

UDC 81'322

DOI: 10.18413/2313-8912-2024-10-4-0-8

Tatiana A. Litvinova¹ 
Viktoriya A. Molchanova² 

**Investigating between-word pause duration
in Russian typed texts using mixture modeling based
on keystroke data**

¹ Voronezh State Pedagogical University,
86 Lenin St., Voronezh, 394043, Russia
E-mail: centr_rus_yaz@mail.ru
ORCID: 0000-0002-6019-3700

² Voronezh State Pedagogical University,
86 Lenin St., Voronezh, 394043, Russia
E-mail: zva0604@yandex.ru
ORCID: 0000-0001-6484-4606

Received 07 October 2023; accepted 15 December 2024; published 30 December 2024

Abstract: Keystroke logging is an objective and scalable methodology that has become the gold standard in writing research for modeling writing processes. A particularly significant aspect of this analysis is the examination of features such as pause duration, as pauses are regarded as indicators of underlying cognitive processes. Traditionally, arbitrary pause thresholds that are universally applied to all writers have been established to differentiate between cognitive and non-cognitive pauses. However, this approach presents considerable limitations and fails to account for the complexity and individual variability inherent in the cognitive processes involved in text production. Furthermore, different scholars employ varying approaches to the calculation of between-word pauses. This study is the first to analyze keystroke logs of Russian typed texts utilizing Gaussian mixture models (GMM) to cluster pause duration values at between-word boundaries. By employing keystroke logs collected from 50 university students who described the views from their home windows, we conducted a cluster analysis of pause duration values before words, after words, and between words separately. It was determined that the distribution of pauses between words cannot be characterised by a single distribution. For the majority of participants, two-component distribution provided the best fit for all three types of pauses. Additionally, we observed a high degree of individual variability in the mixing proportions of different components. This paper underscores the necessity of avoiding the imposition of fixed thresholds in pause analysis that are universally applicable to all writers and advocates for individualized and holistic approach to studying the writing process.

Keywords: Keystroke logging; Keystroke analysis; Pause; Pause duration; Writing research; Writing processes; Mixture modelling.

Acknowledgements: The authors acknowledge the support of the Ministry of Education of the Russian Federation (the research was supported by the Ministry of Education of the Russian Federation within the framework of the State Assignment in the field of science, topic number QRPK-2024-0011).

How to cite: Litvinova, T. A., Molchanova, V. A. (2024). Investigating Between-Word Pause Duration in Russian Typed Texts Using Mixture Modeling Based on

Keystroke Data, *Research Result. Theoretical and Applied Linguistics*, 10 (4), 147-166. DOI: 10.18413/2313-8912-2024-10-4-0-8

Investigating Between-Word Pause Duration in Russian Typed Texts Using Mixture Modeling Based on Keystroke Data

Background	Problem	Results
How do we produce written texts? Keyloggers allows to observe the process of text production	Usually, only cognitive pauses are considered (i.e. pauses above threshold, typically 2 sec) Individual variability is not taken into account	Pauses before words are longer than pauses after words 2-component model shows better fit for most of the texts
Which cognitive processes are involved in written text production? Pauses in writing – up to 75 % of the time of text production Pauses are window to the cognitive processes underlying language production	Solution Data-driven approach Mixture modelling using R How many clusters are there? Pilot study of the process of production of Russian monological texts using keystroke data	1 component - mean 364 ms (most of the between-word pauses), lexical access 2 component – mean 1689 ms (minimal discourse unit construction + reflective thoughts – tail of distribution) High individual variability! All types of pauses should be considered Individual pause thresholds are justified



УДК 81'322

DOI: 10.18413/2313-8912-2024-10-4-0-8

Литвинова Т. А.¹ 
Молчанова В. А.² 

Исследование межсловных пауз в текстах на русском языке по данным кейлоггера с использованием моделей гауссовой смеси

¹ Воронежский государственный педагогический университет
ул. Ленина, 86, Воронеж, 394043, Россия
E-mail: centr_rus_yaz@mail.ru
ORCID: 0000-0002-6019-3700

² Воронежский государственный педагогический университет
ул. Ленина, 86, Воронеж, 394043, Россия
E-mail: zva0604@yandex.ru
ORCID: 0000-0001-6484-4606

Статья поступила 07 октября 2024 г.; принята 15 декабря 2024 г.;
опубликована 30 декабря 2024 г.

Аннотация: Регистрация нажатий клавиш с использованием специальных программ – кейлоггеров – это неинвазивная технология, которая стала золотым стандартом в моделировании процессов порождения текста. Особенно важными характеристиками в таком анализе являются длительности пауз, поскольку именно паузы рассматриваются как индикаторы базовых


когнитивных процессов, лежащих в основе процесса порождения текста. Распространенным подходом является установление произвольных пороговых значений пауз, и разделение на их основе пауз на когнитивные, которые и используются для дальнейшего анализа, и некогнитивные, которые исключаются из дальнейшего анализа как незначимые. Однако такой подход имеет множество недостатков и не позволяет охватить сложность и индивидуальную вариативность когнитивных процессов, связанных с текстопорождением. В статье представлены результаты пилотного исследования, проведенного на основе данных кейлоггера в ходе порождения текстов на русском языке. В исследовании для кластеризации значений продолжительности пауз используются модели гауссовых смесей. Мы обнаружили, что паузы между словами не могут быть охарактеризованы одним распределением. Для описания межсловных пауз в текстах большинства наших участников лучше всего подходит двухкомпонентное распределение, отражающее, вероятно, лексический доступ и рефлексивные процессы. Мы обнаружили высокую индивидуальную вариативность пропорций для выявленных компонентов. В статье подчеркивается необходимость использования индивидуального подхода к установлению паузальных критериев, а также исследования пауз различной продолжительности в их совокупности и взаимосвязи на уровне отдельного текста.

Ключевые слова: Анализ нажатий клавиш; Кейлоггер; Пауза; Продолжительность пауз; Исследования письма; Процесс письма; Модели гауссовой смеси

Информация об источниках финансирования или грантах: Авторы благодарят за финансовую поддержку Министерство просвещения Российской Федерации (работа выполнена при финансовой поддержке Министерства просвещения Российской Федерации в рамках выполнения государственного задания в сфере науки, номер темы QRPK-2024-0011).

Информация для цитирования: Литвинова Т. А., Молчанова В. А. Исследование межсловных пауз в текстах на русском языке по данным кейлоггера с использованием моделей гауссовой смеси // Научный результат. Вопросы теоретической и прикладной лингвистики. 2024. Т. 10. № 4. С. 147-166. DOI: 10.18413/2313-8912-2024-10-4-0-8

ИССЛЕДОВАНИЕ МЕЖСЛОВНЫХ ПАУЗ В ТЕКСТАХ НА РУССКОМ ЯЗЫКЕ ПО ДАННЫМ КЕЙЛОГГЕРА С ИСПОЛЬЗОВАНИЕМ МОДЕЛЕЙ ГАУССОВОЙ СМЕСИ

Контекст	Проблема	Результаты
Как мы порождаем письменные тексты?	Как правило, исследуются когнитивные паузы (паузы определенной длительности, обычно от 2 секунд)	Паузы до слова длиннее, чем паузы после слова
Кейлоггеры позволяют наблюдать за процессом порождения текста	Индивидуальная вариативность не учитывается	2-компонентная модель лучше всего описывает данные (межсловные паузы) для большинства текстов
Какие когнитивные процессы лежат в основе?	Решение	1 компонент - среднее 364 мс (большинство межсловных пауз), лексический доступ
Паузы составляют до 75 % процесса порождения текста	Подход от данных	2 компонент – среднее 1689 мс (minimal discourse unit construction + reflective thoughts – tail of distribution)
Паузы – окно в когнитивную деятельность автора текста	Модели гауссовой смеси, библиотека на языке R	Высокая индивидуальная вариативность!
	Сколько кластеров наших данных?	
	Пилотное исследование процесса порождения монологических текстов на русском языке с использованием данных кейлоггера	Индивидуальные паузальные критерии



Introduction

Keystroke logging software, which captures information regarding each key pressed and released along with timing data during text production via keyboard has gained prominence in writing research for the examination of the writing process and the cognitive mechanisms that underlie it. This trend has been evident since the introduction of academic keyloggers to research community (Leijten, Van Waes, 2006; Leijten, Van Waes, 2013). Keystroke logging has emerged as a gold standard methodology in writing research because, unlike other methods such as screen recording, thinking-aloud protocols, eye-tracking, it offers objective, detailed and easily obtainable data on typing processes during text production, which can be employed non-invasively in naturalistic settings.

One of the most significant types of information yielded by keystroke logging software pertains to the location, duration, and distribution of pauses that occur during the writing process. In the field of writing research pauses are defined as intervals during which no written output is generated (Garcés-

Manzanera, 2024), as periods of inactivity, observable and measurable (Barkaoui, 2019). Pauses in writing have been “assumed to provide us with a window to the cognitive processes underlying language production” (Wengelin, 2006: 108).

In the field of writing research, written composition is often conceptualized as a series of temporal segments, referred to as “bursts” (a minimal units of text production whose linguistic characteristics remain to be fully elucidated) (Alves, Limpo, 2015; Limpo, Alves, 2017) divided by pauses. Various studies suggest that pauses can account for three-quarters of a text production time (Alamargot et al., 2007). This observation is not surprising, given that writing is an inherently complex activity; writers engage in planning, monitoring, revising, and continuously evaluating their work. These processes undoubtedly impose a considerable cognitive load, impacting both attentional and working memory capacities (Alamargot et al., 2007; Medimorec, Risko, 2017).

Pause data have been extensively employed to infer the cognitive processes that

underpin writing since the inception of writing research. It is evident that not all pauses that occur during writing are regarded as indicators of cognitive activity; rather, only those pauses deemed indicative of higher-order thinking are taken into account. A general approach to distinguishing such pauses from “non-cognitive” involves the establishment of a specific threshold.

Classic cognitive models of writing emphasize the reflective processes that occur during cognitive pauses. Researchers in this field examine the locations and distribution of these pauses in relation to various factors, including the writing tasks, the characteristics of the writer, and the quality of the text, etc. The body of research investigating the different characteristics of bursts (Cislaru et al., 2024) is significantly influenced by the cognitive pause definition, as bursts are usually defined as units occurring between two cognitive pauses.

Despite decades of intensive research, little is still known about the exact nature of the cognitive processes underlying written production during “cognitive” pauses, and the results of existing studies are often difficult to compare, as researchers employ different pause duration thresholds to identify cognitive pauses. Currently, there is no objectively and universally accepted pause threshold (Medimorec, Risko, 2017; Galbraith, Baaijen, 2019). While various thresholds are used, the 2000 milliseconds (ms) threshold is widely employed adopted, with pauses above this duration considered indicators of higher-level cognitive processes such as planning new ideas and revising (Wengelin, 2006; Chukharev-Hudilainen, 2014; Chukharev-Hudilainen et al. 2019; Garcés-Manzanera, 2024). A pause lasting between 30 milliseconds and 2 seconds is believed to indicate transcription processes related to typographic skills and spelling, grammatical and lexical decisions (Limpo, Alves, 2017; Spelman-Miller, 2006; Valenzuela, Castillo, 2023). However, it is important to note that

interpretation and linking keystroke logging variables to underlying cognitive processes is not straightforward and easy (Baaijen, Galbraith, 2018; Galbraith, Baaijen, 2019). Overall, the primary focus of writing research utilizing keystroke data has been on higher-level thinking processes rather than on more localized ones.

There are numerous challenges associated with establishing a universal pause threshold. First, this approach does not consider individual differences in typing skills and potentially in the cognitive processes that underlie writing and writing styles (Escorcia et al., 2017; Vandermeulen et al., 2024). Second, it overlooks shorter pauses that could be beneficial for a detailed analysis of the writing process (see Galbraith, Baaijen, 2018 for the dual-process model of writing).

To address the significant limitations of an approach that relies on assigning a predefined and universal pause duration threshold, Baaijen et al. (2012) proposed the use of a mixture models methodology (McLachlan, Peel, 2000) to identify the subcomponents of pause duration distributions. Mixture modeling is a form of cluster analysis that allows researchers to evaluate how many subcomponents – whether of the same type or different types – can be distinguished within an analyzed distribution. This method aims to achieve a better alignment between pause duration values and the underlying cognitive processes. The results obtained by Baaijen et al. (2012) revealed numerous advantages of application of mixture modeling for clustering pauses duration values, and the authors recommend it as a standard practice to analyze pause behavior during writing. However, since the publication of Baaijen et al. (2012) only few studies have employed this method to analyze pauses during writing (Guo et al., 2018; Roeser et al., 2019; Van Waes et al., 2021).

In recent research conducted by Hall et al. (Hall et al., 2024) mixture modeling was employed to cluster pauses at various

locations within texts. Based on the results obtained, the authors assert that traditional threshold-based approaches to pause analysis “fail to capture the complexity of the cognitive processes involved in text production” (Hall et al., 2024). They conclude that pauses at different text locations cannot be adequately described by a single distribution and recommended that researchers “impose a common set of theoretically informed distributions” on pause duration data, rather than impose fixed pause thresholds. However, the aforementioned research was conducted on a small-scale corpus of English texts, highlighting the need for further studies on diverse text types and languages, as emphasized by the authors.

Another problem of analysing pauses in writing studies is related to the calculation of pause duration values. Typically, a pause between words is calculated as a sum of two distinct pauses: one occurring before and one after SPACE. However, these pauses are sometimes treated separately (e.g., Wengelin, 2006), yet the underlying processes occurring during each type of pause remain largely unknown. In the aforementioned work by Hall et al. (Hall et al., 2024) it is asserted that combining these two types of pauses is reasonable; however, no theoretical justification for this combination is provided. As Medimorec and Risko aptly argue, “a potential limitation of this approach is that it implies functional similarity between “after” and “before” pauses” (Medimorec, Risko, 2017: 56). Nevertheless, this issue has received comparatively attention in research. The authors (Medimorec, Risko, 2017) emphasize the importance of investigating the roles of these two types of pauses in text production separately to uncover potential functional differences between them. To the best of our knowledge, there has been no research conducted to reveal such differences to date. It is clear that without modeling the distribution of pauses before and after SPACE

separately, in addition to considering the combined factor, these distinctions would remain unobservable.

The current study is the first to analyze the distribution of between-word pauses in the log files of typed texts in Russian, utilizing mixture modeling to uncover the subcomponents of pause time distribution.

Our main research questions are as follows:

1. How many clusters of pause duration values can be identified among pauses between words?
2. Is the structure of pause duration values differ for pauses occurring after words compared to those before words?
3. If more than one cluster can be identified, do the pause duration thresholds apply to all writers?

Methodology

Participants

All respondents were native Russian speakers. A total of 53 students participated in our experiment and provided written informed consent for their data and texts to be processed for research purposes. For this study, we selected only those participants who wrote texts of all three types resulting in a final sample of 50 participants.

Out of the 50 participants, 48 (94 %) were female, all participants were aged 18 to 20 years old.

The data utilized in this study were collected during a general academic course titled “The Russian Language and Language Culture” for Russian students enrolled in the bachelor’s degree program in pedagogical education in Voronezh State Pedagogical University.

Design

The participants received an electronic version of the questionnaire, which included three writing tasks, demographic questions, and links to psychological tests. They were

encouraged to ask any questions about the experiment details and the tasks assigned to them via Telegram chat with the instructor.

For the first writing assignment, the students were required to compose a detailed description of the view from the window of their home. The second assignment involved writing about their impressions of any event they had attended. For the third text, the students were asked to discuss a topic of personal interest.

Several writing tasks were selected, as it is known that the type of writing task influences the pausing behavior. For instance, research conducted by Medimorec and Risko demonstrated that argumentative essays are more cognitively demanding – being more constrained and requiring greater planning – compared to narratives (Medimorec and Risko, 2017).

No more than one text should be written in a single day; however, the order in which the writing assignments are completed is restricted. Each text must be at least 150 words long. Additionally, it is important to consider that several days should elapse between completion of one text and the writing of the next text.

Before writing each text, the students were required to complete an emotional state assessment consisting of 20 questions (<https://psytests.org/emo/panas.html>). Additionally, during the course of their participation in the study, the students completed the Big Five Questionnaire – a test designed to identify personality traits –

consisting of 44 questions (<https://psytests.org/big5/bfi.html>).

Tests were collected as part of a larger project aimed at assessing the effect of participants' characteristics on their pause behavior during text production.

The study participants provided the finished materials to the moderator via e-mail or Telegram. Each respondent was assigned an individual code. As a result, the anonymous study materials were transferred to the project manager.

All the students used their personal computers for the task. The participants were allowed to use the Internet during the task but were prohibited from copying and pasting texts or their fragments. Students were not restricted in terms of time on the task, but we asked them to write one text per writing session, without interruption.

Before starting to write the texts, the participants had installed the academic keylogger GenoGraphiX-Log 2.0 (abbreviated GGXLog) on their personal computers (Usoof et al., 2020). GGXLog records writing sessions in different writing contexts and stores the data from the writing session as a log file. Furthermore, GGXLog captures and stores informant data, third-party application usage during the writing session, and the final product text.

This software was chosen as the only academic keylogger currently available for download in Russia.

The example of the fragment of the log file provided by GGXLog is presented in Figure 1.

Figure 1. Example of the output of GGXLog keylogger
Рисунок 1. Пример выдачи кейлоггера GGXLog

Index	Device	Input	Start_time	Start_clock	End_Time	End_clock	Event_time	Pause_time	Pause_location	Caret_position	Anchor_position	Mouse_x	Mouse_y	Selected_text	Replaced_text	Gene_type
0	<MOUSE>	<PRIMARY>	0	00:00:00	74	00:00:00	74	687	initial	0	0	114	36			navigate
1	<KEYBOARD>	SHIFT	3884	00:00:03	4478	00:00:04	594	3810	unknown	0	0					unknown
2	<KEYBOARD>	Я	4342	00:00:04	4437	00:00:04	95	458	sentence beginning	0	0					append
3	<KEYBOARD>	BACK_SPACE	5986	00:00:05	6053	00:00:06	67	1549	delete	0	0					delete
4	<KEYBOARD>	BACK_SPACE	6139	00:00:06	6193	00:00:06	54	86	delete	0	0					delete
5	<SYSTEM>	<SESSION_PAUSE>	24336	00:00:24	24336	00:00:24	0	18143	unknown	0	0					unknown
6	<MOUSE>	<PRIMARY>	29732	00:00:29	29818	00:00:29	86	5396	navigation	0	0	131	64			navigate
7	<KEYBOARD>	SHIFT	30824	00:00:30	30854	00:00:30	230	806	unknown	0	0					unknown
8	<KEYBOARD>	Я	30716	00:00:30	30833	00:00:30	117	92	unknown	0	0					append
9	<KEYBOARD>	и	31464	00:00:31	31492	00:00:31	28	631	unknown	1	1					append
10	<KEYBOARD>	BACK_SPACE	32192	00:00:32	32256	00:00:32	64	700	delete	1	1					delete
11	<KEYBOARD>	SPACE	32957	00:00:32	33032	00:00:33	75	701	between words	1	1					append
12	<KEYBOARD>	и	33093	00:00:33	33187	00:00:33	94	61	unknown	2	2					append
13	<KEYBOARD>	н	33197	00:00:33	33257	00:00:33	60	10	unknown	3	3					append
14	<KEYBOARD>	о	33343	00:00:33	33401	00:00:33	58	86	unknown	4	4					append
15	<KEYBOARD>	г	33482	00:00:33	33546	00:00:33	64	81	unknown	5	5					append
16	<KEYBOARD>	д	33689	00:00:33	33757	00:00:33	68	143	unknown	6	6					append
17	<KEYBOARD>	а	33769	00:00:33	33769	00:00:33	50	50	unknown	7	7					append
18	<KEYBOARD>	SPACE	33813	00:00:33	33891	00:00:33	78	44	unknown	8	8					append
19	<KEYBOARD>	о	33974	00:00:33	33989	00:00:33	15	83	unknown	9	9					append
20	<KEYBOARD>	ч	33998	00:00:33	34077	00:00:34	79	9	unknown	10	10					append
21	<KEYBOARD>	е	34106	00:00:34	34168	00:00:34	62	29	unknown	11	11					append
22	<KEYBOARD>	н	34215	00:00:34	34287	00:00:34	72	47	unknown	12	12					append
23	<KEYBOARD>	ь	35515	00:00:35	35578	00:00:35	63	1228	unknown	13	13					append
24	<KEYBOARD>	SPACE	35633	00:00:35	35688	00:00:35	55	55	unknown	14	14					append
25	<KEYBOARD>	л	35715	00:00:35	35776	00:00:35	61	27	unknown	15	15					append
26	<KEYBOARD>	ю	35888	00:00:35	35943	00:00:35	55	112	unknown	16	16					append
27	<KEYBOARD>	б	36027	00:00:36	36093	00:00:36	66	84	unknown	17	17					append
28	<KEYBOARD>	л	36195	00:00:36	36254	00:00:36	59	102	unknown	18	18					append
29	<KEYBOARD>	ю	36356	00:00:36	36443	00:00:36	87	102	unknown	19	19					append
30	<KEYBOARD>	SPACE	36481	00:00:36	36543	00:00:36	62	38	unknown	20	20					append
31	<KEYBOARD>	э	36706	00:00:36	36772	00:00:36	66	163	unknown	21	21					append
32	<KEYBOARD>	а	36783	00:00:36	36838	00:00:36	55	11	unknown	22	22					append
33	<KEYBOARD>	с	36901	00:00:36	36991	00:00:36	90	63	unknown	23	23					append
34	<KEYBOARD>	м	37116	00:00:37	37198	00:00:37	82	125	unknown	24	24					append
35	<KEYBOARD>	о	37216	00:00:37	37216	00:00:37	50	50	unknown	25	25					append
36	<KEYBOARD>	т	37270	00:00:37	37369	00:00:37	99	54	unknown	26	26					append
37	<KEYBOARD>	р	37471	00:00:37	37538	00:00:37	67	102	unknown	27	27					append

Data Analysis and Preprocessing

For this exploratory study, we used only one text type – describing the view from the window – as we assumed this writing task would yield the most homogeneous texts. The dataset used for this study is available on GitHub¹.

We analyzed the log files (50) containing individual participants' keystroke data. In the preliminary stage, we removed outliers from our pause duration values which totaled 111948 entries. We employed the quantile method, eliminating values lower than the 0.1 percentile (2 ms) and higher than 99.9 percentile (50370.28 ms). This process resulted in a final count of 111538 pause duration values.

In the present analysis, we focused solely on pauses between words, specifically instances where a word ended and new word began after a SPACE. This approach meant that we excluded all pauses associated with punctuation marks, revision indicators (such as DELETE and BACKSPACE), and navigation markers. Consequently, we analyzed only those cases where no actions other than pause occurred between consecutive words, eliminating instances of revision, mouse movements, insertion, edits, or punctuation marks. Pauses occurring between sentences, sub-sentence pauses (as defined by Hall et al., 2024 in Table 1, which describes them as “the time between the end of a word that is followed by a comma, and the start of the next word that is preceded by the same comma”) and revision pauses (termed “non-linear events”) are typically regarded as distinct types of pauses in writing research and will be analyzed individually.

1

<https://github.com/Litvinova1984/keystroke-Russian-texts> (accessed on 15.10.2024).

In the current analysis, we analyzed three types of pauses between two concessive words: 1) pauses after words, i.e., after the last letter of preceding word before SPACE (t2 in Figure 2); 2) pauses before words, i.e.,

after SPACE mark (t3 in Figure 2); 3) combined indicator (pause time before SPACE + pause time after SPACE) (i.e., for between-word pause for word2 and word3 is calculated as t2+t3) (Figure 2).

Figure 2. Definition of pauses
Рисунок 2. Определение пауз



Although the SPACE key serves as a marker for pauses between words, the random insertion of SPACE within words is a common issue in the analysis of text production through keystroke logging. Various thresholds for the minimum duration of between-word pauses have been proposed to filter out the random occurrence of SPACE within words. For instance, Van Waes et al. (2021) employed a threshold of 30 ms in their study of typing skills. In contrast, Hall et al. (2024) used a threshold of 50 ms across all text locations to eliminate accidental SPACE transitions within words from their analysis. A manual inspection of our dataset revealed that pauses shorter than the 30 ms threshold were indicative of accidental SPACE within words rather than between words. Consequently, these instances were excluded from further analysis. The resulting dataset comprises 22,300 pauses occurring after words and before word.

Given that pause data are highly positively skewed, with most of pauses being relatively short and a minority being significantly longer, we performed natural log transformations, which is a standard practice in pause analysis (Baaijen et al., 2012).

As the primary analytical approach, we employed mixture modeling which is a form of cluster analysis that enables researchers to

assess how many subcomponents can be identified within an analyzed distribution.

Relying on the results of the studies (Baaijen et al., 2012; Hall et al., 2024) in which between-word pauses were analyzed through clustering via mixture modeling, we constrained the maximum number of components (parameter G) to test to three. Specifically, we constructed models with G=1,2,3, and subsequently selected the best model using the Bayesian Information Criterion (BIC).

Following (Hall et al., 2024), we constructed Gaussian mixture models (GMMs) using the expectation-maximization (EM) algorithm (McLachlan, Peel, 2000) to assess whether the pause times between words, after words, and before words exhibit an underlying structure that is better represented by multiple distributions rather than a single Gaussian distribution. For more details on this algorithm which is also widely used for speech data analysis we refer the reader to the work by Little et al. (Little et al., 2012). To this end, we formally evaluated the relative goodness of fit of single Gaussian distributions compared to two- and three-component models for each participant separately. Consequently, we employed the EM algorithm to fit multiple GMMs to each participant's log-transformed data, including before-word, after-word and between-word

pause duration values, for all pauses longer than 30 ms.

The analysis was conducted using in the R package Mclust version 5.4.7 (Fraley et al., 2020).

Results

Before-word and After-word pauses

Figures 3-4 show that, although log transformations reduced skewness, they did not eliminate it for most writers.

Figure 3. Histograms showing the log-transformed distributions of before-word pause times for some participants (the participants' ID are above the figures)

Рисунок 3. Гистограммы, показывающие распределение продолжительности пауз перед словами после лог-трансформации для отдельных респондентов

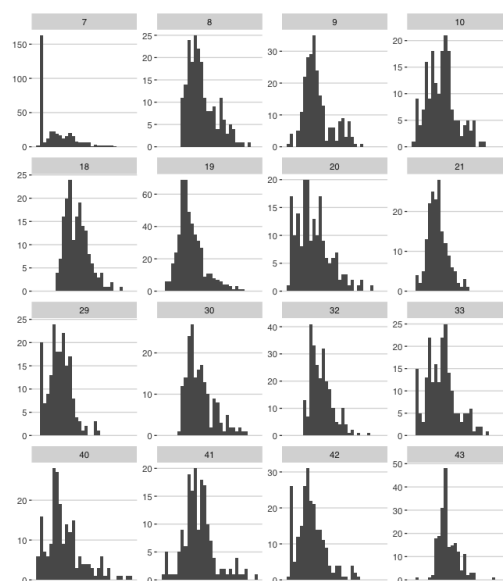


Figure 4. Histograms showing the log-transformed distributions of after-word pause times for some of the participants (the participants' ID are above the figures)

Рисунок 4. Гистограммы, показывающие распределение продолжительности пауз после слов после лог-трансформации для каждого респондента

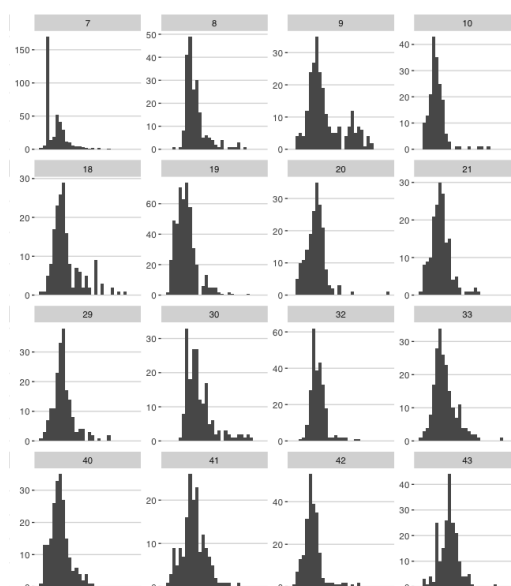


Table 1S² presents Bayesian Information Criterion (BIC) values for various Gaussian Mixture Models (GMM) with G values of 1, 2, and 3, with the best values highlighted in bold. It is important to note that we also evaluated 4-component models; however, their performance was inferior to that of the two- and three-component models (though better than the one-component model) and therefore, their results are not included.

The differences between the best and second-best models are presented in Table 1S. The best models, along with the second-best models that exhibits differences greater than 3.7 compared to the best models, are highlighted and underscored. This threshold was proposed by Kass and Raftery (Kass, Raftery, 1995: 777) who suggested that a difference of this magnitude could be considered significant, while a difference greater than 20 indicates a strong distinction between models.

As for the before-word data, the two-component model demonstrated the best fit for 21 of the respondents (42%), the three-distribution model was most suitable for 11 of them (22%), the remaining 17 writers (34%) were indeterminate between two and three-distribution models. Only for 1 respondent (2 %) the one-distribution model were the best fit.

In the analysis of the afterword pauses, we found that 28 out of 50 texts (56%) were better described by a two-component model. For 9 texts (17.65%) the three-distribution model was the best fit, and the remaining 9 texts (17.65%) being indeterminate between

two and three distributions. Only for 2 respondents (4%) the one-distribution model was the best fit. One writer was intermediate between the one- and three-distribution models, and another writer was intermediate between the one- and two-distribution models.

These results demonstrates that the cognitive processes occurring at word boundaries – both before word and after word – are heterogeneous and cannot be encapsulated by a single threshold.

Next, we constructed a two-component distribution model for pauses occurring before and after words for all writers to compare the properties of these distributions among them. Specifically, we calculated the proportion of pauses within each of the two distributions, as well as the mean values for each distribution (see Table 2S³).

Let us compare the data obtained for the first and second components. The mean duration of pauses for the first component specifically for before-word pauses is log 5.31 (202.35 ms), median value is log 5.40 (221.41 ms). The mean duration for after-word pauses in the first component is log 4.8 (121.51 ms), and median value is log 4.78 (119.1 ms).

We used the **paired samples Wilcoxon test**, also known as **Wilcoxon signed-rank test**, which is a **non-parametric** alternative to the paired t-test used. This method is employed to compare paired data due to the violation of normality in our dataset.

This test confirmed that the differences in pause duration before and after word were statistically significant ($V = 1193$, $p\text{-value} < 0.00001$) (Figure 5).

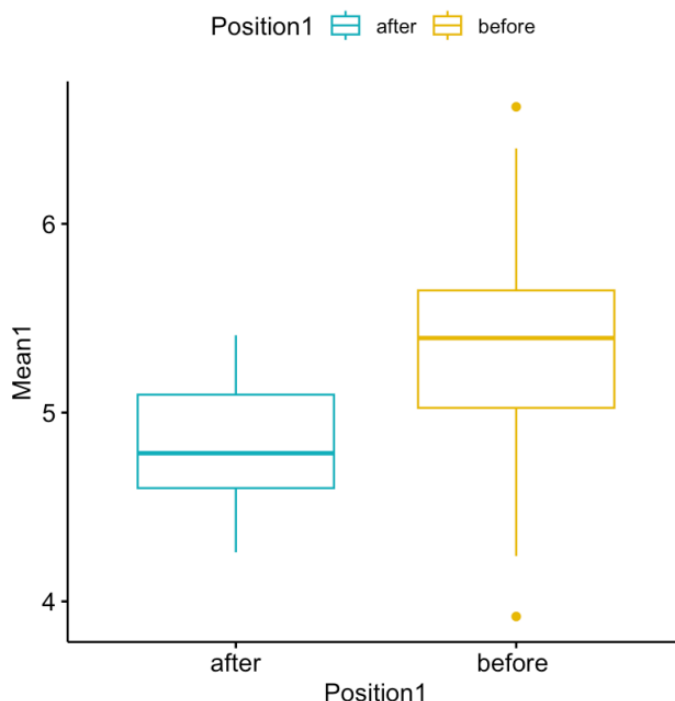
2

https://github.com/Litvinova1984/keystroke_Russian_t_exts (accessed on 15.10.2024).

3

https://github.com/Litvinova1984/keystroke_Russian_t_exts (accessed on 15.10.2024).

Figure 5. Mean pause duration (log ms) before and after words (component 1)
Рисунок 5. Средняя длительность пауз до и после слова (компонент 1)



The second component is characterized by pauses with a median duration of 6.97 log ms (1064 ms) before word, 6.46 log ms (639.06) after word (IQR = 0.85). The differences in pause duration before and after the word in the second component are statistically significant ($V = 1026$, p -value = 0.00018) (Figure 6).

Let us note that the mean mixing proportion of the first component for before-word pauses is 0.692, while for after-word pauses it is 0.827. This indicates that the share

of longer pauses (the second component) is greater in the pause position before words.

Note that the mixing proportions of components vary significantly among the individuals. While some exhibit a strong preference for the first component, accounting for more than 90% of the pauses, others show that the second component also constitutes a significant part of the pause data.

Between-word pauses

The histogram illustrating the distribution of between-word pause duration values is presented in Figure 7.

Figure 6. Mean pause duration (log ms) before and after words (component 2)
Рисунок 6. Средняя длительность пауз до и после слова (компонент 2)

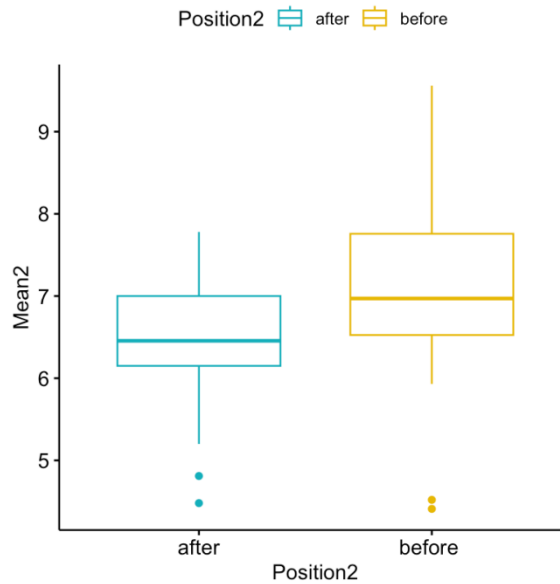
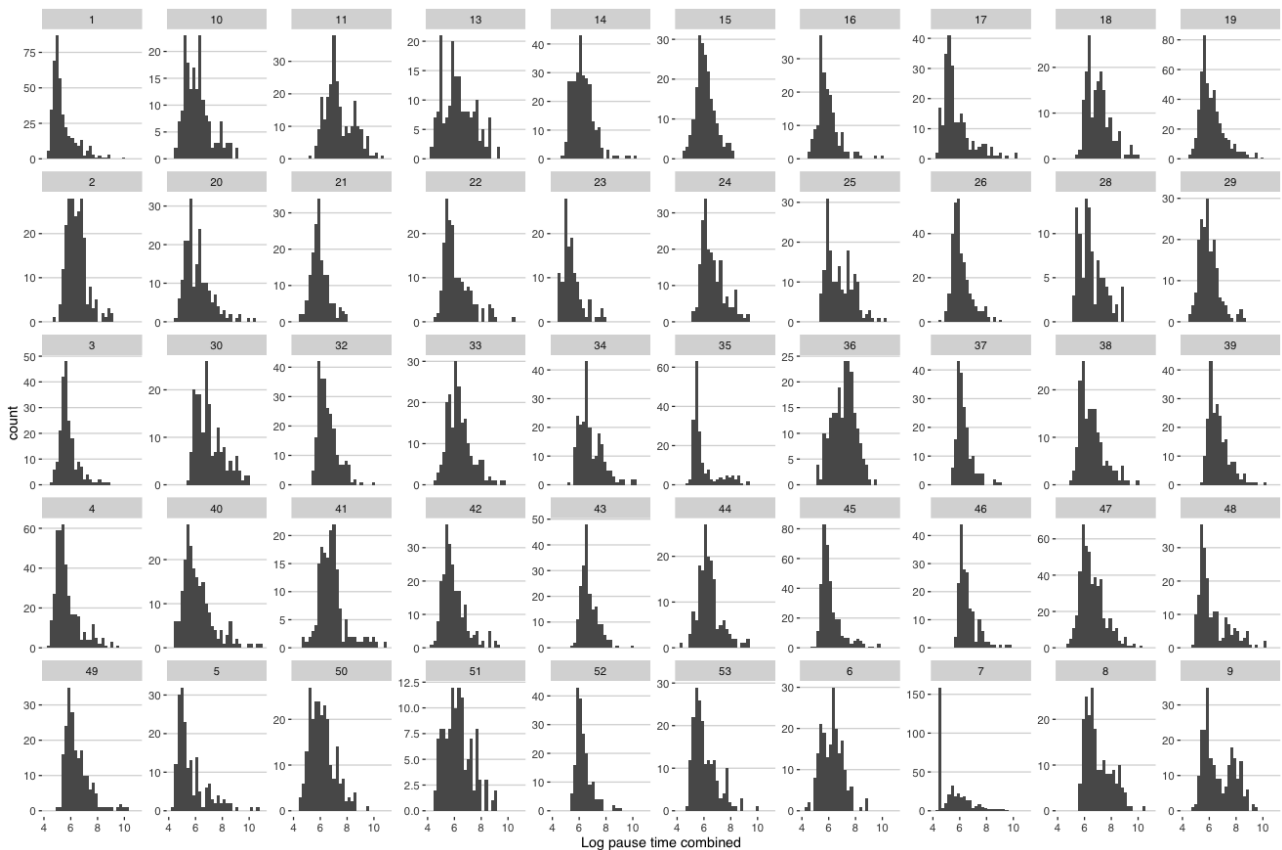


Figure 7. Histograms showing the log-transformed distributions of between-word pause times for each of the participants (participants' ID are above the figures)

Рисунок 7. Гистограммы, показывающие распределение продолжительности межсловных пауз после лог-трансформации для каждого респондента



Again, we observe that even after applying a log transformation pause time data exhibit a multicomponent distribution for the majority of the writers. The pause data from only a few respondents (NN 6 and 36) tend to display a unimodal distribution, while most of the data clearly indicate a multimodal distribution. However, specific characteristics of these distributions vary significantly among the writers. While most of the data demonstrate a two-modal distribution, pause durations for some writers are better described by a three-modal distribution (NN 34, 35, 45, 49).

We constructed GMMs with varying numbers of components ($G = 1, 2, 3$). Three-distributional models provided the best fit for the data from 5 participants (10%), while 11 (22%) were indeterminate between two- and

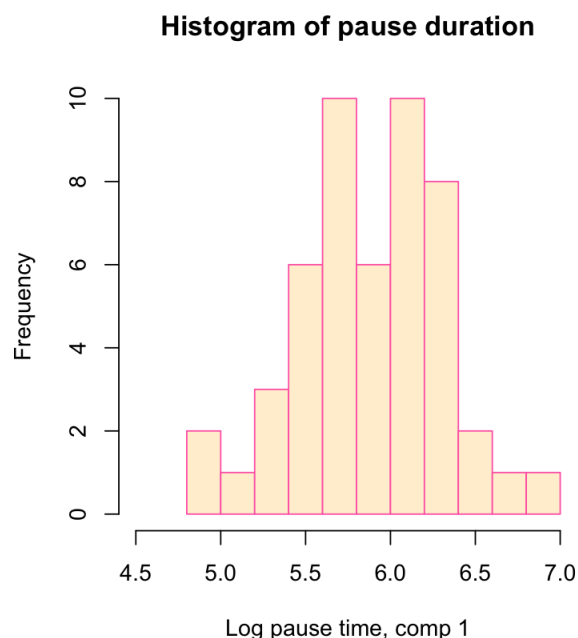
three-component model, one writer (2%) was indeterminate between one- and two-component models. One-distribution model was a best fit for two participants (4%). For the majority of the writers (31, 62%), two-distributional models demonstrated the best fit (see Table 3S⁴).

So, for the sake of comparison, we built a two-component model for all the respondents.

The mean duration of between-word pauses for each mixture component along with the mean proportion of pauses falling within each mixture component for each participant is presented in Table 4S⁵.

The mean duration of pauses between words for the first component is 5.897 log ms, which corresponds to 364 ms (median value is 5.865 log ms, i.e., 352.48 ms) (see Figure 8).

Figure 8. Histogram of between-word pause duration (log ms) for the first component
Рисунок 8. Гистограмма продолжительности межсловных пауз (лог мс) в первом компоненте



4

https://github.com/Litvinova1984/keystroke_Russian_t_exts/tree/main (accessed on 15.10.2024).

5

https://github.com/Litvinova1984/keystroke_Russian_t_exts/tree/main (accessed on 15.10.2024).

The mean mixing proportion of the first component is 0.7 (sd=0.15).

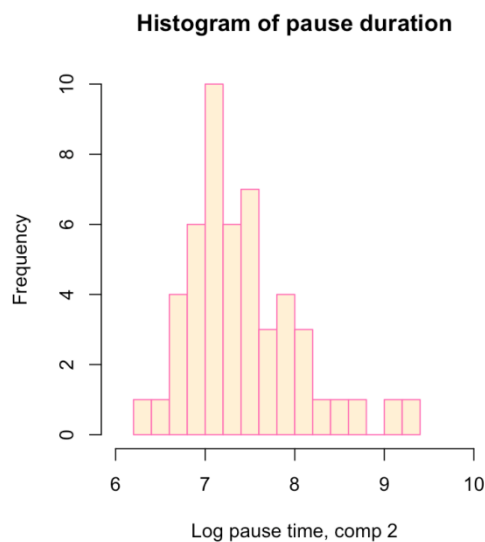
The mean duration of between-word pauses for the second component is 7.4 log ms, which corresponds to 1685.8 ms (median

value is 7.3 log ms, i.e., 1415 ms) (see Figure 9).

The mean mixing proportion of between-word pauses for the second component is 0.3 (sd=0.15).

Figure 9. Histogram of between-word pause duration (log ms) for the second component

Рисунок 9. Гистограмма межсловных пауз (ЛОГ МС) во втором компоненте



Discussion

In this study, we conducted a cluster analysis of pause duration occurring after words, before words and between words in Russian typed texts, clearly outlining our methodology for calculating the pause duration in these contexts. We utilized mixture modeling to analyze the pause time and discovered that for the majority of participants, two-cluster solution demonstrated the best fit.

We also identified differences in the duration of the pauses before word compared to those after words. These results indicate that pauses occurring before and after words reflect distinct cognitive processes, with pauses before words typically being longer. This phenomenon may be related to the increased cognitive load experienced by the writers. As emphasized in the Introduction, the challenge of elucidating the differences in the underlying cognitive processes involved in word retrieval during text production is less

explored in writing research. Nevertheless, our findings replicate those reported in (Mohsen and Qassem, 2020). Their small-scale study, which included data from eight respondents, demonstrated that pauses during the writing of two text genres – descriptive and argumentative essays – varied based on pause location: pauses before words were significantly longer than those after words for genres. Clearly, this issue warrants further investigation.

Our results concerning the multicomponent model, which demonstrated a better fit for between-word pauses than the one-component model, align with those reported in literature (Baaijen et al., 2012; Hall et al., 2024).

As our results indicate, for the majority of the participants, short pauses (up to roughly 500 ms) account for most of the pauses between words. This finding aligns with the results obtained by Baaijen et al. (2012) reported that mean duration of pauses

in the first component is 330 ms, constituting 65% of the pauses in their dataset. Our results are similar to those presented in the papers by E. Chukharev-Hudilainen, (Chukharev-Hudilainen, 2011; Chukharev-Hudilainen, 2014), who, to the best of our knowledge, is the only researcher to have analyzed keystroke data from Russian texts (specifically chat messages) in the context of the writing process. Using ex-Gaussian distribution equation, he demonstrated that two types of pause distributions could be identified. The first distribution has an average pause duration of 386.9 ± 102.9 ms and varies across the subjects, which is close to our results.

It is important to note that in our data two-component distribution was the best fit for modeling between-pause durations for the most (though not all) writers. In this regard, our results are similar to those obtained by Chukharev-Hudilainen. The second type of pauses he identified falls within the range up to 937.9 ± 357.4 ms, which are believed to be associated with the production of predicate expressions (Chukharev-Hudilainen 2014). Therefore, pauses above 1,2 sec threshold are considered indicative of higher-level planning.

In English-language data, three-component models were the best fit for the majority of the writers. A middle distribution, accounting for 26% of the pauses, with mean duration of 735 ms, was revealed by (Baaijen et al., 2012) who suggested that it represents phrase boundary processes. Baaijen et al., 2012 also revealed the right-hand distribution which comprised 9% of the pauses with a mean duration of 2697 ms, indicating a higher-level message planning or reflection. They further proposed that the “long tail” of such pauses should not be treated as a normal distribution but rather as a miscellaneous set of reflective thoughts, with cutoff of around 1.686 ms. The equivalent threshold reported by (Hall et al., 2024) is 1426 ms. These thresholds are comparable to those obtained on our data for the second distribution.

We argue that the assumption of a common set of underlying cognitive processes during writing should not be taken for granted; instead, different models may be necessary to accurately describe the writing processes of various individuals. It is also should be stressed that the relatively high standard deviations for the mixing proportions indicate substantial variation among individuals regarding the relative contributions of different component processes. This suggests the need to consider an individual factor in the analysis of writing behavior.

Our study demonstrated that, counting pauses above an arbitrarily established threshold – particularly the commonly used 2000 ms threshold – fails to encompass the range of processes involved in text production.

In our research, we found that mixture modeling has proven to be an effective method for describing writers’ pausing behavior. It allowed us to identify the relatively strong evidence for a multi-component structure in the linear between-word pause distribution in Russian monological texts for the majority of the writers.

However, the question that remains unresolved in previous studies – regarding the justification for characterizing the distribution of between-word pauses as a variable property of writers also arises in our data. While most writers exhibit two distributions, others display three.

It is essential to model pause distribution at different text locations (within-word, between-word, between-sentences) simultaneously, since these elements are clearly interrelated. This can be accomplished using multilevel mixture modeling (Muthén, Asparouhov, 2009). However, to the best of our knowledge, no research employed this method for a comprehensive analysis of a writer’s pausing behavior during text production has been performed.

It is important to acknowledge that our study has certain limitations. It was conducted

using data from highly homogeneous population, while it is known that individual characteristics, such as demographics, affect writing styles. For instance, Zhang et al. (Zhang et al., 2019) found that female authors produced their texts more fluently, engaged in more extensive macro- and local editing, and exhibited a reduced need to pause at the locations associated with planning (e.g., between bursts of text and at sentence boundaries) compared to their male counterparts. Additionally, our analysis focuses on only one text genre, which, as suggested by previous research, does not impose a high cognitive load (unlike argumentative essays, for example). Clearly, broadening the range of analyzed genres and incorporating greater diversity in the writers' characteristics would enhance the validity of our findings.

In this study, we did not consider the effect of the operating system of the participant's computer (MAC/Windows) on their writing behavior. As one of our reviewers reasonably suggested, this factor could have an impact on the results. In our future study, we will analyze this effect.

Conclusions

The use of keystroke data has become a gold standard in writing research, with pause duration at different text locations being one of the most important features analyzed based on keystroke data, since the analysis of pauses during writing is crucial for modeling the writing processes. To conduct analyses, many researchers rely on predefined pause thresholds. However, this approach has several drawbacks: it fails to account for individual differences in typing skills, writing competencies, styles, and it overlooks pauses below a predetermined threshold, which arguably hinders a comprehensive understanding of the writing process.

Our study is the first to analyze log files containing keystroke data that reflects the process of producing free monologue texts in Russian. We examined pauses before and after words separately and in combination using

mixture modeling methodology. Our findings indicate that the majority of the participants' data fit better with multicomponent (primarily two-component) distributions. The first distribution may be related to lexical assessment and depends on the motor (typing) skills of the writers, while the second distribution could signify underlying reflective thoughts related to text planning. To gain a deeper understanding of the nature of writing process, an individual-based pause threshold should be determined, and linguistic nature of the resulting burst should be examined. This field could greatly benefit from adopting methods for identifying and describing minimal discourse units in oral speech (Kibrik et al., 2020).

Our study clearly demonstrates the necessity to examine all types of pauses during writing and considering the relationships between pause durations at different locations within the text. The issue of individual variations not only regarding the thresholds for different types of pauses but also concerning the structure of distributions should be investigated across various writing tasks.

Our future research directions are manifold. First, we aim to expand our dataset in terms of both text types and the diversity of the writers' characteristics. We will pay special attention to the effects of cognitive load associated with writing tasks on writers' pausing behavior.

Second, we will conduct multilevel mixture modeling of pauses at different locations within text to gain a more comprehensive understanding of pausing behavior during text production.

Third, we will calculate pairwise semantic distance between words and align these results with pause duration to gain a better understanding of the relationship between a semantic flow and text production.

Fourth, we aim to assess the relationship between the revealed characteristics of the writers' pausing behavior and the quality of their texts, which

represents a promising avenue of writing research (Beauvais et al., 2011).

References

- Alamargot, D., Dansac, C., Chesnet, D. and Fayol, M. (2007). Parallel processing before and after pauses: A combined analysis of graphomotor and eye movements during procedural text production, in Torrance, M., Waes, L. van and Galbraith, D. (eds.). *Writing and cognition: Research and applications*, Elsevier Science, 13–29. (In English)
- Alves, R. A. and Limpo, T. (2015). Progress in written language bursts, pauses, transcription, and written composition across schooling, *Scientific Studies of Reading*, 19 (5), 374–391. <https://doi.org/10.1080/10888438.2015.1059838> (In English)
- Baaijen, V. M. and Galbraith, D. (2018). Discovery through writing: Relationships with writing processes and text quality, *Cognition and Instruction*, 36 (3), 199–223. <https://doi.org/10.1080/07370008.2018.1456431> (In English)
- Baaijen, V. M., Galbraith, D. and de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures, *Written Communication*, 29 (3), 246–277. <https://doi.org/10.1177/0741088312451108> (In English)
- Barkaoui, K. (2019). What Can L2 Writers' Pausing Behavior Tell Us About Their L2 Writing Processes?, *Studies in Second Language Acquisition*, 41 (3), 529–554. DOI: 10.1017/S027226311900010X (In English)
- Beauvais, C., Olive, T. and Passerault, J.-M. (2011). Why are some texts good and others not? Relationship between text quality and management of the writing process, *Journal of Educational Psychology*, 103 (2), 415–428. <https://doi.org/10.1037/a0022545> (In English)
- Chukharev-Hudilainen, E. (2014). Pauses in spontaneous written communication: A keystroke logging study, *Journal of Writing Research*, 6 (1), 61–84. DOI: 10.17239/jowr-2014.06.01.3 (In English)
- Chukharev-Hudilainen, E., Saricaoglu, A., Torrance, M. and Feng, H.-H. (2019). Combined Deployable Keystroke Logging and Eyetracking for Investigating L2 Writing Fluency. *Studies in Second Language Acquisition*. 41(3), 583–604. <https://doi.org/10.1017/S027226311900007X> (In English)
- Chukharev-Hudilainen, E. (2011). Local Discourse Structure of Chat Dialogues: Evidence from Keystroke Logging, *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue – Full Papers*, Los Angeles, California, 104–111. (In English)
- Cislaru, G., Feltgen, Q., Khoury, E., Delorme, R. and Bucci, M.P. (2024). Language Processing Units Are Not Equivalent to Sentences: Evidence from Writing Tasks, *Typical and Dyslexic Children, Languages*, 9 (5), 155. <https://doi.org/10.3390/languages9050155> (In English)
- Escorcía, D., Passerault, J.-M., Ros, C. and Pylouster, J. (2017). Profiling writers: Analysis of writing dynamics among college students, *Metacognition and Learning*, 12 (2), 233–273. <https://doi.org/10.1007/s11409-016-9166-6> (In English)
- Fraley, C., Raftery, A. E., Scrucca, L., Murphy, T. B. and Fop, M. (2020). Package “mclust” Title Gaussian Mixture Modelling for Model-Based Clustering, Classification, and Density Estimation, available at: <https://mclust.org.github.io/mclust/> (accessed on 05.11.2024) (In English)
- Galbraith, D. and Baaijen, V. M. (2018). The work of writing: Raiding the inarticulate, *Educational Psychologist*, 53 (4), 238–257. <https://doi.org/10.1080/00461520.2018.1505515> (In English)
- Galbraith, D. and Baaijen, V. M. (2019). Aligning keystrokes with cognitive processes in writing, in Lindgren, E. and Sullivan, K. P. H. (eds.) *Observing Writing: Insights from Keystroke Logging and Handwriting*, Brill, 38, 306–325. https://doi.org/10.1163/9789004392526_015 (In English)
- Garcés-Manzanera, A. (2024). Language bursts and text quality in digital writing by young EFL learners, *Journal of New Approaches in Educational Research*, 13. <https://doi.org/10.1007/s44322-024-00012-x> (In English)
- Guo, H., Deane, P. D., van Rijn, P. W., Zhang, M. and Bennett, R. E. (2018). Modeling basic writing processes from keystroke logs, *Journal of Educational Measurement*, 55 (2), 194–216. <https://doi.org/10.1111/jedm.12172> (In English)
- Hall, S., Baaijen, V. M. and Galbraith, D. (2024). Constructing theoretically informed measures of pause duration in experimentally

manipulated writing, *Reading and Writing*, 37, 329–357. <https://doi.org/10.1007/s11145-022-10284-4> (In English)

Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the American statistical association*, 90 (430), 773–795. (In English)

Kibrik, A. A., Korotaev, N. A. and Podlesskaya, V. I. (2020). Russian spoken discourse: Local structure and prosody. In Izre'el, S., Mello, H., Panunzi, A. and Raso, T. (Eds.). *In search of basic units of spoken language: A corpus-driven approach*, John Benjamins, Amsterdam, 367–382. DOI: [10.1075/scl.94.01kib](https://doi.org/10.1075/scl.94.01kib) (In English)

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

Leijten, M. and Van Waes, L. (2006). Inputlog: New perspectives on the logging of online writing, in Sullivan, K. P. H. and Lindgren, E. (Eds.). *Computer keystroke logging and writing: Methods and applications*, Elsevier, Oxford, 73–93. (In English)

Limpo, T. and Alves, R. A. (2017). Written language bursts mediate the relationship between transcription skills and writing performance, *Written Communication*, 34 (3), 306–332. <https://doi.org/10.1177/0741088317714234> (In English)

Little, D. R., Oehmen, R., Dunn, J. C., Hird, K. and Kirsner, K. (2012). Fluency Profiling System: An automated system for analyzing the temporal properties of speech, *Behavior Research Methods*, 45, 191–202. <https://doi.org/10.3758/s13428-012-0222-0> (In English)

McLachlan, G. J. and Peel, D. (2000). *Finite mixture models*, Wiley, New York. (In English)

Medimorec, S. and Risko, E. F. (2017). Pauses in written composition: on the importance of where writers pause, *Reading and Writing*, 30, 1267–1285. <https://doi.org/10.1007/s11145-017-9723-7> (In English)

Muthén, B. and Asparouhov, T. (2009). Multilevel regression mixture analysis, *Journal of the Royal Statistical Society Series A, Royal Statistical Society*, 172 (3), 639–657. (In English)

Mohsen, M. A. and Qassem, M. (2020). Analyses of L2 Learners' Text Writing Strategy:

Process-Oriented Perspective, *Journal of Psycholinguistic Research*, 49, 435–451. <https://doi.org/10.1007/s10936-020-09693-9> (In English)

Roeser, J., Torrance, M. and Baguley, T. (2019). Advance planning in written and spoken sentence production, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45 (11), 1983–2009. DOI: [10.1037/xlm0000685](https://doi.org/10.1037/xlm0000685) (In English)

Spelman-Miller, K. (2006). The pausological study of written language production, in Sullivan, K. P. H. and Lindgren, E. (eds.). *Computer keystroke logging: Methods and applications*, Elsevier, 11–30. (In English)

Usoof, H. A., Leblay, C. and Caporossi, G. (2020). GenoGraphiX-Log version 2.0 user guide. Les Cahiers du GERAD, available at: https://ggxlog.net/download/User%20Guide_2023_0130.pdf (accessed on 05/11/2024) (In English)

Valenzuela, Á. and Castillo, R. D. (2023). The effect of communicative purpose and reading medium on pauses during different phases of the textualization process. *Read Writ.* 36, 881–908. <https://doi.org/10.1007/s11145-022-10309-y> (In English)

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

Vandermeulen, N., Lindgren, E., Waldmann, Ch. and Levlin, M. (2024). Getting a grip on the writing process: (Effective) approaches to write argumentative and narrative texts in L1 and L2, *Journal of Second Language Writing*, 65. <https://doi.org/10.1016/j.jslw.2024.101113> (In English)

Wengelin, Å. (2006). Examining Pauses in Writing: Theory, Methods and Empirical Data, in *Computer Keystroke Logging and Writing: Methods and Applications*, Brill, Leiden, The Netherlands, 107–130. DOI: 10.1163/9780080460932_008 (In English)

Zhang, M., Bennett, R. E., Deane, P. and van Rijn, P. W. (2019). Are there gender differences in how students write their essays? An analysis of writing processes, *Educational Measurement: Issues and Practice*, 38 (2), 14–26. <https://doi.org/10.1111/emip.12249> (In English)

The authors have read and approved the final manuscript.

Авторы прочитали и одобрили окончательный вариант рукописи.

Conflicts of interests: the authors have no conflicts of interest to declare.

Конфликты интересов: у авторов нет конфликтов интересов для декларации.

Литвинова Татьяна Александровна, доктор филологических наук, заведующий научно-исследовательской лабораторией текстового психолингвистического моделирования, профессор кафедры русского языка, современной русской и зарубежной литературы Воронежского государственного педагогического университета, Воронеж, Россия.

Tatiana A. Litvinova, Doctor of Philological Sciences (D. Sc. in Philology), Head of the Textual Psycholinguistic Modelling

Research Laboratory, Professor of the Department of Russian Language, Modern Russian and Foreign Literature, Voronezh State Pedagogical University, Voronezh, Russia.

Молчанова Виктория Александровна, кандидат филологических наук, младший научный сотрудник научно-исследовательской лаборатории текстового психолингвистического моделирования, ассистент кафедры теории, истории и методики преподавания русского языка и литературы Воронежского государственного педагогического университета, Воронеж, Россия.

Viktoriya A. Molchanova, Candidate of Philological Sciences, Junior researcher in the Textual Psycholinguistic Modelling Research Laboratory, Assistant lecturer of the Department of Theory, History and Methods of Teaching Russian Language and Literature, Voronezh State Pedagogical University, Voronezh, Russia.