than L2
Total number of pauses	103.26 (SD=27.50)	158.03 (SD=40.95)	L1 + fluent than L2
Length of P-bursts (in characters)	13.33 (SD=6.96)	9.74 (SD=4.56)	L1 + fluent than L2
Pause time within words	3.13 (SD=1.94)	3.04 (SD=1.12)	*No difference*
Pause time between words	3.95 (SD=0.96)	3.92 (SD=0.62)	*No difference*
Pause time between sentences	4.84 (SD=3.27)	6.02 (SD=4.78)	L1 + fluent than L2
Total number of deletions	137.39 (SD=63.63)	134.39 (SD=58.98)	**L2 + fluent than L1**
Total number of insertions	76.39 (SD=69.35)	93.55 (SD=78.99)	L1 + fluent than L2
Length of R-bursts (in characters)	18.74 (SD=10.37)	17.34 (SD=9.65)	L1 + fluent than L2
Product/process ratio	65.81% (SD=20.17)	64.34% (SD=17.30)	L1 + fluent than L2

Table 1. Averages and standard deviations (SD) for fluency measures in L1 and L2 PROCEED data

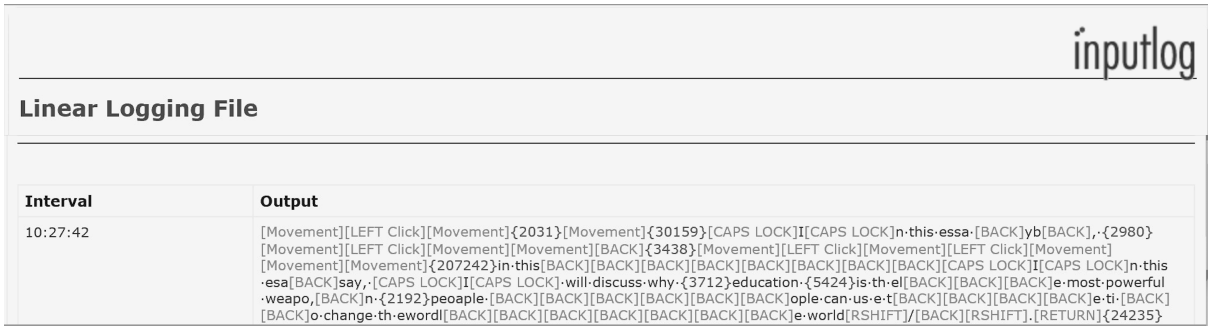
3.2 Cognitive processing of multiword units

Linguists have often sought to make claims about cognitive aspects of language, particularly its mental processing. One linguistic phenomenon that has attracted special attention in this respect is the multiword unit, that is, a sequence of words functioning, at some level, as a single entity (e.g. compound noun, phrasal verb). Alongside psycholinguistic and neurolinguistic evidence (see, e.g., Schmitt, 2004; Siyanova-Chanturia et al., 2017), textual evidence has been used to probe into the processing of multiword units. In corpus linguistics, whose main source of information is the (finished) text, multiword units and, more generally, phraseological phenomena have often been claimed to be processed holistically. Sinclair (1991, p. 110), for example, has put forward the idiom principle, according to which "a language user has available to him or her a large number of semi-preconstructed phrases *that constitute single choices*, even though they might appear to be analysable into segments" (emphasis added). Frequency in texts is regularly taken as a sign that a sequence of words is stored and processed as a whole in the mind, as illustrated by Biber et al.'s (2004, p. 376) quote: "Frequency data […] are one reflection of the extent to which a sequence of words is stored and used as a prefabricated chunk, with higher frequency sequences more likely to be stored as unanalysed chunks than lower frequency sequences". The strength of association between the elements making up a multiword unit has also been considered an indicator of holistic processing. In

Wahl (2015, p. 191), for instance, it is used as an "index of mental storage". Like frequency, however, it is based solely on the occurrence of the words in finished texts.

Writing process data have been argued to offer privileged access to cognitive phenomena. Barkaoui (2019, p. 535) points out that keystroke logging "allows the close analysis of the linguistic and textual output to draw inferences about the cognitive processes from which it is generated". Such inferences have occasionally concerned the processing of multiword units. More particularly, it has been suggested that bursts of writing (introduced above in relation to fluency) could "indicate a certain degree of automatization, which would result in retrieval from long-term memory rather than the implementation of text generation processes" (Cislaru & Olive, 2017, p. 15; my translation from French). This builds on similar claims made for speech, according to which multiword units are likely to be delineated by pauses but not interrupted by them (e.g. Dechert, 1983; Pawley, 1986).

Within the PROCEED project, P-bursts have been used as a basis to identify potential candidates for holistic processing, following a method similar to that adopted by Dahlmann and Adolphs (2007) for speech (see Gilquin, 2024). The study starts from recurrent sequences of two to six words (n-grams) in the finished texts of an L2 PROCEED sample produced by 42 French-speaking learners of English. These n-grams are searched for semi-automatically in the process data, and more precisely in the output of Inputlog's linear analysis, which represents pauses and keystrokes. Figure 1 illustrates the beginning of the linear analysis of a text in PROCEED. The numbers between curly brackets indicate the duration of pauses (in milliseconds), actions such as deletion or capitalization are found between square brackets, and the rest corresponds to the characters that are typed. The linear analysis in Figure 1 includes one of the n-grams investigated, namely *the most* (end of line 4), whose retrieval from the data cannot be fully automatic given the traces of revision found in the linear analysis (th·el[BACK][BACK][BACK]e·most).


inputlog

Linear Logging File

Interval	Output
10:27:42	[Movement][LEFT Click][Movement]{2031}[Movement]{30159}[CAPS LOCK]I[CAPS LOCK]n·this·essa·[BACK]yb[BACK],·{2980} [Movement][LEFT Click][Movement][Movement][BACK]{3438}[Movement][LEFT Click][Movement][LEFT Click][Movement] [Movement][Movement]{207242}in·this[BACK][BACK][BACK][BACK][BACK][BACK][BACK][BACK][CAPS LOCK]I[CAPS LOCK]n·this ·esa[BACK]say,·[CAPS LOCK]I[CAPS LOCK]·will·discuss·why·{3712}education·{5424}is·th·el[BACK][BACK][BACK]e·most·powerful ·weapo,[BACK]n·{2192}peoaple·[BACK][BACK][BACK][BACK][BACK][BACK][BACK]ople·can·us·e·t[BACK][BACK][BACK][BACK]e·ti·[BACK] [BACK]o·change·th·ewordl[BACK][BACK][BACK][BACK][BACK][BACK][BACK]e·world[RSHIFT]/[BACK][RSHIFT].[RETURN]{24235}

Figure 1. Beginning of the linear analysis of a PROCEED sample in Inputlog (Leijten & Van Waes, 2013)

All n-grams extracted from the output of the linear analysis (in bold in the examples below) are annotated manually according to their pause placement pattern: _X_ for n-grams with no immediately adjacent pauses (1), PX_ for n-grams immediately preceded (but not followed) by a pause (2), _XP for n-grams immediately followed (but not preceded) by a pause (3), PXP for n-grams immediately preceded and followed by a pause (corresponding to P-bursts) (4), and XPX for n-grams interrupted by at least one pause (5).

(1) {2680}we[BACK][BACK]be·well-positioned·**in·the**·society·[BACK][LSHIFT].·{4792}
(2) {3592}·**in·order·to**·developo[BACK]·their·critical·mind,·to·make·them·{2440}

(3) {4144}[LSHIFT]Education·is·**the·most**·{4144}
(4) {3080}**point·of·view**{2120}
(5) {2312}**of**·{3248}**the**·small

Thanks to the annotation of n-grams in the process data, a so-called Pause Placement and Processing (PPP) score is computed per n-gram, taking into account the likelihood of each pause placement pattern indicating holistic processing, e.g. a high probability for PXP and a low probability for XPX (see Gilquin, 2024, for further details about the computation of the PPP score). Table 2 lists the five n-grams with the highest PPP scores (on the left) and the five n-grams with the lowest PPP scores (on the right).

N-gram	**PPP score**
in fact	71.43
first of all	67.86
to sum up	65.00
for instance	63.89
in my opinion	60.00

N-gram	**PPP score**
us to	4.76
able to	2.38
that we	0.00
of the	-9.60
to the	-16.07

Table 2. N-grams with the highest (left) and lowest (right) Pause Placement and Processing (PPP) scores in L2 PROCEED data (based on Gilquin, 2024)

It appears that all n-grams with a high PPP score are structurally complete: *in fact*, *first of all*, *to sum up* (which could potentially be followed by a direct object but in the data is mostly used on its own), *for instance*, and *in my opinion*. This could mean that structural coherence goes hand in hand with holistic processing. Interestingly, the literature has shown these five specific n-grams to be overused by learners in general or French-speaking learners in particular, which suggests that they are strongly entrenched in learners' minds. N-grams with a low (or negative) PPP score, by contrast, are structurally incomplete: *us to*, *able to*, *that we*, *of the*, and *to the*. While the degree of holistic processing of these n-grams would have to be checked against more experimental evidence, these results seem plausible. One would expect an internally coherent unit like *in fact* to be stored as a whole in the mind, unlike a structurally incomplete sequence like *to the*. The use of writing process data thus seems promising to study the cognitive processing of multiword units.

3.3 Use of external sources

The use of external sources such as dictionaries or secondary texts has typically been examined retrospectively, that is, after the writing has taken place. This can be done by means of questionnaires that users are required to fill in. In Béjoint (1981), for example, French learners of English are asked questions about their use of (printed) monolingual English dictionaries, including the type of information that they most often look for (meaning, spelling, syntactic information, etc.) and occasions when they were not able to find what they were looking for. Such questions can also be asked in the form of an interview, sometimes in combination with a questionnaire (cf. Knežević et al., 2021). A possible issue with this approach is that writers may not be aware of their consultation behaviour and may not remember the details of what

they did and how. Thus, Béjoint (1981) points out that many participants were unable to recall occasions when they could not find what they were looking for or specific examples of words they could not locate in the dictionary. The information collected in this way may therefore not be entirely reliable. Hatherall (1984, p. 184), in his discussion of the limitations of questionnaire-based studies, asks the following: "Are subjects saying here what they do, or what they think they do, or what they think they ought to do, or indeed a mixture of all three?", which suggests that perception and prescription may bias the results.

To avoid such difficulties, Hatherall (1984, p. 184) advocates direct observation, which he describes as "the only reliable method of collecting data on dictionary user behaviour". His method involves providing students with a form at the beginning of a writing session (more precisely, a translation task) and asking them to complete the form every time they use a dictionary. The form requires the students to supply details such as: search item, whether the item was found or not, and if so, whether the information was useful and where exactly it was found (under what headword). Chon (2009) relies on think-aloud protocols to discover how Korean learners of English use electronic dictionaries when writing in English. During the writing task, learners are required to verbalize what they are doing and thinking, including while consulting a dictionary, and their verbalizations are recorded by means of a digital voice recorder. Such methods offer some access to dictionary consultation. However, they have been criticized for diverting writers' attention from the writing process itself and hence possibly distorting this process (see, e.g., Janssen et al., 1996).

Thanks to technologies that record processes in real time, observation of the use of external sources can be more direct and less disruptive. In Laufer and Levitzky-Aviad (2006), the look-up behaviour of Hebrew learners of English is investigated using the built-in log files of an electronic dictionary, which the authors describe as "an unobtrusive research instrument" (Laufer & Levitzky-Aviad 2006, p. 142). Hamel (2012) relies on screencasting to test an online dictionary prototype and observe how learners of French use it. In Leijten et al. (2019), keylogging helps examine students' use of three types of secondary sources (report, web text, and newspaper article). The keylogging program, Inputlog, enables the scholars to measure (i) the time students spend consulting these source texts, (ii) when during the writing process they do so, (iii) which texts they consult most often, and (iv) how often they switch between texts.

Gilquin and Laporte (2021) analyze a sample from PROCEED, focusing on screencast videos to examine how 84 French-speaking learners use online writing tools (dictionaries, corpus interfaces, language forums, etc.) when writing in L2 English. The videos were annotated by means of the ELAN software. For each video segment showing the use of an online writing tool, 11 tiers of annotation were completed, including the tool used, the type of information searched for, and the effect (i.e. whether the search was successful or not). Figure 2 shows an extract from a video annotated in ELAN. The video plays in the upper left part, and the annotation tiers are found in the lower part (only the first few tiers are visible in the screenshot). In the selected extract, the learner searches for the translation of the French word *salle* in Linguee and then looks up one of the translations, namely *hall*, in Thesaurus.com to find a synonym.

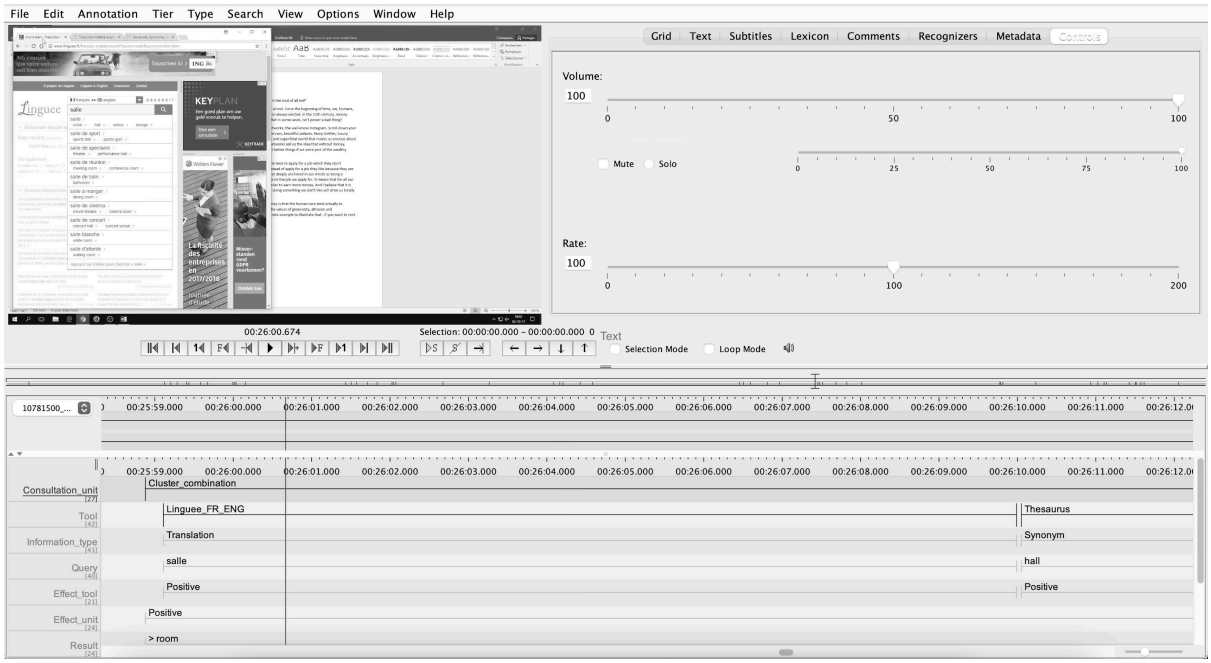


Figure 2. Extract from an annotated PROCEED screencast video in ELAN (Brugman & Russel, 2004)

Table 3 summarizes some of the results of the analysis (more detailed results can be found in Gilquin & Laporte, 2021). It appears that, on average, students spend about 13% of their time using online writing tools and perform 18.4 searches during the whole writing session (lasting about 45 minutes). However, as was the case with the fluency measures in Section 3.1, the SD values point to a high degree of individual variation. Thus, some students carry out almost 50 searches during the writing session (i.e. more than one search per minute), devoting over 25% of their time to this, while other students do not carry out a single search. Table 3 also reveals that, on average, 18% of students' searches involve multiple tools (e.g. searching for the translation of a French word in a French-English bilingual dictionary and then looking up the translation in an English monolingual dictionary). Here again, however, the SD value is quite high. Bilingual dictionaries represent the most often used type of online writing tool, with over 40%. Among searches that are unsuccessful, that is, searches that lead to an incorrect formulation or do not help correct an existing error, a majority concern stylistic issues (37.3%), e.g. register-related errors or unidiomatic language choices.

Variable	**Result**
Average proportion of time spent using online writing tools	13.1% (SD=7.1)
Average number of searches per learner	18.4 (SD=10.9)
Average proportion of multiple-tool searches per learner	18.1% (SD=16.1)
Most often used type of online writing tool	Bilingual dictionaries (40.6%)
Most common type of problem in unsuccessful searches	Stylistic issues (37.3%)

Table 3. Use of online writing tools in L2 PROCEED data (based on Gilquin & Laporte, 2021)

The use of online writing tools can also be correlated with the quality of the finished text. It turns out that two variables correlate positively with the grade obtained by the students in the

PROCEED sample: the proportion of successful searches and the proportion of multiple-tool searches (Gilquin & Laporte, 2021). The former suggests that the number of searches matters less than their efficiency, while the latter shows that students who combine several tools to solve a problem tend to write better texts.

The qualitative analysis of the screencast videos also provides interesting insights into students' use (or non-use) of online writing tools (see Gilquin & Laporte, 2021). Among these insights are the observation that students sometimes pay insufficient attention to detail, with some finding the right information in a dictionary but reproducing it inaccurately in their texts. Students also appear to lack critical thinking (e.g. not checking the context in which a translation is appropriate) and self-awareness (e.g. not realizing that some phrasing in their texts may be problematic and worth checking in a writing tool).

4. Affordances and challenges

The above examples show that keylog files and screencast videos can help gain important insights into L2 writing that are not (or not easily) accessible with more traditional methods. For instance, they reveal what words or word sequences learners tend to produce in one go (bursts of writing) and what effects their use of external sources has on text production. Unlike corpora made up of finished texts, which only reflect the final step of the composition process, keylog files and screencast videos show the whole process, from beginning to end and in real-time. Unlike questionnaires asking writers about their habits during writing, which may elicit writers' perceptions rather than real behaviours, keylog files and screencast videos provide access to what actually happens during the writing process. In addition, they allow for the quantification of various aspects of writing processes, e.g. the average duration of pauses, the percentage of multiword units produced as bursts of writing, or the number of times a writer consults a bilingual dictionary.

Keylog files and screencast videos have been used to observe writing processes and to make inferences about why writers behave in a certain way. However, one should be careful when trying to explain what is observed. As Wengelin et al. (2019, p. 31) point out, "the challenge of interpreting such activities, in the light of more specific cognitive processes, remains". This is why it is often fruitful, as also noted by Wengelin et al. (2019), to combine keylog files and/or screencast videos with other types of data, such as (concurrent or retrospective) think-aloud protocols and eye-tracking. Despite their limitations (see Section 3.3), think-aloud protocols help discover writers' motivations for doing (or not doing) something. One may, for example, find out what the writer was thinking about during a pause or what s/he was planning to write in the case of an incomplete sentence. As for eye-tracking, it provides information about what the writer is looking at, which may be useful for interpreting writing processes, especially when the writer is not typing. It makes it possible, for instance, to check what part of the text the writer is reviewing, whether s/he has looked at an incorrect item (and if so, how long and how many times), or what elements in a dictionary entry s/he has noticed.

5. Conclusion and future directions

Manchón (2018, p. 11) describes "technology and writing" as one of the most promising avenues to be explored in the field of second and foreign language writing. While technology can take many different forms and be used for a variety of purposes, it is expected that

keylogging and screencasting will be increasingly recognized as valuable resources in the study of L2 writing. More generally, it is hoped that writing processes will become more widely accepted as an inherent component of writing. This, incidentally, concerns both typing, which has been the focus of this chapter, and handwriting, which can also be observed by means of special software (e.g. HandSpy; Alves et al., 2019) and is particularly relevant for younger learners who may still engage in more handwriting than typing.

Among the future directions to be considered for the application of keylogging and screencasting to the study of L2, one that comes immediately to mind is language assessment. Typically, assessment of L2 writing has relied exclusively on the final product, with no consideration for the steps leading to this product. Yet, studies that have looked into possible links between writing processes and text quality, including under assessment conditions (see, e.g., Cao et al., 2020), suggest that some features of writing processes could be used to predict text quality. This could open the door to partial automatization of L2 writing assessment through the identification of writing process features in keylogging data.

Going one step further, writing process data could contribute to automatic text classification. Error patterns in (finished) L2 texts have been shown to be reliable clues for the automatic identification of learners' L1 (cf. Bestgen et al., 2012). It can be hypothesized that errors of the type that are immediately corrected (typos) and are only visible in writing process data could help refine L1 identification. The features to be examined could possibly be extended beyond typos, including other characteristics of writing processes, based on the principle of keystroke dynamics that "[t]ypists have unique typing patterns" (Raul et al., 2020, p. 149) and applying this principle to groups of writers rather than individual writers.

These applications, and those described in Section 3, are just some illustrations of how keylogging and screencasting can enrich our understanding of L2 writing. These, hopefully, will be enough to demonstrate the relevance of adding these technologies to one's toolbox.

References

Alves, R. A., Leal, J. P., & Limpo, T. (2019). Using HandSpy to study writing in real time: A comparison between low- and high-quality texts in grade 2. In E. Lindgren & K. P. H. Sullivan (Eds.), *Observing writing: Insights from keystroke logging and handwriting* (pp. 50–70). Brill.

Barkaoui, K. (2019) What can L2 writers' pausing behavior tell us about their L2 writing processes? *Studies in Second Language Acquisition*, *41*(3), 529–554.

Béjoint, H. (1981). The foreign student's use of monolingual English dictionaries: A study of language needs and reference skills. *Applied Linguistics*, *2*(3), 207–222.

Bestgen, Y., Granger, S., & Thewissen, J. (2012). Error patterns and automatic L1 identification. In S. Jarvis & S. A. Crossley (Eds.), *Approaching language transfer through text classification: Explorations in the detection-based approach* (pp. 127–153). Multilingual Matters.

Biber, D., Conrad, S., & Cortes, V. (2004). *If you look at ...*: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371–405.

Brugman, H., & Russel, A. (2004). Annotating multimedia/multi-modal resources with ELAN. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of the fourth international conference on language resources and evaluation (LREC'04)* (pp. 2065–2068). European Language Resources Association.

Cao, Y., Chen, J., Zhang, M., & Li, C. (2020). *Examining the writing processes in scenario-based assessment using regression trees. ETS Research Report No. RR-20-18*. Educational Testing Service.
Celce-Murcia, M. (2001). *Teaching English as a second or foreign language* (3rd ed.). Heinle & Heinle.
Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication*, *18*(1), 80–98.
Chon, Y. V. (2009). The electronic dictionary for writing: A solution or a problem? *International Journal of Lexicography*, *22*(1), 23–54.
Cislaru, G., & Olive, T. (2017). Segments répétés, jets textuels et autres routines. Quel niveau de pré-construction ? *Corpus*, *17*, 1–21.
Dahlmann, I., & Adolphs, S. (2007). Pauses as an indicator of psycholinguistically valid multi-word expressions (MWEs)? In N. Grégoire, S. Evert, & S. N. Kim (Eds.), *Proceedings of the workshop on a broader perspective on multiword expressions* (pp. 49–56). Association for Computational Linguistics.
Dechert, H. W. (1983). How a story is done in a second language. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 175–195). Longman.
Ellis, R. & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*(1), 59–84.
Gilquin, G. (2022). The *Process Corpus of English in Education*: Going beyond the written text. *Research in Corpus Linguistics*, *10*(1), 31–44.
Gilquin, G. (2024). The processing of multiword units by learners of English: Evidence from pause placement in writing process data. *Languages*, *9*(2), 1–22.
Gilquin, G. & Laporte, S. (2021). The use of online writing tools by learners of English: Evidence from a process corpus. *International Journal of Lexicography*, *34*(4), 472–492.
Granger, S., Gilquin, G., & Meunier, F. (Eds.). (2015). *The Cambridge handbook of learner corpus research*. Cambridge University Press.
Hacker, D. J., Keener, M. C., & Kircher, J. C. (2017). TRAKTEXT: Investigating writing processes using eye-tracking technology. *Methodological Innovations*, *10*(2), 1–18.
Hamel, M.-J. (2012). Testing aspects of the usability of an online learner dictionary prototype: A product- and process-oriented study. *Computer Assisted Language Learning*, *25*(4), 339–365.
Hatherall, G. (1984). Studying dictionary use: Some findings and proposals. In R. R. K. Hartmann (Ed.), *LEXeter '83 Proceedings: Papers from the international conference on lexicography at Exeter, 9-12 September 1983* (pp. 183–189). Max Niemeyer.
Janssen, D., van Waes, L., & van den Bergh, H. (1996). Effects of thinking aloud on writing processes. In C. M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 233–250). Erlbaum.
Knežević, L., Halupka-Rešetar, S., Miškeljin, I., & Milić, M. (2021). Millennials as dictionary users: A study of dictionary use habits of Serbian EFL students. *SAGE Open*, *11*(2), 1–11.
Laufer, B., & Levitzky-Aviad, T. (2006). Examining the effectiveness of 'Bilingual Dictionary Plus' — A dictionary for production in a foreign language. *International Journal of Lexicography*, *19*(2), 135–155.
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.
Leijten, M., & Van Waes, L. (2020). Designing keystroke logging research in writing studies. *Chinese Journal of Second Language Writing*, *1*(1), 18–39.
Leijten, M., Van Waes, L., Schrijver, I., Bernolet, S., & Vangehuchten, L. (2019) Mapping master's students' use of external sources in source-based writing in L1 and L2. *Studies in Second Language Acquisition*, *41*(3), 555–582.

Manchón, R. M. (2018). Introduction: Past and future of L2 writing research. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 1–15). de Gruyter.

Pawley, A. (1986). Lexicalization. In D. Tannen & J. E. Alatis (Eds.), *Language and linguistics: The interdependence of theory, data, and application* (pp. 98–120). Georgetown University Press.

Raul, N., Shankarmani, R., & Joshi, P. (2020). A comprehensive review of keystroke dynamics-based authentication mechanism. In A. Khanna, D. Gupta, S. Bhattacharyya, V. Snasel, J. Platos, & A. E. Hassanien (Eds.), *International conference on innovative computing and communications. Proceedings of ICICC 2019, Volume 2* (pp. 149–162). Springer.

Roca De Larios, J., Murphy, L., & Marin, J. (2002). A critical examination of L2 writing process research. In S. Ransdell & M.-L. Barbier (Eds.), *New directions for research in L2 writing* (pp. 11–47). Springer.

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, *9*(3), 259–291.

Schilperoord, J. (2002). On the cognitive status of pauses in discourse production. In T. Olive & C. M. Levy (Eds.), *Contemporary tools and techniques for studying writing* (pp. 61–87). Springer.

Schmidt, T., & Wörner, K. (2014). EXMARaLDA. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford handbook on corpus phonology* (pp. 402–419). Oxford University Press.

Schmitt, N. (Ed.). (2004). *Formulaic sequences. Acquisition, processing and use*. John Benjamins.

Seow, A. (2002). The writing process and process writing. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 315–320). Cambridge University Press.

Siahaan, C. R. P., & Chowanda, A. (2022). Spoofing keystroke dynamics authentication through synthetic typing pattern extracted from screen-recorded video. *Journal of Big Data*, *9*, 111.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., & van Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language*, *175*, 111–122.

Spelman Miller, K. (2000). Academic writers on-line: Investigating pausing in the production of text. *Language Teaching Research*, *4*(2), 123–148.

Spelman Miller, K., Lindgren, E., & Sullivan, K. P. H. (2008). The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging. *TESOL Quarterly*, *42*(3), 433–454.

Strömqvist, S., & Malmsten, L. (1997). Scriptlog Pro 1.04—User's manual. Department of Linguistics, University of Göteborg.

Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Computers and Composition*, *21*(2), 217–235.

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

Van Waes, L., Leijten, M., Roeser, J., Olive, T., & Grabowski, J. (2021). Designing a copy task to measure and assess typing skills in writing research. *Journal of Writing research*, *13*(1), 107–153.

Wahl, A. (2015). Intonation unit boundaries and the storage of bigrams: Evidence from bidirectional and directional association measures. *Review of Cognitive Linguistics*, *13*(1), 191–219.

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. P. H. Sullivan (Eds.), *Observing writing: Insights from keystroke logging and handwriting* (pp. 30–49). Brill.

Xu, C., & Ding, Y. (2014). An exploratory study of pauses in computer-assisted EFL writing. *Language Learning & Technology*, *18*(3), 80–96.

Xu, C., & Qi, Y. (2017). Analyzing pauses in computer-assisted EFL writing — A computer-keystroke-log perspective. *Journal of Educational Technology & Society*, *20*(4), 24–34.