RiCL Research in Corpus Linguistics

The *Process Corpus of English in Education*: Going beyond the written text

Gaëtanelle Gilquin
Université catholique de Louvain / Belgium

Abstract – The *Process Corpus of English in Education* (PROCEED) is a learner corpus of English which, in addition to written texts, consists of data that make the writing process visible in the form of keystroke log files and screencast videos. It comes with rich metadata about each learner, among which indices of exposure to the target language and cognitive measures such as working memory or fluid intelligence. It also includes an L1 component which is made up of similar data produced by the learners in their mother tongue. PROCEED opens new perspectives in the study of learner writing, by going beyond the written product. It makes it possible to investigate aspects such as writing fluency, use of online resources, cognitive phenomena like automaticity and avoidance, or theoretical modelling of the writing process. It also has applications for teaching, e.g. by showing students screencast video clips from the corpus illustrating effective writing strategies, as well as for testing, e.g. by establishing a corpus-derived standard of writing fluency for learners at a certain proficiency level.

Keywords – Learner corpus research; process learner corpus; writing process; keylogging; screencasting; metadata

1. Introduction: From written product to writing process

The first electronic corpus ever, the *Brown Corpus*, was a corpus of written English. Since then, many corpora have been collected that represent written language. Among learner corpora, i.e. corpora consisting of language produced by foreign or second language (L2) learners, 64 per cent are made up of written texts only (and 12% of both written texts and spoken transcripts) according to the current version of the *Learner Corpora around the World* list maintained by the *Centre for English Corpus Linguistics* (2020). Examples of written learner corpora include the *International Corpus of Learner English*, the *Longman Learners' Corpus*, the *International Corpus of Crosslinguistic Interlanguage* or the *Written Corpus of Learner English*. These and other written corpora have yielded invaluable insights into writing: its lexico-

grammatical features, the way sentences and paragraphs are organised, how genres can be characterised linguistically, what errors writers tend to make, etc.

What these corpora give access to is the written product, that is, the final output of the writing act. Most written texts, however, go through several stages of editing and revision before they reach the final stage, when the text is offered to the reader. These intermediate states of the text are lost in a typical written corpus. The aim of the resource that is introduced in this article, the *PROcess Corpus of English in EDucation* (PROCEED),[1] is to make the whole writing process visible. To illustrate the difference between written product and writing process, one can consider example (1), a sentence taken from PROCEED and produced by a French-speaking learner of English. This sentence is the result of as many as twenty-eight different stages, as visible in PROCEED and as represented in (2), where strikethrough indicates text that has been deleted and the grey font shows a word in which one or several letters have been inserted.

(1) Our actual society is dominted by technology and science. A lot of experiments concentrate lately on the effects of those new developments on the human being.

(2) a. In
b. ~~In~~
c. Our moder
d. ~~Our moder~~
e. It is o
f. ~~o~~
g. nowadays a fact: our modern society is dominated by
h. ~~It is nowadays a fact: our modern society is dominated by~~
i. Our actual society is dominted by num
j. ~~num~~
k. techonology and scin
l. ~~n~~
m. encee
n. ~~e~~
o. . A lot of experiments concetrate lately to the effect of those
p. concentrate
q. ~~to~~
r. on
s. effects
t. ~~of those~~
u. that it could cause for the human beings.
v. ~~that it could cause for the human beings.~~

[1] See https://uclouvain.be/en/research-institutes/ilc/cecl/proceed.html (8 March, 2021.)

w. that
x. ~~that~~
y. of those new devem
z. ~~m~~
aa. lopment on the human being.
ab. developments

The intermediate stages reveal, among others, errors that have been corrected, for example *concentrate to* (2q) which has been replaced by *concentrate on* (2r), but also, more surprisingly, correct phrases that have been replaced by incorrect ones, as appears from the transformation of *modern society* into *actual society* (2h-i), with the use of a false friend (in French *actuel* means 'current'). Although (2g) is not kept in the finished text, it is interesting because it includes the correct form *dominated*, which suggests that *dominted*, used in the final version of the sentence, is probably only a typo, since the learner is clearly able to spell the word correctly. What has not been represented in (2), but is visible in the PROCEED data, is the fact that the learner has paused on several occasions while typing this sentence. For example, in (2u), there is a long pause of 23 seconds just after *that*, which may be indicative of the learner's difficulty in finishing the sentence. There is also a seven-second pause before the insertion of the *s*-letter at the end of *development* (2ab), which seems to correspond to a reviewing of the whole sentence, resulting in a last correction.

This example is an illustration of Murray's (1980: 3) witty remark that "process can not be inferred from product any more than a pig can be inferred from a sausage." It also points to the importance of considering the writing process next to the written product. Indeed, there have been calls in the literature to pay attention to the writing process. Back in the 1980s, Hairston (1982: 84) thus claimed that

> we have to try to understand what goes on during the act of writing [...] if we want to affect its outcome. We have to do the hard thing, examine the intangible process, rather than the easy thing, evaluate the tangible product.

The use of computers as well as technologies like screencasting (recording of the screen activity) and keylogging (recording of the keys struck on the keyboard) have made the intangible more tangible: it is now possible to see the writing process unfold before our eyes, with all its deletions, insertions, substitutions, pauses, etc. Several recent studies have relied on such information to approach writing and have demonstrated its

usefulness for descriptive, theoretical or pedagogical purposes (see, e.g., Cislaru 2015; Lindgren and Sullivan 2019; Révész and Michel 2019).

Among the studies that use writing process data, the setting tends to be experimental, with data being collected specifically for this particular study, often among a small group of participants. In Breuer (2019), for example, the keystroke log files produced by 10 German students writing three texts in English and two in German are used to investigate the students' fluency in L1 (mother tongue) and L2 writing, revealing a higher degree of fluency in L1 than in L2 for most students. Sullivan and Lindgren (2002) test the pedagogical use of keystroke log files among four learners of English required to write a narrative text and demonstrate the positive effect of observing one's own composing process. In Elola and Mikulski (2016), a comparison is drawn between the screen activity of six learners of Spanish as a foreign language and 12 learners of Spanish as a heritage language, which brings to light similarities between the two groups (e.g. transfer of writing processes from the L1) as well as differences (e.g. more surface revisions but fewer meaning revisions in Spanish as a foreign language). The term *corpus* is hardly ever used in such studies, which may suggest that the data are not meant as a durable and reusable resource. A notable exception is Wengelin (2006), who describes her data sets, consisting of keystroke log files for Swedish texts, as corpora. Moreover, she shows how the techniques of corpus linguistics can be applied to the study of pauses in writing by looking for 'micro-contexts' made up of a pause preceded and followed by certain elements (e.g. a pause preceded by a typed letter and followed by a deletion). Cislaru and Olive (2018) similarly refer to their process data (different versions of texts in French, together with the keystroke log files) as a corpus. In addition, they explicitly mention corpus linguistics as one of the frameworks they draw inspiration from. Hamel and Séror (2016: 156) also use the term *corpus* to describe a collection of screencast videos showing the writing process of L2 learners of French and English. They point out that

> such corpora represent new and exciting forms of empirical data which, once anonymized, could contribute to learner corpus projects that might be shared with others.

The *Process Corpus of English in Education* (PROCEED), as its name indicates, was designed as a corpus right from the start, meant as a durable and reusable resource bringing together a substantial amount of data supposed to be representative of a larger population. It relies on both keylogging and screencasting. It also comes with rich

metadata and comparable data in the learners' L1. The resource is described in more detail in Section 2, while Section 3 provides an overview of some of the research perspectives that the corpus offers. Section 4 concludes the article.

2. THE CORPUS

2.1. A project in learner corpus research

PROCEED can be described as a new type of learner corpus in the typology of learner corpora (cf. Gilquin 2015), namely a 'process learner corpus', which shows the process through which a text is composed on computer by language learners. It makes the writing process visible through keylogging and screencasting, two complementary methods to record the activity of writing a text on computer. The corpus aims to contribute to learner corpus research by providing a resource that allows for a novel and fine-grained approach to written performance, in the original sense of 'performance', that is, the *process* of doing something (in this case, writing a text).

The corpus project started in February 2017 with the collection of writing process data among a group of higher intermediate to advanced, mostly French-speaking students majoring in English at the University of Louvain (Belgium). Since then, additional data have been collected at least once a year among a new cohort of students each year. This is seen as the first step towards setting up an international project that seeks to collect similar data in other countries, among learners of English with different mother tongue backgrounds.

2.2. The data

Like traditional written learner corpora, PROCEED includes texts written by learners. These learner texts are written in English and are of the argumentative type, as this genre is thought to involve more complex writing processes than other text types like narrative texts (cf. Roca de Larios *et al.* 2002). Each learner begins by choosing a topic or quote among several options offered. They then have about 45 minutes to write a text of approximately 350 words defending their point of view. They are allowed to use online reference tools but are asked not to draw on secondary sources. These texts represent the written product.

In addition to the written product, the corpus includes writing process data. With the learners' permission, the keys struck on the keyboard are recorded by means of *Inputlog* (Leijten and Van Waes 2013) and the screen activity is recorded by means of *OBS Studio* during the whole writing task.[2] The *Inputlog* data take the form of log files, one per text, representing the different actions performed (letters typed, deletion, capitalisation, mouse click or movement, pauses, transition between Word document and other windows, etc.). These files can serve as a basis to carry out different types of analyses and to compute various statistics within *Inputlog*, e.g. linear analysis (with one action per line), revision matrix (a list of all the revisions), writing time, pausing time or number of revisions. Since they involve textual/numerical data, they can be searched by means of the techniques of corpus linguistics, although with adapted queries (cf. Wengelin's (2006) study, mentioned in Section 1). *Inputlog* has a replay function, which makes it possible to reconstruct the writing process in a video-like manner on the basis of the stored data. However, the function comes with a warning that an error-free replay of the process files cannot be guaranteed and with a recommendation for researchers relying on replay to resort to screencasting.

Screencasting with *OBS Studio* produces a faithful representation of the screen activity during the writing task. The *OBS Studio* data take the form of screencast videos, one per text. The videos can be navigated easily, and played at different speeds, using any multimedia player. While videos as such cannot be queried directly with the usual tools and techniques of corpus linguistics, they may be amenable to queries via alignment with the keystroke log files or via annotation. The *OBS Studio* videos can be aligned with the *Inputlog* data thanks to the video timeline and the timestamps associated with each action in *Inputlog*. A search on the textual data from *Inputlog* with text retrieval software could therefore generate hits from the *Inputlog* file that are linked to the corresponding part of the video. Annotation is another way of querying the screencast videos. A program like *ELAN* (Wittenburg *et al.* 2006) makes it possible to annotate videos with written information that describes their contents, by inserting annotation tiers which include attributes assigned to specific video segments (e.g. segments without any typing or involving the use of an online dictionary; see Laporte and Gilquin 2018 for an illustration). The information provided in the annotation can then be searched by means of text retrieval software.

[2] https://obsproject.com (12 March, 2021.)

Although PROCEED is first and foremost a learner corpus, consisting of non-native data produced by language learners, it was deemed relevant to include L1 data representing the learners' writing process in their mother tongue. This is because writing processes are said to display "conspicuous individual differences" (Sasaki 2000: 262), which may partly be the result of idiosyncratic behaviours that are language-independent, and hence valid regardless of whether the writer is writing in their L1 or in an L2. Comparing writers' behaviours in L1 and L2 is not only intrinsically interesting (cf. Thorson 2000; Stevenson *et al.* 2006), but it can also help distinguish these language-independent features from those that are due to the non-native nature of the writing process. The L1 data are collected according to the same principles as the L2 data: the learners have about 45 minutes to write a 350-word argumentative text on one of several set topics/quotes, while their screen and keyboard activity is recorded with their permission.

2.3. The metadata

As is the case with most learner corpora, PROCEED comes with rich metadata describing learners' profiles and collected via a questionnaire to be filled in by each participant. It includes personal information (age, gender, nationality, country of residence, etc.) as well as information about the learner's use and knowledge of languages (native language, parents' native languages, language(s) used in everyday life, language(s) of instruction at school, knowledge of foreign languages, etc.). Particular attention is paid to learners' exposure to English (number of years of English at school/university, proportion of classes taught in English, time spent in an English-speaking country, varieties of English they have been exposed to, etc.) as well as the kind of contexts in which they use English (estimation of the time spent doing certain activities, such as reading, watching TV or doing homework, in English). Learners are also asked to evaluate their (speaking, writing, listening, reading, pronunciation, grammar and spelling) skills in English. This comes as a complement to their score on the *LexTALE* vocabulary test, which has been shown to correlate with general tests of English proficiency (Lemhöfer and Broersma 2012). Finally, the questionnaire includes a few questions that are specifically related to the kind of corpus data collected, such as the type of keyboard learners usually use or whether they have been diagnosed with dyslexia.

Because typing speed is essential when considering aspects of the writing process such as fluency, learners are required to carry out a copy task, both in English and in their L1. The copy task was designed by the developers of *Inputlog*, within which the results of the task can be analysed. It can be done online, with the output file being directly downloadable from the website.[3] It involves several activities: pressing two keys one after the other as quickly as possible, copying a sentence as many times as possible, copying combinations of three words and copying blocks of consonants.

The analysis of writing process data can provide insights into more cognitive aspects of language performance (cf. Section 3.1). For this reason, the PROCEED metadata also include measures of learners' cognitive abilities, which can be related to the writing process data and possibly account for some of the individual variation. These measures are collected by subjecting the participants to a battery of tests. Learners' verbal aptitudes (including vocabulary learning and grammatical inferencing) are tested through some of the *LLAMA* (language-independent) tests (Meara and Rogers 2019). Their non-verbal aptitudes are tested by means of *Raven's Matrices* (Raven and Raven 2003), which measure abstract reasoning (fluid intelligence). In addition, the *Psychology Experiment Building Language* (PEBL) interface (Mueller 2012) is used to assess working memory capacity (by means of the *Operation Span* task; cf. Hegarty and Dufflecoat Enterprises 2014) as well as response inhibition and interference suppression (by means of the *Flanker* and *Simon* tasks; cf. Mueller 2011a, 2011b).

3. Research and pedagogical perspectives

3.1. Writing process research

Besides the kind of research that is traditionally possible on the basis of written learner corpora, the PROCEED data have great potential for research into the writing process. By combining keylogging and screencasting, they present an accurate picture of the way learners of English compose their texts, with unprecedented detail on the actual mechanics of the process. This information can be used for descriptive, explanatory and theoretical purposes.

In terms of description, the keylogging data provide comprehensive statistics about aspects that have to do with writing fluency (number, duration and location of

[3] http://inputlog.ua.ac.be/Website/copytask/tasks.html (8 March, 2021.)

pauses, type and number of revisions, etc.). As against the conventional approaches that measure fluency as the number of words produced overall or the mean number of words produced per minute (cf. Sasaki 2004), the keylogging-based approach considers writing fluency in its multidimensionality (cf. Van Waes and Leijten 2015). This focus on the notion of fluency also opens up new possibilities for comparing learner writing and speech. In addition, keylogging and screencasting data make it possible to examine the use of online resources during the writing process, such as secondary sources (Leijten *et al.* 2019) or writing tools (Gilquin and Laporte forthcoming, based on the annotation of PROCEED videos with *ELAN*). The data could also be used to carry out a dynamic discourse analysis, looking at how discourse is created in real time (e.g. paragraph formation, development of rhetorical functions) or what strategies learners adopt to compose a text (e.g. linear composition or outline that is progressively fleshed out).

A further use of PROCEED is for explanatory purposes. The writing process data can help account for the origin of certain features of the finished texts. A lack of tense agreement between main clause and subclause, for example, may turn out to be due to the fact that the tense of the main verb was changed at some stage but the writer failed to adapt the tense of the verb in the subclause (cf. Gilquin 2021). The data can also help uncover more cognitive aspects of writing performance (cf. Spelman Miller *et al.* 2008). Revisions may thus point to a lack of automaticity for certain language components (e.g. the subject-verb agreement rule, if the verb form regularly needs to be revised) or to phenomena of avoidance (e.g. avoidance of the passive, if passive structures are systematically aborted), which are typically very difficult to discover on the basis of written texts only. Seeing what words are produced together in one go (the so-called ‘bursts’, see Chenoweth and Hayes 2001) can also give an indication of the constructions that are stored as wholes in the mind (Gilquin 2020).

From a theoretical perspective, writing process data such as those found in PROCEED can help develop or improve models of writing, as shown in Leijten *et al.* (2014) with keylogging data. The design of PROCEED, consisting of texts produced by the same writers in their L1 and in L2 English, could lead to the development of bilingual writing models representing native and non-native writing, and showing how L1 and L2 writing abilities interact with each other. The metadata associated with each writer might even make it possible to adapt a general writing model to individual

variation, most notably through the empirical measures of working memory, which is part and parcel of most writing models (cf. Kellogg 1996; Hayes 2012).

3.2. Teaching and testing applications

Next to its use for research purposes, PROCEED also has potential applications for teaching and testing. The most immediate pedagogical application is to use PROCEED as a local learner corpus, that is, a corpus that is collected by the teacher among —and for the benefit of— his or her own students (Seidlhofer 2002). In other words, the learners are both contributors to and users of PROCEED. After collecting data from a group of learners, they can each be given access to their screencast video and be required to watch (part of) it, so as to become aware of how they actually compose a text. Additionally, clips from some learners' videos can be selected and shown to the members of the group, to illustrate effective strategies that could be useful to them (e.g. highlighting words to be checked later in a dictionary, so that the flow of ideas does not get interrupted). Learners can also be presented with some statistics describing their writing behaviour. On the basis of a keystroke log file, *Inputlog* can generate a user report that summarises some important facts about the user's writing process, such as the time they have been writing vs. pausing or the number of revisions they have made (Vandermeulen *et al.* 2020). The report also includes a graph representing the writing process which, with some explanations, could help learners visualise their own writing behaviour, and possibly compare it with the behaviour of other learners in the group or that of native writers (see Gilquin 2019 for a pedagogical intervention based on PROCEED as a local learner corpus). The PROCEED data can also be used as pedagogical materials for learners other than those among whom the data were collected. Video clips illustrating different writing strategies (effective or less effective) could be shown to learners to help them reflect on the act of writing and how best to compose a text. The process graphs generated by *Inputlog* could also be used as a basis to exemplify various writing behaviours (e.g. revising the text as one goes along or leaving some time at the end to revise the whole of it).

The writing process data from PROCEED can also serve testing purposes. While the testing of writing skills typically only relies on the quality assessment of the finished text, considering the writing process too could result in a more fine-grained evaluation of writing performance (cf. Ranalli *et al.* 2018). Thus, it would make sense, as is the

case for speech, to include a criterion like writing fluency, which would aim to assess how smooth the writing process is. The PROCEED data, and in particular the analysis of the keystroke log files, could provide the necessary statistics to empirically assess the writing fluency of the learners who contributed to the corpus. Their writing fluency in the mother tongue could even be taken into account to provide a tailor-made yardstick for each learner. Another aspect that could be relevant to the evaluation of writing skills is consultation behaviour, that is, the way in which learners resort to online writing tools like dictionaries or thesauri, as using these tools effectively may be seen as an important component of writing performance. Again, this can be examined empirically for the contributors to the corpus, using the screencast videos. The analysis of such aspects of the writing process in PROCEED could also help improve writing assessment on a more general level, for other learners than those who contributed to the corpus. By bringing together data from a large number of participants, PROCEED can be said to be representative of a certain population of learners. It can therefore be exploited to determine the typical writing behaviour of learners at a given proficiency level, for example in terms of pausing time or number of revisions, and to set this as the expected standard. Other learners with a similar profile can then be evaluated against this corpus-derived standard.

4. Conclusion

This article has introduced a new resource, PROCEED, which also represents a new type of corpus to investigate learner writing. Its unique combination of written texts, screencast videos, keystroke log files, rich metadata including cognitive measures, and equivalent L1 data offers an unparalleled opportunity to study the process through which learners write texts. It also opens new perspectives in terms of research and applications: study of writing fluency and comparison with spoken fluency; analysis of learners' use of online writing tools; dynamic discourse analysis taking the development of discourse into account; exploration of cognitive aspects of writing performance; theoretical modelling of the bilingual writing process; pedagogical interventions involving learners' examination of their own writing behaviour; addition of a 'process' component to the assessment of writing skills, based on corpus-derived standards; etc.

While collecting and analysing corpus data of the PROCEED type implies different routines than those followed in traditional learner corpus research, this

description of the PROCEED project will hopefully have demonstrated the value of what could be referred to as 'process learner corpus research', and the significance of its possible applications. The potential of PROCEED will arguably continue to increase as the corpus keeps growing in size and in diversity of learner profiles.

REFERENCES

Breuer, Esther Odilia. 2019. Fluency in L1 and FL writing: An analysis of planning, essay writing and final revision. In Eva Lindgren and Kirk P. H. Sullivan eds. *Observing Writing: Insights from Keystroke Logging and Handwriting*. Leiden: Brill, 190–211.

Centre for English Corpus Linguistics. 2020. Learner corpora around the world. Louvain-la-Neuve: Université catholique de Louvain. https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html (23 December, 2020.)

Chenoweth, N. Ann and John R. Hayes. 2001. Fluency in writing: Generating text in L1 and L2. *Written Communication* 18/1: 80–98.

Cislaru, Georgeta ed. 2015. *Writing(s) at the Crossroads: The Process–Product Interface*. Amsterdam: John Benjamins.

Cislaru, Georgeta and Thierry Olive. 2018. *Le processus de textualisation. Analyse des unités linguistiques de performance écrite*. Louvain-la-Neuve: De Boeck Supérieur.

Elola, Idoia and Ariana M. Mikulski. 2016. Similar and/or different writing processes? A study of Spanish foreign language and heritage language learners. *Hispania* 99/1: 87–102.

Gilquin, Gaëtanelle. 2015. From design to collection of learner corpora. In Sylviane Granger, Gaëtanelle Gilquin and Fanny Meunier eds. *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 9–34.

Gilquin, Gaëtanelle. 2019. Screencasting and keylogging as pedagogical tools to enhance writing skill development. Paper presented at *EUROCALL 2019, Louvain-la-Neuve, Belgium, 28–31 August 2019*.

Gilquin, Gaëtanelle. 2020. In search of constructions in writing process data. *Belgian Journal of Linguistics* 34: 99–109.

Gilquin, Gaëtanelle. 2021. Hic sunt dracones: Exploring some *terra incognita* in learner corpus research. In Anna Čermáková and Marketa Malá eds. *Variation in Time and Space: Observing the World through Corpora*. Berlin: De Gruyter, 65–86.

Gilquin, Gaëtanelle and Samantha Laporte. Forthcoming. The use of online writing tools by learners of English: Evidence from a process corpus. *International Journal of Lexicography*.

Hairston, Maxine. 1982. The winds of change: Thomas Kuhn and the revolution in the teaching of writing. *College Composition and Communication* 33/1: 76–88.

Hamel, Marie-Josée and Jérémie Séror. 2016. Video screen capture to document and scaffold the L2 writing process. In Catherine Caws and Marie-Josée Hamel eds. *Language-Learner Computer Interactions: Theory, Methodology and CALL Applications*. Amsterdam: John Benjamins, 137–162.

Hayes, John R. 2012. Modeling and remodeling writing. *Written Communication* 29/3: 369–388.

Hegarty, David L. and Dufflecoat Enterprises. 2014. *The PEBL Operation Span Task*. http://pebl.sourceforge.net/battery.html (12 March, 2021.)

Kellogg, Ronald T. 1996. A model of working memory in writing. In C. Michael Levy and Sarah Ransdell eds. *The Science of Writing: Theories, Methods, Individual Differences, and Applications*. Mahwah, NJ: Erlbaum, 57–71.

Laporte, Samantha and Gaëtanelle Gilquin. 2018. Annotating the use of online writing resources in a video corpus of written process data in ELAN. Annotation manual version 1.1. CECL Papers 2. Louvain-la-Neuve: Université catholique de Louvain. https://uclouvain.be/en/research-institutes/ilc/cecl/cecl-papers.html (8 March, 2021.)

Leijten, Mariëlle and Luuk Van Waes. 2013. Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication* 30/3: 358–392.

Leijten, Mariëlle, Luuk Van Waes, Iris Schrijver, Sarah Bernolet and Lieve Vangehuchten. 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.

Leijten, Mariëlle, Luuk Van Waes, Karen Schriver and John R. Hayes. 2014. Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research* 5/3: 285–337.

Lemhöfer, Kristin and Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods* 44/2: 325–343.

Lindgren, Eva and Kirk P. H. Sullivan eds. 2019. *Observing Writing: Insights from Keystroke Logging and Handwriting*. Leiden: Brill.

Meara, Paul M. and Vivienne E. Rogers. 2019. *The LLAMA Tests v3*. Cardiff: Lognostics.

Mueller, Shane T. 2011a. *The PEBL Flanker Task*. http://pebl.sourceforge.net/battery.html (12 March, 2021.)

Mueller, Shane T. 2011b. *The PEBL Simon Interference Task*. http://pebl.sourceforge.net/battery.html (12 March, 2021.)

Mueller, Shane T. 2012. The Psychology Experiment Building Language, Version 0.13. http://pebl.sourceforge.net (23 December, 2020.)

Murray, Donald M. 1980. Writing as process: How writing finds its own meaning. In Timothy R. Donovan and Ben W. McClelland eds. *Eight Approaches to Teaching Composition*. Urbana, IL: National Council of Teachers of English, 3–20.

Ranalli, Jim, Hui-Hsien Feng and Evgeny Chukharev-Hudilainen. 2018. Exploring the potential of process-tracing technologies to support assessment for learning of L2 writing. *Assessing Writing* 36: 77–89.

Raven, John and Jean Raven. 2003. Raven Progressive Matrices. In R. Steve McCallum ed. *Handbook of Nonverbal Assessment*. Boston, MA: Springer, 223–237.

Révész, Andrea and Marije Michel. 2019. State of the scholarship: Introduction. *Studies in Second Language Acquisition* 41/3: 491–501.

Roca de Larios, Julio, Liz Murphy and Javier Marín. 2002. A critical examination of L2 writing process research. In Sarah Ransdell and Marie-Laure Barbier eds. *New Directions for Research in L2 Writing*. Dordrecht: Kluwer Academic Publishers, 11–47.

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

Sasaki, Miyuki. 2004. A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning* 54/3: 525–582.

Seidlhofer, Barbara. 2002. Pedagogy and local learner corpora: Working with learning-driven data. In Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson eds. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 213–234.

Spelman Miller, Kristyan, Eva Lindgren and Kirk P. H. Sullivan. 2008. The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging. *TESOL Quarterly* 42/3: 433–454.

Stevenson, Marie, Rob Schoonen and Kees de Glopper. 2006. Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing* 15/3: 201–233.

Sullivan, Kirk and Eva Lindgren. 2002. Self-assessment in autonomous computer-aided second language writing. *ELT Journal* 56/3: 258–266.

Thorson, Helga. 2000. Using the computer to compare foreign and native language writing processes: A statistical and case study approach. *Modern Language Journal* 84/2: 155–170.

Vandermeulen, Nina, Mariëlle Leijten and Luuk Van Waes. 2020. Reporting writing process feedback in the classroom: Using keystroke logging data to reflect on writing processes. *Journal of Writing Research* 12/1: 109–139.

Van Waes, Luuk and Mariëlle Leijten. 2015. Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition* 38/A: 79–95.

Wengelin, Åsa. 2006. Examining pauses in writing: Theory, methods and empirical data. In Kirk P. H. Sullivan and Eva Lindgren eds. *Computer Keystroke Logging and Writing: Methods and Applications*. Amsterdam: Elsevier, 107–130.

Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann and Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, 1556–1559.

Corresponding author
Gaëtanelle Gilquin
Université catholique de Louvain
Collège Erasme
Place Blaise Pascal 1, bte L3.03.33
B-1348 Louvain-la-Neuve
Belgium
Email: gaetanelle.gilquin@uclouvain.be

received: December 2020
accepted: March 2021
published online: April 2021