Combined Deployable Keystroke Logging and Eyetracking for Investigating L2 Writing Fluency

SCHOLARONE™ Manuscripts

Cambridge University Press

Combined Deployable Keystroke Logging and Eyetracking

for Investigating L2 Writing Fluency

Evgeny Chukharev-Hudilainen

Department of English, Iowa State University, USA

Aysel Saricaoglu

Department of English Language and Literature, Social Sciences University of Ankara, Turkey

Mark Torrance

Department of Psychology, Nottingham Trent University, UK

Hui-Hsien Feng

Department of English, National Kaohsiung University of Science and Technology, Taiwan

Abstract

Although fluency is an important sub-construct of language proficiency, it has not received as much attention in L2 writing research as complexity and accuracy have, in part due to the lack of methodological approaches for the analysis of large datasets of writing-process data. This article presents a method of time-aligned keystroke logging and eye tracking and reports an empirical study investigating L2 writing fluency through this method. Twenty-four undergraduate students at a private university in Turkey performed two writing tasks delivered through a web text editor with embedded keystroke logging and eye-tracking capabilities. Linear mixed-effects models were fit to predict indices of pausing and reading behaviors based on language status (L1 vs. L2) and linguistic context factors. Findings revealed differences between pausing and eye-fixation behavior in L1 and L2 writing processes. The paper concludes by discussing the affordances of the proposed method from the theoretical and practical standpoints.

Keywords: L2 writing fluency, writing processes, keystroke logging, eye tracking

Introduction

According to a consensus in second language acquisition research, language proficiency can be captured by the constructs of accuracy (i.e., lack of errors), complexity (i.e., being able to produce adequately complex linguistic structures), and fluency (Pallotti, 2009; Wolfe-Quintero, Inagaki, & Kim, 1998). While fluency is central to studies of L2 speaking proficiency (de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012; Derwing, Munro, & Thomson, 2008; Lennon, 1990; Révész, Ekiert, & Torgersen, 2016), in L2 writing research fluency is frequently ignored. When it is considered, it is often measured across whole writing sessions as the number of words written per minute (Baba & Nitta, 2014; Chenoweth & Hayes, 2001). Such global measures of writing fluency, however, lack both explanatory power and instructional value. Knowing that a student writes slowly overall is less useful, from both the research and the practical perspective, than being able to identify what specifically it is that the student struggles with when a drop in production fluency occurs. To draw a parallel with the construct of accuracy, knowing how many errors, in total, the text produced by a student contains is less useful than knowing what the specific types of errors are and how they are distributed across sentences. Therefore, useful measures of fluency in writing need to (a) identify specific locations where normal fluent production breaks down and (b) support inferences about the cognitive causes of these disfluencies. Sometimes, of course, pausing to stop and think is a normal, and desirable, feature of written production. However, pausing may also indicate underlying difficulties with the written medium (e.g., spelling) or the language in which the text is being written.

In this paper, we describe and illustrate a method for distributed collection and automated analyses of L2 writing processes through deployable, low-cost, concurrent keystroke logging and

COMBINED DEPLOYABLE KEYSTROKE LOGGING   4

eye tracking. We demonstrate the application of this method to a study of L1 and L2 writing

processes in the setting of an English-medium university in Turkey. We investigate differences

in the pausing and eye-fixation behavior of students who wrote in their L1 (Turkish) and their L2

(English).

Investigating Writing Fluency

Investigating the fluency of written text production is important for both theoretical and

practical reasons. Accounts of the mental processes that underlie writing (Hayes, 2012; Olive,

2014; van Galen, 1991), following models of speech (Dell, 1986; Levelt, 1999; Skehan & Foster,

1997), describe text production as a cascade of cognitive processes, starting from the writer's

intended message and ending with keystrokes (or pen strokes). Early, high-level processes in this

cascade are responsible for generating ideas and making decisions about appropriate rhetorical

strategies and discourse structures. This processing can be a conscious, deliberate, and effortful

"problem-solving" activity (Flower & Hayes, 1980; Hayes & Nash, 1996). This message is then

passed on to low-level processes that are responsible for lexical retrieval, grammatical encoding,

and spelling. In young and developing writers these processes are also effortful, but with

instruction and practice, they become increasingly automatized, allowing attention to be devoted

to high-level thinking and reasoning. In principle, if low-level processes run smoothly and

without demanding attention, the writer can, simultaneously and in parallel, attend to deciding on

what to say next. The resulting message is then delivered to low-level processing.

The parallel functioning of the high-level and the low-level processes in language

production lead to a fundamental "now-or-never" bottleneck, which Christiansen and Chater

(2016) explain using a stock-control metaphor: Information from upstream processes is delivered

"just-in-time" to the processes necessary for output. If low-level processing is delayed—the writer struggles with the syntax, word-retrieval or spelling necessary for expressing their message—then the message may be lost: A writer who stops in the middle of a sentence to worry about making verbs agree with nouns may, in a literal sense, forget what they were going to say next (or be unable to come up with new ideas).

Previous studies of L2 writing fluency largely employed methodologies similar to those used in oral fluency studies. The measure of fluency that researchers have most commonly depended on is number of words written per unit of time. Chenoweth and Hayes (2001), who defined written fluency as "the rate of production of text" (p. 81), measured it by words written per minute. Similarly, in their study on the effect of planning on fluency, complexity and accuracy in second language narrative writing, Ellis and Yuan (2004) used oral fluency measures to measure written fluency: the number of syllables produced per minute and the rate of disfluencies, operationalized as the number of reformulated words divided by the total number of words produced. They were not able to measure length of pauses in writing. In contrast to these studies, Chandler (2003) looked at the total amount of time students spent writing an assignment and measured the fluency of written production as the the number of minutes it took writers to produce 100 words (self-reported by student participants).

It has been a common finding that, even though writers may write proficiently in their L1 and their basic L1 literary operations do not need to be reacquired in another language (Cummins, 1980; De Larios, Murphy, & Marín, 2002), their fluency declines when they produce text in an L2: The same low-level language skills that work smoothly in L1 may become disfluent in L2, largely due to poor automation of L2-specific lexical retrieval and syntactic

COMBINED DEPLOYABLE KEYSTROKE LOGGING   6

planning (Chenoweth & Hayes, 2001; Ransdell, Arecco, & Levy, 2001; Wolfersberger, 2003).

From the theoretical perspective, therefore, being able to identify local disfluencies and infer

their cognitive causes is important for developing an understanding of the cognitive processes

that underlie writing, and particularly the development of L2 writing skills. From the practical

perspective, we argue (a) that information about the moment-by-moment fluency of a student's

written output gives valuable insights into the aspects of the students' L2 competence that

require remediation, and (b) that developing fluent written production can be, in and of itself, an

important focus of intervention. Such interventions, however, need to rely on more detailed

measures of fluency than the global speed (rate) of text production.

Keystroke Logging for Measuring Writing Fluency

The question then is what methods can provide such measures of fluency. Analyzing the

text produced by the writer provides few insights into the cognitive processes whereby the text

was produced. What is needed is a method that can directly capture the dynamics of the text

production.

Since for most adult writers keyboarding is the dominant method of writing, keystroke

logging has seen widespread adoption as a method of capturing the observable aspects of writing

process. The keystroke-logging method involves accurate (millisecond) timing of each keypress

made during text production. Inter-keystroke intervals (IKIs) can then be automatically

calculated from the keystroke log. An IKI is defined as the time that elapsed between two

consecutive key-presses. For example, if the writer is fluently typing the word "hello," the first

two letters would be produced by consecutively pressing and releasing the keys "h" and "e" on

the keyboard. If the key "h" was pressed at time t_1 and the key "e" was pressed at the time t_2,

COMBINED DEPLOYABLE KEYSTROKE LOGGING 7

then the IKI associated with the latter key would be $IKI = t_2 - t_1$. The entire time-course of text

production can thus be segmented into sequential, non-overlapping IKIs, and any event that took

place during this time-course can be mapped to exactly one IKI during which such event

occurred.

IKIs, arguably, can give considerable insight into the cognitive processes that underlie

written composition. Obviously, minimum IKI is determined by the motor constraint of typing:

the brief time that it takes to move fingers between keys. As we discussed above, behind each

key press is a cascade of cognitive processes that start with from content generation, go through

syntactic, lexical, and orthographic processing, and end with the motor planning of finger

movements. All of these cognitive processes may occur in parallel with output (i.e. the actual

implementation of the motor program), providing a steady flow of information from mind to

fingers, in which case text production will be fluent. However, if there is a delay in any of the

upstream processes (e.g., search for content, syntax, spelling, or keys), then this delay propagates

down the cascade and is observed as a longer IKI. Therefore, disfluencies in writing can be

identified by prolonged IKIs. Such IKIs that are indicative of a disfluency are often termed

"pauses," however operational definitions of "pauses" vary. In many studies, "pauses" are

defined as IKIs above particular threshold (Alves & Limpo, 2015; Chukharev-Hudilainen, 2014;

Connelly, Dockrell, Walter, & Critten, 2012; Leijten & Waes, 2013; Torrance, Rønneberg,

Johansson, & Uppstad, 2016).

Inferences about the cognitive cause of a disfluency must necessarily be made on the

basis of the location of the corresponding IKI within the text: An IKI at a particular

linguistically-relevant location is associated, in part, with the cognitive activity of planning the

linguistic unit that follows. Thus, for example, word-initial IKI is likely to be partially associated with the planning of this word (Torrance, Rønneberg, et al., 2016). Since L2 writers' processes are poorly automatized compared to L1 writers, it is expected that L2 writers would have longer IKIs before the linguistic units that require more planning effort. This was confirmed by Spelman Miller (2000), who used keystroke logging to investigate writing-process differences between native English speakers (L1 group) and ESL learners (L2 group) composing text in English. The study reported the L2 group paused longer at all locations, but especially at the beginning of clauses and sentences.

Some of the activity that occurs during an IKI may be related to reading, defined here loosely as looking back at the text produced so far. Therefore, further inferences about the writer's cognitive activity during an IKI can be made if keystroke data are combined with information about where the writer looks between consecutive keystrokes (Torrance, Johansson, Johansson, & Wengelin, 2016; Wengelin et al., 2009). For example, if the writer's eye fixations remain within the unfinished sentence containing the current point of inscription, cognitive processing during the pause is likely to be concerned with this sentence. On the other hand, looking back at stretches of text beyond the current sentence will indicate a different kind of processing, which might be more ideational than linguistic in nature (e.g. deciding what to say next based on what has just been said in the text).

Eye-Tracking for Measuring Writing Fluency

Although eye-tracking per se is an established technique in psycholinguistic research, it has not seen widespread adoption in the study of written production. Arguably, the application of eye-tracking to the study of writing processes has been impeded by two factors. First,

commercially available eye movement analysis software is tailored to the most common use cases of the eye-tracking technology: reading and visual perception studies and user-experience research. In this case, the standard approach to analyzing the eye-tracking data is by predefining "areas of interest" within the stimulus images and then using software to calculate the duration and sequencing of visual fixations that fall within each area of interest. This approach will not work for writing, where each participant composes a different text, and the onscreen configuration of the text constantly changes as composition progresses. Editing, line wrapping, and scrolling operations all mean that words will not retain their original position on the screen. Therefore, using the *(x, y)* coordinates of gaze point *on the screen*—the data provided by the eye tracker—to identify the character (and word, and sentence) *within the text* that the writer is looking at is a non-trivial issue.

The second factor is the prohibitively high cost of eye tracking hardware and software. A single research-grade unit may cost tens of thousands of dollars, making it impossible to collect data at large scales and in settings other than a psycholinguistic laboratory. While much work has been done on logging and analyzing keystrokes in naturalistic writing environments (Chukharev-Hudilainen, 2014; Lindgren, Spelman Miller, & Sullivan, 2008; Schoonen et al., 2003; Uppstad & Solheim, 2007; Van Waes & Leijten, 2015), it has not been logistically and financially feasible to augment keystroke data with gaze-point information outside of laboratory settings. Collecting data at scale, however, would radically increase sample sizes and thus the noise possibly associated with loosely-controlled (i.e., noisy) data collection environments may be compensated by the increased statistical power. Collecting data in settings that are closer to

real-world classroom environments is necessary for applied research that seeks to improve

language learning practices.

The first issue has been addressed in multiple ways. The most obvious approach is to

overlap a screen recording of the composition process with visualizations of gaze-point

movements, such as fixation paths or heat maps, and then manually annotate the resultant video

stream (Revesz et al., present issue). This, however, is very time-consuming due to the amount of

manual coding involved. A more technologically advanced approach is to automatically capture

and save a bitmap screenshot each time the writer strikes the space bar or return key, or moves

the cursor in the text. These bitmaps can then be annotated to map the coordinates of each word

on the screen to eye-tracking measurements (Hacker, Keener, & Kircher, 2017). The ideal

approach, however, would be to allow for fully automatic mapping of eye fixations to the

characters of the text being edited without any need for manual annotation. This approach is

clearly preferable, especially for research involving quantitative hypothesis testing, because any

need for additional manual annotation substantially increases the cost of analyses, diminishes

scalability, introduces human-related sources of error, and reduces the reproducibility and

transparency of the research. To our knowledge, the only system that follows this approach is

EyeWrite (Simpson & Torrance, 2007; Torrance, 2012), which is designed for experimental,

laboratory-based research and only works in conjunction with high-end eye trackers

manufactured by SR Research Ltd. Thus, the EyeWrite system does not solve the second

problem: the need for use of expensive research-grade equipment.

In this paper, we describe and illustrate a method for scalable, distributed collection and

automated analyses of L2 writing processes through deployable, low-cost, concurrent keystroke

logging and eye tracking. This method resolves both of the issues identified in the previous section. The data collection instrument that we present takes the form of a specially developed web-based text editor with embedded keystroke logging capabilities and an interface with eye-tracking hardware. Time-aligned keystroke and eye data allow for a deterministic reconstruction of all observable aspects of the text-production process. Such reconstructions can take two forms: (1) animated visualizations for qualitative review and (2) machine-readable log files for quantitative analysis.

In the remaining parts of this paper, we will first introduce our prototype system implementing the proposed method of data capture and analysis, focusing on qualitative visualizations and on the extraction of quantitative data (i.e. variables of interest). Then we will proceed to report the empirical study demonstrating the method. The article will conclude with a discussion of both the results of the empirical study and the affordances of the methodology employed.

The CyWrite system

Data Capture

The proposed research method is implemented in a prototype system named CyWrite. The key innovation of the approach is the collection of machine-readable data that capture all observable aspects of text production and allow for its deterministic reconstruction and automated analyses. These data include concurrent, time-aligned logs of (1) keystrokes, (2) eye fixations, and (3) moment-by-moment changes to the contents and layout of the text on the screen.

In CyWrite, control over the layout of the text (formatting, scrolling, wrapping) is achieved by using a custom text editor with a clearly defined, deterministic (i.e. stably replicable) text-rendering behavior. To enable scalable and distributed data collection, the text editor is implemented as a client-server web application. This affords easy deployment not requiring any specialized software (beyond eyetracker drivers) to be installed on user computers. The editor provides a familiar writing experience identical to any other low-feature word processor (e.g., Microsoft WordPad): It supports typing, clipboard operations (copy, cut, paste), simple character formatting (bold, italic, underline) and paragraph justification (left, center, right). By contrast, laboratory software (EyeWrite) is restricted to typing, deleting, and keyboard-based cursor navigation—a limitation that may not be an issue for controlled environments, but may diminish the usability of the software in classroom settings.

The editor application consists of the client part that works in the user's browser, and the server part that is deployed in the cloud (i.e. on any machine that is accessible from the internet via the HTTP/S and WebSocket protocols). Instead of relying on the browser's internal text-editing capabilities (such as HTML forms or ContentEditable elements), the editor programmatically implements text editing and visualization. At all times, the internal state of the editor is represented as $E_t = \{T_t, D_t\}$, where E_t is the state of the editor at the time point t, T_t is the text which has been composed as of this time point, and D_t is the vector of display parameters, such as screen size, window size, font size, and the scrolling position of the viewport within the text. Each time the user interacts with the editor by pressing a key or clicking a mouse button, a custom JavaScript event handler is invoked within the client part of the editor application. The event handler obtains a timestamp for the input event and generates an object representing an

editing operation (e.g. "insert character z into paragraph x at offset y") which transforms the text, or a display operation (e.g. "move the scrolling viewport n lines up in the text") which modifies the display parameters. Both editing and display operations are similar in that they modify the editor state: $E_{i+1} = o_j(E_i)$. Here E_i is the original state of the editor before the operation is applied, o_j is the operation, and E_{i+1} is the resultant state of the editor after the operation has been applied. Crucially, this process is deterministic: Applying the same sequence of operations to the editor state guarantees an identical resultant state.

JavaScript code is further used to programmatically render the text in the browser window each time an editing or display operation is applied. To simplify text rendering procedures, nonproportional fonts are used in the editor so that each character occupies exactly one cell in a rectangular grid. When the text is rendered, associations between characters and grid cells are computed. As a result, at any particular point in time the character in each square is known.

Thus, editor states themselves do not need to be stored in a log file. Instead, it is sufficient to store the timestamped sequence of operations that occurred during the editing session in order to reconstruct, for any given time point, both the current text (including portions scrolled out of the current viewport) and the characters displayed at any given (x, y) coordinates within the on-screen viewport. It should be noted that most keypresses produce two operations: a text-editing operation "add a character" and a display operation "move cursor forward." Some keypresses might not produce any operations at all (e.g., a modifier key, such as Shift, does not itself produce an editing operation, but rather affects the character typed by the alphanumeric key that is pressed while the modifier key is held down). Thus, generally speaking, keypresses and

operations are related as "many-to-many." To allow for IKI analyses, a separate timestamped log

of keypresses is recorded in addition to the log of operations. The two logs are cross-indexed to

simplify analyses. This method of keystroke data collection and analysis allows for the

measurement of IKIs with the standard deviation of 5 ms from the true value

(Chukharev-Hudilainen, forthcoming), which is similar to other software-based techniques (Frid,

Wengelin, Johansson, Johansson, & Johansson, 2012).

Eye tracking in CyWrite can be performed, in principle, by any device that supports an

application programming interface (API) for the real-time streaming of gaze data, which includes

most of commercially available eye-tracking devices. Eye fixations, similarly to keypresses, are

recorded in a timestamped log. In this study, GazePoint GP3 devices (0.5–1 degree of visual

angle accuracy, 60 Hz sampling rate) were used due to their low cost: At the time the study was

conducted, they were available from the manufacturer at 495 USD per unit, which is a fraction of

the cost of a research-grade system. At this price, both the lead institution in the U.S. and the

remote data-collection site in Turkey were able to afford multiple devices to be installed in

on-campus teaching labs, classrooms, and faculty offices. The editor software interfaces with the

GP3 devices via the Open Gaze protocol (Hennessey & Duchowski, 2010) which provides a

real-time feed of eye fixation coordinates.

Timestamped records of keypresses, eye fixations, and editing/display operations are

streamed live to the server side via a WebSocket connection. The server-side application is

implemented in JavaScript (with NodeJS) and shares most of its codebase with the client-side

application running in the user's browser. The server application can run in the online and the

offline mode. In the online mode, the server receives timestamped records from the client, stores

them in a log file for further analysis, and applies editing operation records to the server-side copy of the text. This way, the server maintains a real-time copy of the editor's internal state E_i identical to the one in the browser. The fact that both the server and the client software is implemented in a single programming language and share the same codebase substantially simplifies this implementation. A custom application-level communications protocol, implemented on top of WebSockets, provides data integrity and consistency checks and seamlessly renegotiates the connection between the client and the server should the WebSocket be closed. In the offline mode, the server is instructed to read records from a log file and "replay" them as if they were being received from the client side over the live connection. This allows for repeated computational reconstructions of the writing session for various analyses.

For the present study, while the data collection site was located in Turkey, the server was deployed on the lead institution's campus in the U.S. No technical issues due to network connectivity or latency were noted. Research team members in the U.S. were able to assist the researcher in Turkey with data collection procedures in real time, including monitoring the quality of eye-tracking data.

Visualizations

A second client-side web application, called "the viewer," was developed to generate visualizations of the process of composition. The viewer connects to the server, running in either the online or the offline mode, and obtains the editor-state information E_i from the server. This results in an animation looking just as if it were a high-fidelity screen-captured video of composition except that it is stored in a log file that is much more compact than an actual video file would be and, more importantly, permits various automated analyses.

COMBINED DEPLOYABLE KEYSTROKE LOGGING   16

In the offline mode, the viewer also plots the entire writing session as a "process graph"

similar to one that is plotted by InputLog (Leijten & Waes, 2013), but different in that the graph

is dynamically linked to the animated playback. The current position on the playback timeline is

indicated by a movable playhead in the graph. The playback of any desired fragment of the

writing session can be immediately viewed by clicking on the corresponding area in the graph.

This idea was borrowed from the user interface of many audio editors, where the entire timeline

of an audio file is shown as as a waveform or a spectrogram, and clicking on a point in that

timeline allows the user to quickly jump to the corresponding time point in the file. In our case,

the "process graph" of a writing session takes place of the waveform visualization of audio

(Figure 1).

(Figure 1 near here.)

The graph plots the current length of text (green line), the total number of characters

produced so far including those that have been subsequently deleted (blue line), the current

location of the cursor (red line), the location within the text of the character that is currently

displayed in the top-left corner of the viewport (pink line), and the position of the eye fixation

(yellow line). All in-text locations are represented as character-wise offsets from the beginning

of the text (the y axis) changing over time that elapsed since the beginning of the writing session

(the x axis, in minutes). In the playback, the current eye fixation is marked by the yellow circle.

Another visualization generated by CyWrite that is useful for exploring the pausing

behavior is shown in Figure 2.

(Figure 2 near here)

In the pause analysis visualization, the intensity of the red shade corresponds to the duration of the IKI that occurred before producing the shaded character: the darker the color, the longer the IKI. As it is evident from Figure 2, longer IKIs are more prevalent in L2 writing than in L1. The patterning of IKIs is also different across languages: Longer IKIs tend to occur at the beginning of sentences and clauses (or around revisions that are indicated with strike-through text) in L1, while in L2 they appear more frequently in mid-sentence locations. These observations provide some tentative and qualitative insights into the differences between L1 and L2 production.

Quantitative analyses

To permit automated quantitative analyses of keystroke and eye-fixation data, the server code, when running in the offline mode, provides interfaces for controlling the replay of the logged data programmatically. This allows the programmer to supply a callback function that will be called each time a log entry is processed or the editor-state information is updated. An analysis program (script) can then be written to output datasets in arbitrary formats based on what is required for further analyses. The analysis script has programmatic access to both the log entries (i.e. the operations being replayed) and the resultant states of the editor.

Comparing Writing Processes in L1 (Turkish) and L2 (English)

in University-Level Writers

To illustrate the use of the combined eye-tracking and keystroke logging methods that we have just described, we report an exploratory study in which we compared L1 (Turkish) and L2 (English) writing in adult English language learners. Our general aim in this study is to establish whether, as we suggested in our discussion above, it is indeed possible to differentiate, in systematic and predictable ways, writing processes in L1 and L2 on the basis of (1) output timecourse as measured by IKIs and (2) writers' reading activity as captured by eye-tracking measures.

More specifically, we would expect that if L2 writers require more cognitive effort for syntactic planning, lexical retrieval, and orthographic production, then longer IKIs would occur at those points in the text where this processing is expected to occur, namely, at clause and word boundaries. While L1 writers can engage in message planning (i.e. thinking what to say next) in parallel with outputting the previous text segment, L2 writers are less likely to do so due to processing demands imposed by poorly automated low-level language skills. Assuming that ideational planning happens at a sentence level, we would expect longer IKIs at sentence boundaries in L2.

Secondly, if reading during writing serves to help the writer refresh in their memory the contents of what they have previously written, and lower L2 proficiency would reduce chances of parallel processing and increase chances of the writer forgetting what they have just said, then more reading activity would be expected when writers pause in L2 than when they do so in L1.

Therefore, the empirical study reported in this paper was driven by the following two research questions:

RQ1: To what extent does students' writing fluency in L1 (Turkish) and L2 (English) differ, as determined by length of IKIs at word, clause, and sentence boundaries?

RQ2: To what extent do engage in reading activity when they write in L1 (Turkish) and L2 (English), as determined by eye-tracking measures?

Participants

Participants were 24 native speakers of Turkish (20 female) selected through convenience and volunteer sampling. They were all undergraduate students enrolled in three education majors at a private university in Turkey with English as the medium of instruction: 14 students majored in Guidance and Psychological Counseling, six in English Language Education, and four in Early Childhood Education. They were in different years of their undergraduate programs of study: two were in their first year, 17 in their second year, four in their third year, and one in their fourth year. The age of the participants ranged from 19 to 24 years.

Instruments

Participants completed two writing tasks, one in Turkish (L1) and one in English (L2), with language and topic counterbalanced across participants. The prompts were adapted from the Test of English as a Foreign Language (TOEFL), a widely adopted high-stakes standardized test. Both prompts were translated into Turkish.

Prompt 1: "Some people say that music not only entertains us but changes how we think and feel about ourselves. Do you think music has the power to influence as well to entertain

people? Support your views with specific examples from your experience, observations or

reading."

Prompt 2: "Some people say that computer technology is a barrier to developing real

friendships. To what extent do you agree or disagree with the statement above? Support your

views with specific examples from your experience, observations or reading."

Data Collection Procedures

After completing the informed consent process, participants were seated in front of a

computer (Lenovo Ideapad 100 with an external USB keyboard and an external DELL SE2216H

21.5-inch monitor screen set to the resolution of 1920x1080). A 60 Hz version of the Gazepoint

GP3 eye-tracking device was mounted on a standard tripod below the external monitor screen.

First, the eye-tracker was calibrated, which involved participants looking at a moving target

displayed on the screen (a nine-point calibration pattern was used similar to the standard one

provided by the Gazepoint software, but modified to randomly rearrange the sequence of fixation

points and thus prevent participants from anticipating the movement of the target). After the

calibration, students composed the two essays. The prompt was placed at the top of the editable

file. The paragraph containing the prompt was specially marked to permit separate tracking and

analysis of cursor movements and eye fixations within that paragraph. The editor was configured

to use the fixed-width Consolas font (font size 35 pixels, with inter-linear spacing increased by

26 pixels). A screenshot of the interface is presented in Figure 3.

(Figure 3 near here.)

Participants were given a maximum of 40 minutes to compose each essay, but they were free to finish writing earlier if they were satisfied with their text. They received a monetary compensation for taking part in the study.

Data analyses

Analysis of log files. An analysis script was developed to generate two types of output: (1) a tabular file describing each character that was added to the text, including those that remained in the final version of the text and those that were removed at some point in the composition process; (2) a plain-text version of the final text produced in the editing session. Each character in the text was labeled with a unique identifier that allowed for easy automatic mapping between the two types of output.

The first type of output included variables that described the character produced, the duration of the IKI immediately preceding the production of this character, and eye fixations that occurred during this IKI. When multiple keys had to be pressed to produce a single character (e.g., the Shift key pressed first, followed by a letter key to produce a capital letter), the IKIs were measured to the initial keypress in the sequence (i.e., the Shift key in the example given).

For present purposes we report analysis of disfluencies that involved the writer pausing but did not then involve revision of the text to, for example, correct an error. Thus, this first type of the output was filtered to eliminate all keystrokes that were used for moving the cursor around in the text and for deleting characters. We also excluded from analysis character keypresses that occurred immediately after the cursor was moved or after one or more characters were deleted.

In the second type of output, clauses and orthographic sentences were annotated. Clause boundaries were annotated manually. The beginning of each clause was marked in the text file

with an opening bracket, and the end of each clause was marked with a corresponding closing

brackets. Square brackets were used for finite clauses (i.e. those where the main verb is marked

for mood or modality), and curly brackets were used for non-finite clauses (i.e. those that do not

show mood or modality, such as infinitive and participle clauses in English). Embedded clauses

were marked as being part of the matrix clauses (e.g. "[The house [that Jack built] was

destroyed.]"). In those instances where two clauses were connected by a coordinating

conjunction, such conjunction was deemed, by convention, part of the clause immediately

following the conjunction. When a dependent clauses was introduced by a subordinator, such

subordinator was included in the dependent clause. Similarly, when one subordinator introduced

multiple dependent clauses, joined in turn by coordinating conjunctions, the subordinator was

marked as part of the first dependent clause it immediately preceded, and coordinating

conjunctions were marked as parts of the dependent clauses they immediately preceded.

Prepositions preceding non-finite dependent clauses (with the latter acting as the objects of such

prepositions), however, were not considered part of such clauses but rather deemed part of the

matrix clause. When sentence-level punctuation was at odds with syntactic constituency,

constituency information prevailed in annotation. When a grammatical error precluded

annotation, the minimal and simplest edit(s) that would make the sentence grammatically

acceptable were considered, and clause boundaries were marked based on such potential edits

(without actually applying any edits to the text).

Independently from clause-level annotations, texts were split into orthographic sentences

using NLTK sentence tokenizers for English and Turkish (Perkins, 2010). Sentence splitting then

was manually corrected. All conflicts between sentence splitting and clause annotation (i.e. when

an orthographic sentence did not begin with an independent clause) were manually resolved, and the few instances that were not resolvable due to syntactic irregularities of the text were marked as ambiguous and excluded from further analysis.

Linguistic annotations were mapped onto IKI data to label each IKI with a factor variable describing the linguistic location of the character produced immediately after the IKI. The levels of the Location factor were defined as follows: "within-word" for IKIs that occurred between the typing of two word characters, "word-initial" for those after typing the space and before the first letter of a word that did not start a new clause, "non-finite-clause-initial" for the beginning of a non-finite clause that did not start a new sentence, "finite-clause-initial" for the beginning of a finite clause that did not start a new sentence, and "sentence-initial" for the beginning of a new orthographic sentence. As discussed above, we assume that an IKI at a particular location is associated, in part, with the cognitive activity of planning the linguistic unit that follows. This, for example, would predict longer "word-initial" IKIs than "within-word" IKIs.

Next, log files were used to determine the position of eye fixations in the text relative to the current point of inscription (i.e. the location of the cursor) at the time of fixation onset. Three eye-tracking measures were derived: probability of looking back, mean character-wise lookback distance, and lookback distance in terms of the types of linguistic units fixated. For all three measures, only fixations on the text five or more characters behind the cursor were included to eliminate fixations on the cursor (i.e. to exclude the monitoring behavior while typing and only retain fixations that represent looking back at the text produced so far). Off-screen fixations, fixations on the writing prompt, and fixations on user-interface elements other than the writer's own text were excluded. For each character of the text, we identified all eye fixations that began

during the IKI immediately preceding the production of this character. Mean character-wise lookback distance then was measured for each IKI as the mean number of characters in the text between the points of fixation and the point of inscription, counted in the version of the text that existed at the time of fixation onset.

Character-wise lookback distance, however, does not account for the distribution of the linguistic structures in the text: A fixation that is a certain number of characters behind the cursor can be both linguistically close to the point of inscription (e.g. within the same clause) and linguistically remote (e.g. in a different sentence). To account for linguistic structures, each fixation was automatically coded with one of the following four labels: "same-clause" if the point of fixation and the point of inscription belonged to the same clause; "crossing-clause" if the fixation was on a character in a different subordinate clause yet within the same matrix clause as the point of inscription; "different-clause" if the fixation was on a character within a different matrix clause than the point of inscription, but still within the same orthographic sentence; and "different-sentence" if the fixation was on a character within the different orthographic sentence than the point of inscription. Whenever coding was ambiguous (e.g. if the space character between two clauses was fixated), by convention the code representing the shortest linguistic distance was preferred (e.g. "different-clause" over "different-sentence").

Statistical analyses. Our analyses were based on fitting linear mixed-effects regression (MER) models to predict dependent variables related to pausing and reading behaviors based on the two factors described above: Location (with levels corresponding to the linguistic location of the IKI based on the linguistic unit that immediately follows) and Language (L1 vs. L2). Specifically, for each dependent variable of interest, four nested MER models were fit to the

data: M0, an intercept-only model; M1, a model adding a fixed effect for Location; M2, a model adding a fixed main effect for Language; and M3, a model adding an interaction between Language and Location. All models included random by-subject intercepts and random slopes for Language and Location. Gains in goodness of fit of successive models were evaluated by chi-square change. Fixed-effect parameters of the full model were used to estimate means of the dependent variables at different levels of Language and Location. Wald estimates of the confidence intervals (CIs) for means were then derived from the model.

Results

IKIs at linguistically-relevant locations

The distribution of IKIs showed a strong negative skew typical for response time data. We therefore trimmed the data to remove IKIs longer than 8 s (0.14% of IKIs), and then log-transformed the remaining values prior to inferential testing.

Each consecutive model, of the four nested MER models, resulted in a significantly better fit to the data than the previous one: M1, $\chi^2(4) = 59.3$, $p < 0.001$; M2, $\chi^2(1) = 16$, $p < 0.001$; M3, $\chi^2(4) = 247$, $p < 0.001$. Table 1 gives estimated mean IKIs and 95% CIs derived from this model.

Table 1. *Estimated inter-keystroke intervals, in milliseconds, by Language and Location within the text, and 95% CIs*

Location	L1 (Turkish)		L2 (English)	
	Mean	**95% CI**	**Mean**	**95% CI**
within-word	224	209, 240	235	221, 250
word-initial	399	363, 437	519	472, 572
non-finite-clause-initial	602	522, 693	520	458, 590

| finite-clause-initial | 612 | 532, 705 | 549 | 474, 636 |
| sentence-initial | 996 | 810, 1224 | 1117 | 913, 1366 |

Participants typed more slowly when they composed text in English than when they did it in Turkish, except at clause boundaries. Their mean within-word IKIs, however, were longer by just 11 ms. When writing in their L1, participants were slower at the beginnings of words than within words, slower at the beginnings of clauses than at the beginnings of words, and slower at the beginnings of sentences than at the beginning of non-sentence-initial clauses.

The improved fit of the nested models allows to infer that patterns of participants' pausing behavior differed significantly between the two languages. Contrary to what was predicted by Language alone, at the beginning of non-finite clauses participants were 82 ms faster, and at the beginning of finite clauses they were 63 ms faster. However, for words that were not at the start of clauses initial IKI was slower by 121 ms in English.

Eye fixations

Fixation probabilities. Mixed-effects logistic regression models were fit to predict the probability of one or more lookback fixations occurring during an IKI. Four nested models were evaluated, starting with an intercept-only model (M0) and then adding the fixed effects. Each consecutive model resulted in a significantly better fit: $\chi^2(4) = 40.36$, $p < 0.001$; $\chi^2(1) = 5.46$, $p = 0.02$; $\chi^2(4) = 10.79$, $p = 0.03$. Estimates derived from the full model (M3) are given in Table 2.

In L1, lookback fixations were more probable at the start of larger linguistic units, increasing almost 8-fold from within-word locations to sentence-initial locations. Probabilities of looking back when writing in L2 were generally higher than in L1, except for locations at the

start of embedded clauses: Interestingly, the mean probability of looking back in L2 was higher at the start of a non-finite clause than at the start of a finite clause, which is different from the pattern observed in L1.

Table 2

Probabilities of one or more lookback fixations during an inter-keystroke interval, by Language and Location, and 95% CIs

Location	L1 (Turkish)		L2 (English)	
	Mean	95% CI	Mean	95% CI
within-word	0.040	0.029, 0.054	0.052	0.038, 0.069
word-initial	0.081	0.063, 0.103	0.114	0.092, 0.141
non-finite-clause-initial	0.090	0.051, 0.155	0.138	0.091, 0.204
finite-clause-initial	0.142	0.101, 0.197	0.092	0.058, 0.143
sentence-initial	0.302	0.213, 0.409	0.355	0.255, 0.470

Character-wise lookback distances. The distribution of mean lookback distances was negatively skewed, but approximated the normal distribution after log-transformation. The four nested MER models were fit to the data. M1 (the model adding a fixed effect for Location) resulted in a significantly better fit than M0 (the intercept-only model): $\chi^2(4) = 17.97$, $p = 0.001$. However, adding the fixed effect of Language (M2) and the interaction of Language and Location (M3) did not provide a further improvement: $\chi^2(1) = 2.15$, $p = 0.14$; $\chi^2(4) = 2.87$, $p =$

0.58. This permits inferring that fixation behavior during IKIs, in terms of the character-wise distance of lookbacks, was similar in L1 and L2.

Estimated mean character-wise lookback distances (Table 3) during IKIs before sub-sentence units (i.e. at all locations except "sentence-initial") were similar, with numerically close means and largely overlapping CIs. Before starting a new sentence, however, participants would look significantly farther back in their text.

Table 3

Lookback distances (characters), by Language and Location, and 95% CIs

Location	L1 (Turkish)		L2 (English)	
	Mean	95% CI	Mean	95% CI
within-word	89	73, 107	102	80, 131
word-initial	80	65, 99	101	80, 128
non-finite-clause-initial	78	46, 132	115	78, 171
finite-clause-initial	82	60, 113	89	59, 135
sentence-initial	119	87, 164	151	113, 202

Linguistic fixation distances. Next, we analyzed lookbacks in terms of their linguistic distances from the point of inscription. A separate series of nested mixed-effects logistic regression models was built for each type of linguistic distance (i.e., "same-clause," "crossing-clause," "different-clause," "different-sentence;" see Measures above for

COMBINED DEPLOYABLE KEYSTROKE LOGGING   29

operationalizations) to predict probabilities of one or more fixations occurring during an IKI at

this linguistic distance from the point of inscription. Within each series, every model added a

fixed effect to the previous model, and improvement of model fit was evaluated. Estimated mean

probabilities and the respective CIs are given in Table 4, and contributions of the fixed effects to

the model fit are presented in Table 5.

Table 4

Probabilities of fixations at a given linguistic distance from the point of inscription, by Language

and Location, and 95% CIs (in brackets)

Location	Language	Linguistic distance from point of inscription			
		same-clause	crossing-clause	different-clause	different-sentence
within-word	L1	0.007 [0.005, 0.010]	0.008 [0.006, 0.012]	0.002 [0.001, 0.003]	0.021 [0.015, 0.029]
	L2	0.007 [0.005, 0.010]	0.006 [0.004, 0.010]	0.003 [0.002, 0.006]	0.030 [0.022, 0.040]
word-initial	L1	0.028 [0.020, 0.038]	0.019 [0.014, 0.026]	0.004 [0.002, 0.007]	0.038 [0.028, 0.051]
	L2	0.024 [0.018, 0.032]	0.019 [0.013, 0.028]	0.004 [0.002, 0.007]	0.069 [0.055, 0.086]
non-finite-clause-initial	L1	0.027 [0.011, 0.068]	0.013 [0.004, 0.047]	0.007 [0.002, 0.033]	0.051 [0.025, 0.103]
	L2	0.035 [0.016, 0.075]	0.026 [0.011, 0.060]	0.007 [0.002, 0.0028]	0.067 [0.038, 0.115]
finite-claus	L1	0.060 [0.036,	0.019 [0.008,	0.009 [0.003,	0.084 [0.055,

e-initial		0.100]	0.043]	0.0026]	0.126]
	L2	0.030 [0.014, 0.063]	0.000	0.010 [0.003, 0.031]	0.056 [0.033, 0.100]
sentence-initial	L1	0.118 [0.070, 0.191]	0.102 [0.054, 0.186]	0.019 [0.007, 0.053]	0.147 [0.095, 0.220]
	L2	0.139 [0.085, 0.220]	0.066 [0.028, 0.146]	0.036 [0.014, 0.089]	0.188 [0.124, 0.275]

Table 5

Contribution of fixed effects to the fit of models predicting probabilities of fixations at a given

linguistic distance from the point of inscription

Effect	Linguistic distance from point of inscription			
	same-clause	crossing-clause	different-clause	different-sentence
Location	$\chi^2(4)=56.168$, $p<0.001$	$\chi^2(4)=30.201$, $p<0.001$	$\chi^2(4)=21.284$, $p<0.001$	$\chi^2(4)=33.267$, $p<0.001$
Language	$\chi^2(1)=0.167$, $p=0.683$	$\chi^2(1)=0.997$, $p=0.318$	$\chi^2(1)=3.559$, $p=0.060$	$\chi^2(1)=7.345$, $p=0.007$
Interaction	$\chi^2(4)=4.786$, $p=0.310$	$\chi^2(4)=14.128$, $p=0.007$	$\chi^2(4)=1.235$, $p=0.870$	$\chi^2(4)=11.822$, $p=0.020$

The probability of one or more fixations occurring on the text of the clause currently

being composed ("within-clause" label) varied based on the Location of the point of inscription,

but not based on Language. Specifically, fixation probabilities increased with the increasing

hierarchical status of the linguistic unit of text that followed. For example, within-clause

COMBINED DEPLOYABLE KEYSTROKE LOGGING   31

fixations were more likely to occur during word-initial IKIs than during within-word IKIs, etc. A

similar pattern was evident for fixations on other clauses within the same sentence

("crossing-clause" and "different-clause" labels).

However, when it came to looking back at a previous sentence, participants' behavior

varied across languages with a significant main effect and interaction. For example, at the

beginning of a mid-sentence word, participants were almost twice as likely to look back at a

previous sentence when they wrote in L2 vs. L1. Similar patterns (L2 > L1) were observed for all

other locations except finite-clause-initial, where, interestingly, fixations on a different sentence

were more likely in L1 than in L2.

Discussion and Conclusion

This paper focused on describing and illustrating a method for scalable, distributed

collection and automated analyses of L1 and L2 writing processes through deployable, low-cost,

concurrent keystroke logging and eye tracking. As demonstrated in the paper, the method allows

for the collection of data through a web application deployed at a remote site while using

inexpensive eye-tracking hardware. Although inexpensive hardware has relatively low

spatio-temporal resolution, statistical power can be boosted by the larger size of samples that can

be collected in this manner. Writing-process data can be automatically triangulated with

linguistic annotations of texts produced by participants. In the reported study, certain types of

annotation were conducted manually, in part because manual annotation was necessary for a

separate corpus-based project utilizing the same dataset. In principle, however, natural language

processing (NLP) tools can be used to perform such annotations automatically and in real time.

This shows the potential for the method to be used in practical applications, such as automated

writing evaluation (AWE) and computer-assisted language learning (CALL).

The reported empirical study shed some light on the two research questions that we

posed. Specifically, RQ1 asked to what extent students' writing fluency would differ when

writing in L1 (Turkish) and L2 (English). We predicted that IKIs at word, clause, and sentence

boundaries would be longer in L2 than in L1. In fact, pausing behavior did differ significantly

between L1 and L2: Participants were generally slower, with longer IKIs, when writing in L2.

This finding is consistent with previous research demonstrating that written production tends to

be less fluent in L2 than in L1 (Schoonen et al., 2003; Van Waes & Leijten, 2015), and L2

writers tend to pause more at all linguistically-relevant locations (Spelman Miller, 2000).

However, contrary to our prediction, at the beginning of non-sentence-initial clauses

participants were significantly *faster* in L2 than in L1. Even though they spent less time planning

the clause that they were about to produce, writers were subsequently less fluent in producing

individual words within the clause, as evidenced by the increased word-initial IKIs. This is an

interesting finding indicating that planning effort might be reallocated in L2 compared to L1:

Participants may be more thoroughly pre-planning the structure of the clauses they are about to

produce in L1 (hence longer clause-initial IKIs), while in L2 they may be "jumping" into

outputting the clause before it has been sufficiently planned, and postponing some of the

planning decisions until after part of the clause has already been output. This could be described

as a shift in planning effort from clause-initial locations to within-clause locations.

This finding was incidental to the study reported in this paper, and thus calls for more

thorough investigation in a follow-up study. If this shift in planning effort is confirmed, it may

have crucial practical implications as it may be signaling the redistribution of effort that may be

potentially disruptive to the flow of text production: A writer who does not sufficiently plan

ahead at clause boundaries but pauses at the beginning of words within the clause instead might

be distracted from the ideas that are being expressed in the clause and thus might end up

producing a lower-quality text. If this is the case, then interventions directly and strategically

modifying the distribution of planning effort during L2 production might be warranted to

minimize associated disfluencies.

RQ2 asked whether reading behavior would differ between L1 and L2. We predicted,

specifically, that more reading activity would occur in the (less automatized) L2 than in the

(more automatized) L1. Indeed, participants were found to look back more frequently during L2

production than they did during L1 production. Look-back fixations, however, would stay within

the same mean distance (character-wise) from the point of inscription in both languages.

Similarly, the probability of looking back within the current sentence did not differ between

languages. On the other hand, participants were significantly more likely to look at a previous

sentence when they composed in L2 than in L1. This was especially evident at the beginning of

mid-sentence words: At these locations participants were almost twice as likely to look at a

previous sentence in L2 than they were in L1. This finding calls for additional investigation in

future work. If looking back at a previous sentence serves as a memory refresher (or as

scaffolding for planning future content), it may be especially problematic from the practical

standpoint: Writers who not only pause longer mid-clause, but also use the pause time to look

back at the previous sentence may thus be distracted from the idea package they are currently

trying to express. This, again, may warrant interventions that explicitly modify behaviors that contribute to disfluencies in L2 writing.

The present results are interesting from both the theoretical and the applied standpoints. Theoretically, they shed light onto the differences between the cognitive processes that underlie written production in L1 and L2. Because each IKI is associated, in part, with planning the linguistic unit that follows, differences in IKI patterns indicate that the planning effort is distributed differently based on the language in which the text is being produced. Much of this planning, in either language, occurs at very short time-scales (under 2 seconds) and may not, therefore, reach awareness in the writer (Torrance, 2015). The proposed method thus permits the detection of differences that might not be detectable using conventional verbal reporting protocols.

Practically, as was mentioned above, explicit fluency-focused interventions may be warranted in L2 writing instruction. On the one hand, such interventions at this point may seem far-fetched, because more research first needs to be done in order to understand what may be causing these shifts in pausing and reading behaviors in L2 text production. On the other hand, interventions based on compensatory L2 writing strategies which implicitly target writing fluency (e.g. using L1, to a different extent, for ideational planning) have already been proposed (Kim & Yoon, 2014; Wolfersberger, 2003). It is important to note that existing intervention are blind to the actual writing process and thus have to infer disfluencies from indirect data. The deployability of combined eye-tracking and keystroke-logging technology being a central feature of the proposed method of data collection and analysis, new interventions may make direct use of the real-time writing-process data similar to those analyzed in the present paper, to provide for

adaptive computer-assisted learning experiences that would otherwise be impossible even in principle.

This study is not without limitations. Specifically, the differences in orthographic depth (Durgunoğlu, 2006) and language typology (morphological and syntactic) between English and Turkish were not considered, only one type of writing prompt was used, and the sample size was smaller than what the technology could have afforded. The relatively low resolution of the eye tracker should also be acknowledged as a limitation of the study. However, it is important to underline that the main purpose of this study was to illustrate the method and to demonstrate that it can provide potentially interesting data that would not be otherwise readily available. We believe that this purpose was satisfactorily achieved: This paper serves to prepare the methodological ground necessary for subsequent investigation of writing processes in L1 and L2.

Acknowledgements

References

Alves, R. A., & Limpo, T. (2015). Progress in written language bursts, pauses, transcription, and written composition across schooling. *Scientific Studies of Reading: The Official Journal of the Society for the Scientific Study of Reading*, *19*(5), 374–391.

Baba, K., & Nitta, R. (2014). Phase transitions in development of writing fluency from a complex dynamic systems perspective. *Language Learning*, *64*(1), 1–35.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, *12*(3), 267–296.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication*, *18*(1), 80–98.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *The Behavioral and Brain Sciences*, *39*, 1–72.

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

Chukharev-Hudilainen, E. (forthcoming). Empowering automated writing evaluation with keystroke logging. In E. Lindgren & K. P. H. Sullivan (Eds.), *Observing writing: insights from keystroke logging and handwriting*. Leiden, Netherlands: Brill Publishing.

Connelly, V., Dockrell, J. E., Walter, K., & Critten, S. (2012). Predicting the quality of composition and written language bursts from oral language, spelling, and handwriting skills in children with and without specific language impairment. *Written Communication*, *29*(3), 278–302.

Cummins, J. (1980). The Cross-Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Age Issue. *TESOL Quarterly*, *14*(2), 175–187.

de Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking

proficiency. *Studies in Second Language Acquisition*, *34*(01), 5–34.

De Larios, J. R., Murphy, L., & Marín, 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). Netherlands: Springer.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*(3), 359–380.

Durgunoğlu, A. Y. (2006). Learning to read in Turkish. *Developmental Science*, *9*(5), 437–439.

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.

Flower, L., & Hayes, J. (1980). The dynamics of composing : Making plans and juggling constraints. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31–50). Hillside, NJ: Lawrence Erlbaum Associates.

Frid, J., Wengelin, Å., Johansson, V., Johansson, R., & Johansson, M. (2012). Testing the temporal accuracy of keystroke logging using the sound card. In *13th International EARLI SIG Writing Conference, Porto, Portugal*. Retrieved from http://portal.research.lu.se/portal/files/5892439/3412110.pdf

Hacker, D. J., Keener, M. C., & Kircher, J. C. (2017). TRAKTEXT: Investigating writing processes using eye-tracking technology. *Methodological Innovations*, *10*(2), 1–18.

Hayes, J. R. (2012). Evidence from language bursts, revision, and transcription for translation and its relation to other writing processes. In M. Fayol, D. Alamargot, & V. Berninger (Eds.), *Translation of thought to written text while composing* (pp. 15–25). New York, NY: Psychology Press.

Hayes, J. R., & Nash, J. G. (1996). On the nature of planning in writing. In C. Levy & S. Ransdell (Eds.),

The science of writing: Theories, methods, individual differences, and applications (pp. 29–55).

Hillside, NJ: Lawrence Erlbaum Associates.

Hennessey, C., & Duchowski, A. T. (2010). An open source eye-gaze interface: Expanding the adoption

of eye-gaze in everyday applications. *Proceedings of the 2010 Symposium on Eye-Tracking

Research and Applications*. Retrieved from http://dl.acm.org/citation.cfm?id=1743686

Kim, Y., & Yoon, H. (2014). The use of L1 as a writing strategy in L2 writing tasks. *GEMA Online

Journal of Language Studies*, *14*(03), 33–50.

Leijten, M., & Waes, L. V. (2013). Keystroke logging in writing research. *Written Communication*, *30*(3),

358–392.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3),

387–417.

Levelt, W. (1999). Producing spoken language: A Blueprint of the speaker. In C. Brown & P. Hagoort

(Eds.), *The neurocognition of language* (pp. 83–122). Oxford, UK: Oxford University Press.

Lindgren, E., Spelman Miller, K., & Sullivan, K. P. H. (2008). Development of fluency and revision in L1

and L2 writing in Swedish High School years eight and nine. *ITL International Journal of Applied

Linguistics*, *156*, 133–151.

Olive, T. (2014). Toward a Parallel and Cascading Model of the Writing System: A Review of Research

on Writing Processes Coordination. *Journal of Writing Research*, *6*(2), 173–194.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, *30*(4),

590–601.

Perkins, J. (2010). *Python text processing with NLTK 2.0 Cookbook*. Birmingham, UK: Packt Publishing

Ltd.

Ransdell, S., Arecco, M. R., & Levy, C. M. (2001). Bilingual long-term working memory: The effects of

working memory loads on writing quality and fluency. *Applied Psycholinguistics*, *22*(1), 113–128.

Révész, A., Ekiert, M., & Torgersen, E. N. (2016). The effects of complexity, accuracy, and fluency on

communicative adequacy in oral task performance. *Applied Linguistics*, *37*(6), 828–848.

Schoonen, R., Gelderen, A. van, Glopper, K. de, Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M.

(2003). First Language and Second Language Writing: The Role of Linguistic Knowledge, Speed of

Processing, and Metacognitive Knowledge. *Language Learning*, *53*(1), 165–202.

Simpson, S., & Torrance, M. (2007). EyeWrite (Version 5.1).

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign

language performance. *Language Teaching Research*, *1*(3), 185–211.

Spelman Miller, K. (2000). Academic writers on-line: investigating pausing in the production of text.

Language Teaching Research, *4*(2), 123–148.

Torrance, M. (2012). EyeWrite—A tool for recording writers' eye movements. *Learning to Write

Effectively: Current Trends in European Research*, *25*, 355.

Torrance, M. (2015). Understanding planning in text production. In C. A. MacArthur, S. Graham, & J.

Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 72–87). New York, NY: Guilford Press.

Torrance, M., Johansson, R., Johansson, V., & Wengelin, Å. (2016). Reading during the composition of

multi-sentence texts: an eye-movement study. *Psychological Research*, *80*(5), 729–743.

Torrance, M., Rønneberg, V., Johansson, C., & Uppstad, P. H. (2016). Adolescent Weak Decoders

Writing in a Shallow Orthography: Process and Product. *Scientific Studies of Reading: The Official

Journal of the Society for the Scientific Study of Reading*, *20*(5), 375–388.

Uppstad, P. H., & Solheim, O. J. (2007). Aspects of fluency in writing. *Journal of Psycholinguistic

Research*, *36*(2), 79–87.

van Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. *Human Movement Science*, *10*(2),

165–191.

Van Waes, L., & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing

fluency applied to L1 and L2. *Computers and Composition*, *38, Part A*, 79–95.

Wengelin, A., Torrance, M., Holmqvist, K., Simpson, S., Galbraith, D., Johansson, V., & Johansson, R. (2009). Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior Research Methods*, *41*(2), 337–351.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. Honolulu, HI: University of Hawaii Press.

Wolfersberger, M. (2003). L1 to L2 writing process and strategy transfer: A look at lower proficiency writers. *TESL-EJ*, *7*(2), 1–12.
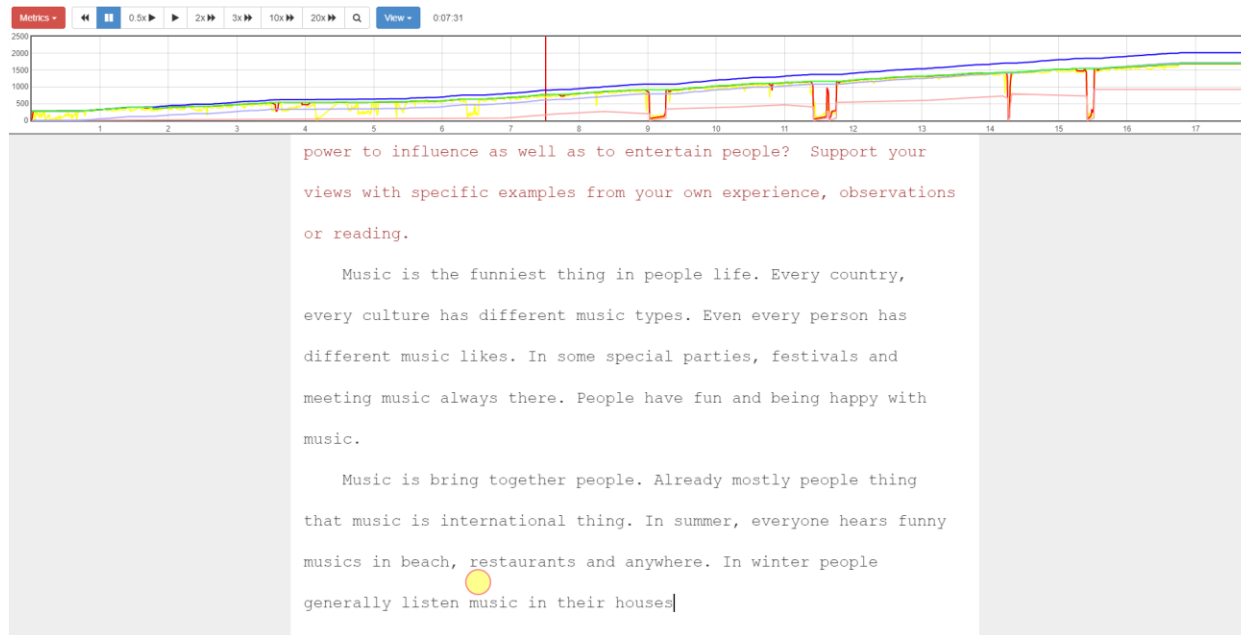
Figure 1. Process graph visualization

Music is impthe funniest thing in people life. Every country-, every culture even every person haslike different mucsic has differen-t musics ty types. Even every person is has different music likjees. IIn some partiespecial parties, fests and ivals and meetingsng becmusic always there. People have fun and nd being happy with music. I think

Music is ebring together people. INIn the some aLAldready allmostly people thigng that music is International thing. INn summer, eveeveryone hears funny musics in beach restoraaurants and soanywhere-. In winter people generally helisten music in their hoursses and everyin sipring people started to go out and in streets we can heard musics. When people heard thisese musicas IIn INin my opiinonnion they stabecome more enerjtjytscgytic. BoNOn the otjeher hand we can look different point. For example, ehehere is a girl who broke up with her boyfriend. Sge-he heardas some kidnd of music-, she becomes a unhappy. SOo, music cghahanges our mode. WQe cane can not say that music everyeverytime music is oois become good things. I have an experiences about that. One day I and my father went to celebrate. ItIHEThe place was very good but the musscic was vso bad. FFirstly, we liked plACE ace and to but after obehalf or one ourhour we got bore bored and we wanted to go homnee. This isexample explains what music do to te how music influcence peos people mode.

As a resduult, music mostluy brings good memoriesmode but sometimess it bebr brings bas-d mode-. This is inrTIt edepennds on people memories- and which place they people in. Solf people In pary-ty or festivals probabyly they bewhen thet-y heard music het-y become good mode but some situation can be different.

Tek Teknoloji hepimizin hayatında uzun süredir var olan bir durumuzun süredir hayatımız bulunuyor. Dünyamız geliştikçe teknoloji iyi eye aa da kötüğ nimetlerini bize daha fazla sunuyor. Teknolojiyle birlikte hayatımıza giren ve belki herher dakika kulolanlandığımız internet, internet aracılığıyla da sosyal medyayla tanışmış oluyoruz. Sosyal medya üezerinde yaşadığımız dünyadan kat kat büyük öggörünüyor gözümüze. Okulda, işteI-, sokakta iletişim kurduğumuz insan sayısından fazla sosyal medyadan edindiğimiz arkadaşlarımız oluyor.

Sosayal nmeemedya arkadaşlıklarına dair bir çok fikir akımı varç-. Kimi insanlar bunun nornmal ve çağımıza uygun olduğunu düşünürken, kimi insanlar yanlış ve çarpık buluyor. Benim fikrime gel geşleve deneyimlerime gekelecek olursak, sosyal emedyadan edinilen arkadaşlıkları doğru bulmuyorum. YÜZ üz yüze iletyişişimin minsanları anlamak ve aanaliz etmekte hala bir numara olduğğpunu düşünüyorum. Internet ve süzerindeinden sosyal medyada ta da da tanıştığımız isnsan, bize kedndini dilediği gibi anlatabilir. Yalan syöyleme alışkanlığıe varsa-yalana karşıymış gibi, değişken bir ruh hali içindeyse sabit tutumlu gibi gösterebilir. Kişiyi tanıma da en iyi gyöntemin hbakışlar ve vücut dili oloduğuduğunu düşünüyorum. İnternet üzerinden tanıdığımız kişilerin yalnızca kendileriinnin biz bize anlattıkları şekilde tanıyabiliriz. Bu da bize doğru sonuç vermez. Geçen yıllarda atanıdığpım bir arkadaşım sinternet üzerinden bir arkadaşlık kurrmuştu. Tanıştığı kişi ona fiziksel özelliklerini, aslında olduğundan daha iyi tanıtm farklı tanıtmıştı. Ayerırıca ekendini üniversitre e meznunu ve çalışmakta olan bir birey olarak tanımlamıştı. Fakartt arkadaşım rastlandtı eseri onun kendsisi hakkında yanlış şeyler söylediğini saptadı. SEğer bunu farketmeşeydi büyük bir yan kendini yanlış bir romantkik ilişkiinnin iiçinde bulabilirdi. Internet arkadaş edGördüğüm bu olay bana zaten yanlış olduğunu düşündüğüm internet arkadaşlıklarının ge bir kere daha yanIsakıncalı olduğunu gösterdi.

Anlattıklarımı ölzeezetleyecek olursam, ibnnternmeet ve sosyal medya üzerinden kurulan arkadaşlık ilişkilerinin samimiyetten ve gerçeklikten uzak olduğunu, kişiyi hayal kırıklığına sürükyeleleyebileecceğini düşünüyorum.

Figure 2. Pause analysis visualization of texts produced by the same participant in L2 (English) and L1 (Turkish)

Some people have said that music not only entertains us but changes how we think and feel about ourselves. Do you think music has the power to influence as well as to entertain people? Support your views with specific examples from your own experience, observations or reading.
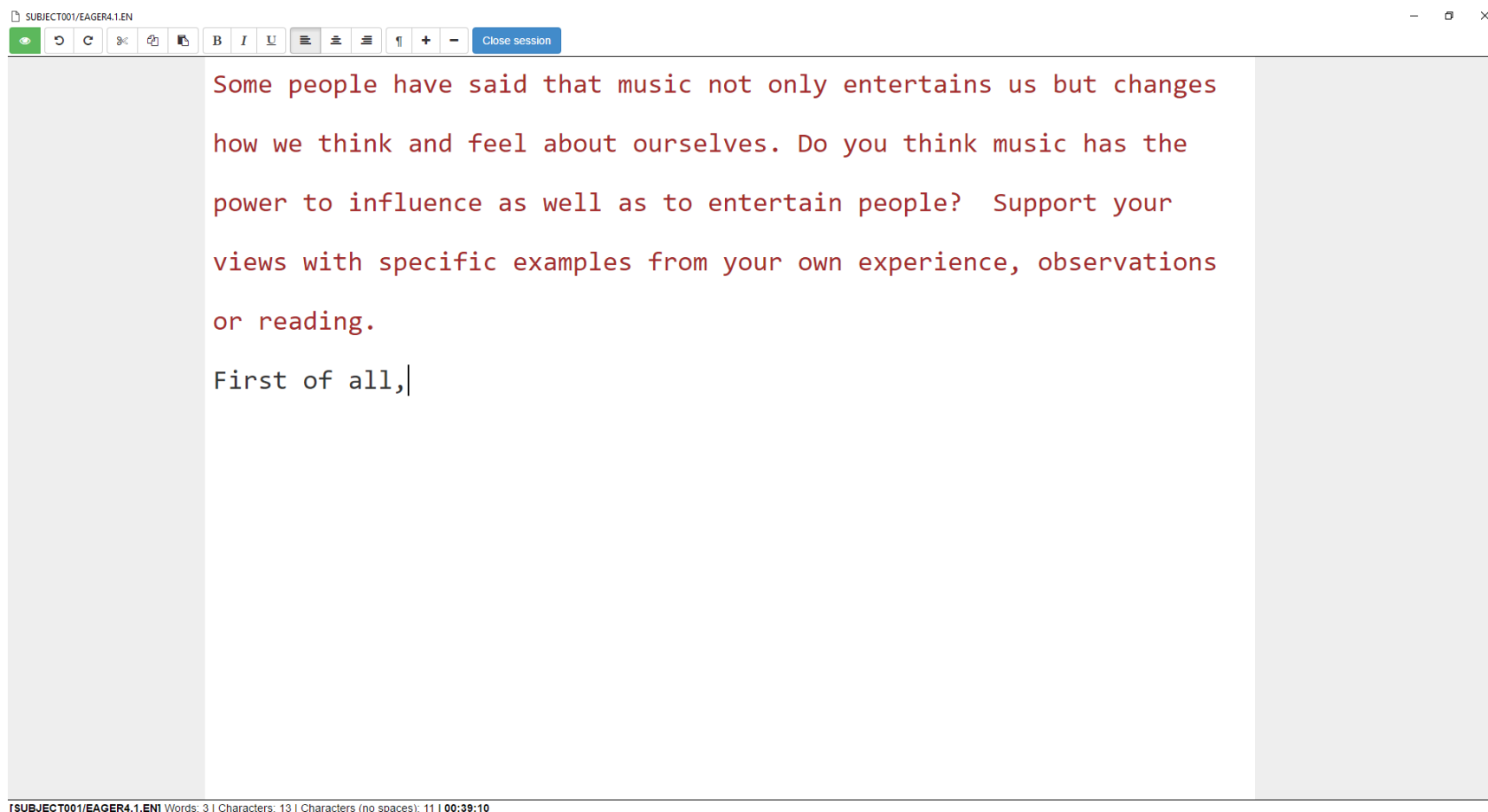
First of all,

Figure 3. Screenshot of the CyWrite interface