Известия Воронежского государственного педагогического университета. 2025. № 4 (309). С. 225–232.
Izvestia Voronezh State Pedagogical University. 2025. (4), 225–232.

Научная статья
УДК 81-112.4
DOI: 10.47438/2309-7078_2025_4_225

ВОСПРИЯТИЕ ПИСЬМЕННОГО ТЕКСТА И ПРОЦЕСС ЕГО ПОРОЖДЕНИЯ: АНАЛИЗ НА ОСНОВЕ ДАННЫХ КЕЙЛОГГЕРА

Татьяна Александровна Литвинова[1], Елена Сергеевна Котлярова[2]

Воронежский государственный педагогический университет[1, 2]
Воронеж, Россия

[1]*Доктор филологических наук, профессор кафедры русского языка, современной русской и зарубежной литературы,*
ORCID ID: 0000-0002-6019-3700, тел.: +7-473-264-44-17, e-mail: centr_rus_yaz@mail.ru
[2]*Аспирант кафедры русского языка, современной русской и зарубежной литературы,*
ORCID ID: 0000-0002-4162-5002, e-mail: kray-voronezh@.ru

Аннотация. В современных исследованиях письменной речи всё большее внимание уделяется не только анализу текста как финального продукта речепорождения, но и изучению – с использованием специальных программ – академических кейлоггеров, регистрирующих нажатия клавиш и все связанные с ними темпоральные данные – самого процесса его порождения, выявлению различных характеристик этого процесса и построению стратегий текстопорождения, в том числе в связи с различными характеристиками финального текста. Особенно актуальными являются исследования связи процессуальных характеристик текста и субъективных оценок его качества. Подобные исследования имеют не только теоретическую, но и выраженную прикладную значимость, поскольку их результаты используются для усовершенствования методик обучения связной письменной речи. Однако до настоящего времени не исследованы в полной мере, в том числе с использованием методов многомерного анализа, сложные взаимосвязи между процессуальными характеристиками текста и текста как продукта. Кроме того, большинство исследований выполнены на материале английского языка. В статье представлены результаты первого пилотного исследования на материале текстов на русском языке, направленного на выявление связи между особенностями процесса порождения, рассчитанными на основе логов, полученных с использованием академического кейлоггера, и экспертными оценками качества текстов, а также числом языковых ошибок разных типов. Был выполнен корреляционный отбор процессуальных признаков, после чего был проведен анализ методом главных компонент (PCA), направленный на установление взаимосвязи между процессуальными признаками и оценкой качества текста, при этом показатели числа ошибок и общая оценка качества использовались как дополнительные переменные для интерпретации. Было выявлено, что процесс письма характеризуется устойчивой многомерной структурой, отражающей индивидуальные различия в ритме письма и стратегиях редактирования текста. Общая экспертная оценка качества умеренно коррелирует с измерением, связанным с ритмом и непрерывностью письма, однако не определяется им линейно. Полученные результаты подтверждают эвристическую ценность процессно-ориентированного анализа для исследования письменной речи и разработки эффективных методик обучения созданию эффективных текстов.

Ключевые слова: порождение текста, письменная речь, академический кейлоггер, процессуальные характеристики текста, качество текста, восприятие текста, математические методы в лингвистике, анализ главных компонент, стратегии порождения текста, паузация, редактирование текста.

Для цитирования: *Литвинова Т.А., Котлярова Е.С.* Восприятие письменного текста и процесс его порождения: анализ на основе данных кейлоггера // Известия Воронежского государственного педагогического университета. 2025. № 4. С. 225–232. DOI: 10.47438/2309-7078_2025_4_225

Введение

Оценка качества письменного текста традиционно производится на основе характеристик порожденного текста и складывается из разных показателей: языковая грамотность (орфография, пунктуация, грамматика), лексическое разнообразие, связность, структура, содержание (см. подробнее об оценке качества письменных текстов: (Sehlström et al., 2023 Self-efficacy; Vandermeulen getting a grip)). В педагогической практике и редакторской экспертизе именно итоговый текст выступает главным объектом контроля и анализа. Обучение созданию качественных, эффективных текстов является

© Литвинова Т.А., Котлярова Е.С., 2025

одной из ключевых задач, стоящих перед системой образования. Однако анализ текста как продукта в аспекте оценки его качества фиксирует лишь конечный результат и не объясняет, как именно он был получен: какие трудности возникали в ходе его порождения, насколько интенсивным было редактирование, где происходили «сбои» потока письма. В связи с этим в международных исследованиях письма активно используются методы, позволяющие изучать порождение текста как процесс. Одним из наиболее активно используемых инструментов являются академические кейлоггеры – программы, позволяющие фиксировать нажатия клавиш и все действия пользователя в процессе порождения текста с точной временной разметкой и возможностью восстановить стратегии создания связных письменных текстов (Литвинова, Дехнич, 2024).

Основным преимуществом данного метода является то, что он позволяет наблюдать за процессом создания текста, не вмешиваясь в него, анализировать «тонкие» процессуальные показатели (скорость порождения текста, распределение длительность пауз и их расположение, события редактирования, возвраты курсора и т.д.). Отметим, однако, что сам состав признаков, создаваемых на их основе, варьируется от работы к работе.

Ряд исследований в данной области посвящен связи признаков процесса порождения текста и качественным характеристикам порожденных текстов. В подобных работах исследуются отличия в процессах порождения текстов авторов с разным опытом создания текстов либо же предпринимаются попытки предсказать оценки качества текста на основе процессуальных признаков его создания (Baaijen; Vandermeulen Getting a grip on).

Особое внимание в подобных исследованиях уделяется анализу распределения и длительности пауз, а также паттернам редактирования текста, что в совокупности составляет ритм письма. Паузы в цифровом письме рассматриваются как индикатор распределения когнитивных усилий между формулированием, планированием и редактированием, а также как маркер особых зон текста (например, границы предложений, старт смысловых отрезков).

Исследуются также причины формирования тех или иных стратегий текстопорождения. К ним относят прежде всего степень развитости языковых навыков, а также состояние когнитивных функций пишущего (Haake, L. (2025). RELATIONSHIP BETWEEN EXECUTIVE FUNCTIONS AND L2WRITING). В указанной работе отмечается, что высоко развитые языковые навыки дают возможность пишущему планировать длинные лингвистические последовательности, что отражается в меньшем числе прерываний потока письма. Авторы со слабо развитыми языковыми навыками, напротив, делают больше коротких пауз (200-300 мс) и пауз между словами, что указывает на упор на планировании малых лингвистических единиц, тогда как авторы с высокоразвитыми языковыми навыками делают больше длинных пауз, связанных с планированием общей структуры текста. Низкая развитость языковых навыков, а также когнитивных ресурсов не позволяет координировать и реализовывать эти процессы, связанные с порождением текста, что влияет и на качество текста в итоге.

Различаются и стратегии редактирования у этих групп: авторы с менее развитыми языковыми навыками, как правило, редактируют отдельные языковые аспекты текстов, тогда как авторы с развитыми языковыми навыками редактируют структуру текста. В анализируемых работах указывается, что редактирование структуры повышает качество текста, тогда как частые и частные ревизии могут быть менее эффективными.

Обзор работ, связанных с исследованием взаимосвязи между процессуальными характеристиками и качеством финального текста, показывает, что те авторы, которые порождают текст с малым число пауз, как правило, порождают тексты высокого качества, но длинные паузы и ревизии могут улучшить качество текста. Отметим, однако, что большинство подобных исследований выполнены на материале текстов, созданных на неродном для авторов языке (как правило, английском), однако совсем необязательно наблюдения, сделанные на подобно материале, применимы к текстам на родном языке. Тем не менее проведенный нами анализ работ показал ряд сходств в аспекте связи между процессуальными характеристиками текста и его качеством независимо от того, на родном или неродном для автора языке они были написаны. Во-первых, время, потраченное на подготовку и непосредственно продуцирование текста, положительно связано с его качеством (Bennett et al., 2020, Guo et al., 2018 and Van Weijen et al., 2008). Более эффективные авторы продуцируют больше слов и минимальных единиц порождения текста (в англоязычной литературе – bursts), меньше пауз и с меньшей их общей продолжительностью (Kim et al., 2021). Авторы с более высокими оценками качества текстов продуцируют более длинные "bursts", и паузы в их текстах часто связаны с планированием (например, более длинные паузы встречаются на границах предложений), тогда как авторы с менее качественными текстами делают больше пауз между словами либо внутри слов, что связано с трудностями лексического доступа либо с когнитивной нагрузкой (Medimorec and Risko, 2017, Barkaoui, 2019). Что касается редактирования, то исследователи сходятся в том, что более опытные авторы делают меньше ревизий, их правки связаны с содержанием, тогда как менее успешные авторы делают больше локальных коррекций, связанных с исправлением ошибок в словах (Xu, 2018).

Как показывает этот краткий анализ, большинство исследователей сходятся во мнении о том, что процессуальные характеристики порождения и качество итогового текста связаны между собой, хотя и характер этой связи нуждается в дальнейших исследованиях. Данная научная область, как показывает библиометрический обзор (Литвинова, Дехнич 2024), находится на пике исследовательского интереса; предпринимаются даже попытки включить характеристики порождения текста в системы автоматической оценки качества готового текста (Xinyun He).

Несмотря на международный исследовательский интерес, для русского языка данная тематика является малоразработанной; только в последние годы стали создаваться соответствующие датасеты, а таже исследуется вопрос о границах коротких и длинных (когнитивных) пауз, отделяющих минимальные единицы текстопорождения (Литвинова, Молчанова, 2024). Насколько нам известно, работы, в которых бы анализировалась связь между экспертной оценкой качества текстов на русском языке и характеристиками их порождения, а также их соответствию языковым нормам, отсутствуют. Между тем именно для русского письма, где нормы орфографии и пунктуации тесно связаны с синтаксической организацией текста, сопоставление процесса порождения с типами ошибок и воспринимаемым субъективным качеством текста может дать ценную эмпирическую базу – как для лингвистики письменного текста, так и для решения прикладных задач (диагностика трудностей порождения связного письменного текста, редакторская работа, педагогические

рекомендации по улучшению навыков создания эффективных письменных текстов).

Настоящая работа встраивается в международную исследовательскую парадигму, но решает её в формате case study, ориентированном на демонстрацию похода к решению задачи поиска связей между процессуальными характеристиками порождения текста и субъективным качеством текстов на русском языке. Мы исследуем девять текстов трех авторов, а также файлы логов и впервые для русского языка сопоставляем процессуальные показатели с экспертной разметкой ошибок (пунктуация, орфография, грамматика, речевые нарушения, описки, логико-структурные нарушения) и с интегральной экспертной оценкой качества текста.

Цель исследования – описать и эмпирически продемонстрировать, какие процессуальные характеристики порождения текста (прежде всего паузальные показатели, скорость набора и признаки редактирования/навигации), извлекаемые из логов, соотносятся: а) с интегральной экспертной оценкой качества текста и б) с распределением типов ошибок. В силу малого объёма материала результаты интерпретируются как тенденции и как методологическая иллюстрация подхода, задающая основу для последующего расширения корпуса. В то же время наши результаты говорят о том, что даже на малых данных процессуальные признаки позволяют перейти от простого утверждения «в тексте больше/меньше ошибок» к более содержательному описанию того, каким способом авторы создают текстовые продукты разного качества.

В исследовании использованы тексты, извлеченные случайным образом из созданного в Воронежском государственном педагогическом университете под руководством Т.А. Литвиновой корпуса текстов разных жанров, содержащих файлы логов, отражающих все события набора текстов, а также метаданные об авторах. Для данного исследования нами были проанализированы 9 текстов, полученных от трёх авторов (далее – автор 16, автор 38, автор 51 – по уникальным номерам авторов). Каждый автор создавал тексты трех жанров – текст-рассуждение на одну из предложенных тем (например, напишите эссе-рассуждение о проблеме, которая Вас волнует), текст-описание вида из окна, текст-повествование о ярком событии, прочитанной книге, просмотренном фильме и т.д. что даёт 3Ч3=9 текстов. Нами намеренно были выбраны несколько текстов одного автора, созданные в разных жанрах, так как анализ научной литературы свидетельствует о важности учета жанра текста при анализе его процессуальных характеристик.

Для каждого текста доступны:

1. Лог-файл набора текста (данные академического кейлоггера), который содержит последовательность событий, фиксирующую действия автора при наборе (нажатия клавиш, действия мышью), время событий и автоматически рассчитанные параметры пауз и положения курсора.

2. Экспертная разметка ошибок итогового текста по категориям (то есть число ошибок): пунктуация, орфография, грамматика, речевые ошибки, описки, логика/структура, а также интегральная общая оценка текста, отражающая целостное экспертное впечатление от текста при чтении (далее – quality). *Общая оценка текста* производилась по следующим критериям: 0-3 – идея текста непонятна/несвязные предложения; 4-6 – идея понятна, но она не раскрыта до конца/повествование не заинтересовало; 7 – текст понятен, соответствует норме, но повествование не привлекло особого внимания; 8-9 – проблема и тема понятны, они раскрыты, текст представляет собой цельное произведение; 10 – текст представляет собой цельное произведение, повествование заставило задуматься, привлекло внимание, хотелось бы перечитать его снова.

Для экспертной оценки текстов было привлечено два разметчика – авторы этой статьи: профессиональный редактор и доктор филологических наук с большим опытом редакторской работы. На первом этапе тексты были оценены разметчиками независимо друг от друга; была получены высокая сходимость результатов оценки; спорные случаи были обсуждены, и в итоге был достигнут консенсус по ним. Итоговая разметка была использована в дальнейшем анализе.

При формировании списка процессуальных признаков мы учитывали накопленный опыт исследований в данной области, а также возможность автоматического извлечения значений признаков из логов программными средствами (весь анализ производился нами в среде R). Изначально мы извлекли 18 признаков, отражающих скорость набора, длительность и тип пауз, признаки редактирования и навигации. Далее для обработки данных и выявления на основе этих признаков комплексных метрик, отражающих основные характеристики процесса письма, и различий между авторами/текстами, мы использовали метод главных компонент (PCA). Известно, что при анализе небольших массивов данных важно избегать ситуации, когда в модель одновременно попадают очень близкие по смыслу, высокоррелирующие признаки. Это приводит к тому, что многомерная модель (в том числе PCA) начинает отражать не разные стороны процесса, а один и тот же эффект в нескольких измерениях и оси (компоненты) становятся менее интерпретируемыми. Поэтому перед выполнением анализа методом PCA мы удалили избыточно коррелирующие признаки. После проверки нулевой дисперсии и удаления избыточных коррелирующих параметров в качестве активных (то есть влияющих на построение осей PCA) использован следующий набор (таблица 1).

Таблица 1. Процессуальные признаки текстопорождения

Группа признаков	**Название признака**	**Описание признака**
Скорость набора (интенсивность ввода текста)	gross_cpm	скорость набора как количество клавиатурных событий в минуту (CPM, characters/keystrokes per minute; фактически скорость действий)
Паузы (“ритм” письма и распределение остановок в ходе набора)	pause_median_ms	медианная длительность паузы (более устойчивый показатель, чем среднее, менее чувствителен к редким сверхдлинным паузам)
	pause_p90_ms	90-й процентиль длительности пауз (отражает “хвост” распределения, характерную длительность наиболее длинных пауз в процессе письма и

Группа признаков	Название признака	Описание признака
		позволяет оценить степень фрагментации письма без опоры на единичные экстремальные остановки)
	pause_sentence_begin_sum	Суммарное значение пауз в позициях начала предложения
Признаки редактирования и навигации (признаки “возвратов” и технических действий во время набора)	caret_back_moves	количество случаев, когда позиция курсора (caret position) двигалась назад относительно предыдущего шага (возвраты к уже набранному фрагменту)
	caret_jump_abs	число “скачков” курсора (когда модуль изменения позиции курсора превышает 1) – индикатор нетипичных перемещений, не соответствующих обычному последовательному набору
	shift_ctrl_alt_n	число нажатий служебных клавиш (Shift/Ctrl/Alt).
	arrow_n	число нажатий стрелок (LEFT/RIGHT/UP/DOWN), если они фиксируются как отдельные события (навигация по тексту с клавиатуры)
	delete_n	число нажатий Delete

В анализе присутствуют два класса переменных: 1) активные переменные (см. таблицу выше) – они формируют “геометрию” пространства PCA; 2) дополнительные количественные переменные (quanti.sup) – число ошибок по категориям и общая оценка качества текста; эти признаки не участвуют в построении главных осей, но “проецируются” на пространство PCA, чтобы понять, как они соотносятся с процессуальными характеристиками (punct, ortho, gram, speech, typos, logic, quality); 3) дополнительная качественная переменная (quali.sup) – author_id, идентификатор автора (16 / 38 / 51); используется только для последующей интерпретации: раскраска точек, проверка группировки текстов по авторам для оценки устойчивости авторских профилей создания текста при смене жанра.

Отметим, что роль PCA в нашем анализе – “сжать” набор взаимосвязанных количественных признаков в 2–3 обобщённых измерения, чтобы увидеть, есть ли естественная группировка текстов по процессу набора; понять, какие признаки делают тексты “похожими” или “разными”; проверить, как на этой карте расположены: авторы (как категории), качество текста (как числовая оценка), ошибки (как числовые показатели). PCA показывает структуру взаимосвязей и профили различий между наблюдениями (логами).

Все активные признаки были стандартизированы (приведены к сопоставимому масштабу). После получения координат наблюдений на первых двух осях (Dim1, Dim2) вычислялась корреляция между координатой текста на оси (например, Dim1) и его общей оценкой качества (quality).

Результаты

Анализ показал, что тексты, созданные разными авторами, существенно различаются как по экспертной оценке качества, так и по структуре ошибок. При этом вариативность наблюдается не только между авторами, но и между разными текстами одного и того же автора, что указывает на наличие как индивидуальных стратегий письма, так и ситуативных различий, связанных с типом задания (рис. 1).

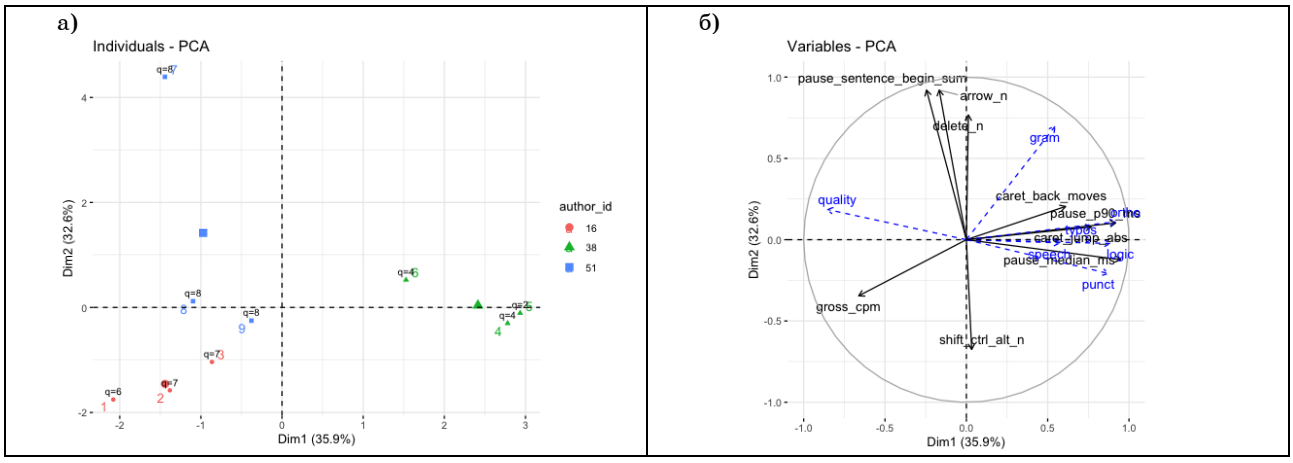


Рис. 1. Результаты анализы с использованием PCA:
а) расположение текстов в многомерном пространстве (q = значения оценок качества текстов),
б) процессуальные признаки в пространстве PCA

Общая экспертная оценка качества варьирует в диапазоне от низких значений (тексты автора 38) до стабильно высоких (тексты автора 51), тогда как тексты автора 16 занимают промежуточное положение. Эта трёхчленная градация («низкое – среднее – высокое качество») воспроизводится и при анализе структуры ошибок, и при анализе процессуальных характеристик набора.

Первая главная компонента (Dim1) аккумулирует вариативность, связанную с темпом и непрерывностью набора. В направлении одного полюса этой оси располагаются тексты, характеризующиеся более высокой скоростью ввода и меньшей длительностью пауз. В противоположном направлении концентрируются тексты с более выраженной фрагментацией процесса набора: более длинными и частыми паузами, а также большим числом возвратов курсора.

Вторая главная компонента (Dim2) отражает различия в стратегиях редактирования текста. Эта ось позволяет различить тексты, в которых процесс письма носит преимущественно линейный характер, и тексты, где наблюдается активная переработка уже набранного материала.

Как показывает рисунок, в пространстве главных компонент тексты разных авторов образуют устойчивые кластеры. Тексты автора 38, характеризующиеся низкой общей оценкой качества, располагаются в области пространства, ассоциированной с высокой фрагментацией процесса письма, большим количеством пауз и активными возвратами к ранее набранным фрагментам. Этот профиль сочетается с высоким числом языковых ошибок.

Тексты автора 51, напротив, группируются в области пространства, связанной с более структурированной работой с текстом: относительно умеренным темпом набора, паузами в позициях начала предложения и активной, но целенаправленной переработкой текста. Эти тексты получают стабильно высокую экспертную оценку качества.

Тексты автора 16 занимают промежуточное положение как по координатам в пространстве PCA, так и по качественным характеристикам. Их процесс письма характеризуется более высоким темпом по сравнению с автором 38, но меньшей выраженностью стратегической переработки по сравнению с автором 51.

Таким образом, PCA позволяет выделить индивидуальные профили письма, которые воспроизводятся на уровне нескольких текстов одного автора. Отметим, что тексты наиболее "эффективного" автора (51) образуют более разреженный кластер (по Dim 2), чем тексты двух других авторов, что говорит о разных применяемых им стратегиях редактирования для текстов разных жанров.

Показатели ошибок и общая оценка качества были проецированы на пространство главных компонент как дополнительные количественные переменные. Это позволило оценить, как характеристики результата письма соотносятся с выявленной структурой процессуальных различий.

Проекция общей оценки качества показывает, что она направлена в сторону области пространства, связанной с более высокой непрерывностью набора и меньшей фрагментацией процесса. При этом направление вектора качества не совпадает полностью ни с одной из главных компонент, что указывает на отсутствие простой линейной зависимости между качеством текста и отдельным измерением процесса письма.

Типы ошибок располагаются в пространстве неоднородно. Пунктуационные и орфографические ошибки проецируются в направлении, противоположном вектору качества, тогда как грамматические ошибки оказываются ближе к области, связанной с навигацией и переработкой текста. Это указывает на то, что различные типы ошибок связаны с разными аспектами процесса письма.

Для дополнительной интерпретации осей PCA была рассчитана корреляция между координатами текстов на главных компонентах и общей экспертной оценкой качества. Первая главная компонента демонстрирует умеренную отрицательную связь с качеством текста ($\rho = 0.6$) на уровне тенденции ($p = 0.091$). Вторая компонента не показывает связи с качеством.

Полученный результат свидетельствует о том, что качество текста частично соотносится с доминирующим измерением процессуальных различий, однако не может быть сведено к нему. Иными словами, характеристики процесса письма создают условия для формирования текста определённого качества, но не определяют его однозначно.

Обсуждение

Полученные результаты подтверждают, что процесс письма не может быть описан одной шкалой «быстро – медленно» или «хорошо – плохо». Даже при анализе небольшого корпуса текстов обнаруживается устойчивая многомерная структура процессуальных различий, отражающая разные стратегии письма. Анализ главных компонент позволил выявить эти различия и связать их с экспертным восприятием качества текста.

Важно подчеркнуть, что оси PCA формировались исключительно на основе характеристик процесса набора, тогда как показатели качества и ошибок использовались лишь для интерпретации. Это принципиально отделяет анализ процесса письма от анализа его результата.

Первая главная компонента (Dim1) в наибольшей степени определяется показателями, связанными с временной организацией письма и нарушением линейности набора. Наибольший вклад в формирование этой оси вносят медианная длительность пауз и 90-й процентиль длительности пауз, а также показатели скачков и возвратов курсора и скорость набора. Такое сочетание признаков позволяет интерпретировать первую компоненту как измерение, отражающее ритм письма и степень его непрерывности. В одном полюсе этой оси располагаются тексты, создаваемые в быстром темпе, с относительно короткими паузами и меньшим числом резких возвратов курсора. В противоположном полюсе находятся тексты, для которых характерны более длительные и вариативные паузы, частые скачки курсора и выраженная фрагментация процесса набора.

Важно отметить, что скорость набора не доминирует в этой компоненте (занимает 4-е место по вкладу в компоненту), а выступает лишь одним из элементов более общего паттерна. Это указывает на то, что не сама по себе скорость, а её сочетание с паузами и возвратами определяет место текста в пространстве первой компоненты.

Обнаруженная умеренная отрицательная связь между координатами текстов по первой компоненте и общей экспертной оценкой качества указывает на то, что качество текста действительно соотносится с ритмической организацией письма, но не сводится к ней. Тексты, характеризующиеся более выраженной фрагментацией процесса (длинные паузы, более низкая скорость набора, частые возвраты курсора), в среднем получают более низкие оценки качества. При этом речь идёт не о прямой причинно-следственной связи, а о тенденции, проявляющейся на уровне индивидуальных профилей письма.

Вторая главная компонента (Dim2) формируется за счёт совершенно иного набора признаков. Наибольший вклад в неё вносят показатели пауз в начале предложения, числа навигационных действий, операций удаления и использования служебных клавиш.

Эта ось может быть интерпретирована как измерение, отражающее стратегию планирования и переработки текста во время набора. В одном направлении располагаются тексты, для которых характерны остановки в позициях начала предложений, активное перемещение по тексту и его редактирование. В противоположном направлении находятся тексты, создаваемые более линейно, с меньшим числом возвратов, но большим числом использования служебных клавиш.

В отличие от первой компоненты, вторая ось не связана с общей оценкой качества. Это позволяет предположить, что активная переработка текста и навигация по нему не являются однозначными маркерами хорошего или плохого письма, а отражают индивидуальные стили и привычки авторов. Однако отметим, что по данной компоненте противопоставляются тексты "среднего" и "хорошего" авторов.

Сопоставление координат текстов с идентификаторами авторов показывает, что каждый автор демонстрирует относительно устойчивый профиль в пространстве главных компонент. Это особенно заметно для авторов с низкими и средними оценками качества.

Полученные результаты демонстрируют, что анализ логов набора текста может служить продуктивным инструментом для изучения письменной деятельности, особенно в сочетании с экспертной оценкой качества текста. Использование PCA позволило перейти от анализа отдельных показателей к рассмотрению целостных профилей письма, что особенно важно в гуманитарных исследованиях.

В то же время формат "кейс-стади" и малый объём данных накладывают ограничения на универсальность выводов. Перспективным направлением дальнейших исследований является расширение корпуса текстов, включение большего числа авторов и типов текстов, а также сопоставление процессуальных характеристик письма с более тонкими лингвистическими параметрами текста.

Выводы

В рамках данного исследования была предпринята попытка комплексного анализа процесса письма на основе логов набора текста в сочетании с экспертной оценкой качества и структуры ошибок. Полученные результаты позволяют сформулировать ряд выводов методологического и содержательного характера.

Во-первых, показано, что процесс письма обладает выраженной многомерной структурой, которая не сводится к отдельным показателям скорости или количеству ошибок. Даже на небольшом материале выявляются устойчивые различия между текстами, отражающие разные стратегии письма и способы организации письменной деятельности.

Во-вторых, анализ главных компонент, выполненный исключительно на основе процессуальных признаков набора текста, позволяет выделить латентные измерения, связанные, с одной стороны, с ритмом и непрерывностью письма, а с другой – со стратегиями планирования и переработки текста. Эти измерения отражают качественно различные аспекты письменной деятельности и могут рассматриваться как характеристики индивидуального стиля письма.

В-третьих, проекция показателей ошибок и общей экспертной оценки качества на пространство процессуальных признаков показывает, что качество текста систематически соотносится с особенностями процесса письма, однако не определяется ими напрямую. Обнаруженная умеренная связь между качеством текста и первой главной компонентой указывает на тенденцию: тексты, создаваемые в более устойчивом и менее фрагментированном режиме, в среднем воспринимаются экспертами как более качественные. Вместе с тем отсутствие жёсткой линейной зависимости подчёркивает сложный и многофакторный характер письменного качества.

В-четвёртых, выявленные различия между авторами подтверждают наличие устойчивых индивидуальных профилей письма, воспроизводимых на уровне нескольких текстов одного автора. Эти профили проявляются как в характеристиках процесса набора, так и в структуре ошибок и итоговой оценке качества, что позволяет рассматривать процессуальные признаки как потенциальные индикаторы индивидуального письменно-речевого стиля.

В-пятых, результаты показывают, что различные типы ошибок по-разному соотносятся с процессуальными характеристиками письма. Это позволяет рассматривать ошибки не как простые следствия отдельных моторных или временных параметров, а как результат взаимодействия когнитивных, языковых и стратегических компонентов письменной деятельности.

В целом исследование демонстрирует перспективность использования логов набора текста для гуманитарного анализа письменной речи. Предложенный подход позволяет связать процесс и результат письма, сохраняя при этом методологическое различие между ними. Несмотря на ограниченный объём данных и формат кейс-стади, полученные результаты могут служить основой для дальнейших исследований, направленных на расширение корпуса, уточнение набора процессуальных признаков и углублённое изучение индивидуальных стратегий письма в разных жанрах и типах заданий.

Конфликт интересов

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Библиографический список

1. Baaijen V.M., Galbraith D. Discovery Through Writing: Relationships with Writing Processes and Text Quality // Cognition and Instruction. 2018. № 36 (3). P. 199–223. DOI: 10.1080/07370008.2018.1456431

2. Vandermeulen N., Lindgren E., Waldmann C., Levlin M. Getting a grip on the writing process: (Effective) approaches to write argumentative and narrative texts in L1 and L2 // Journal of Second Language Writing. 2024. Vol. 65. P. 101113. DOI: 10.1016/j.jslw.2024.101113.

3. Xinyun He, Qi Shu, Mo Zhang, Wei Huang, Han Zhao, and Mengxiao Zhu. Beyond Final Products: Multi-Dimensional Essay Scoring Using Keystroke Logs and Deep Learning // Proceedings of the 15th International Learning Analytics and Knowledge Conference (LAK '25). Association for Computing Machinery. New York, 2025. P. 601–610. DOI: 10.1145/3706468.3706548

4. Haake, L. Does the relationship between executive functions and L2 writing depend on language proficiency? // Journal of Writing Research. 2025. Vol. 17(1). P. 115-153. DOI: 10.17239/jowr-2025.17.01.05

5. Sehlstrцm P., Waldmann C., Levlin M. Self-efficacy for writing and written text quality of upper secondary students with and without reading difficulties // Front Psychol. 2023. Sep 22;14:1231817. DOI: 10.3389/fpsyg.2023.1231817.

6. Dekhnich O.V., Litvinova T.A. A keystroke analysis to study writing: a bibliometric review using R and VOSviewer // Research Result. Theoretical and Applied Linguistics. 2024. Vol. 10, no. 3. P. 91-115. DOI: 10.18413/2313-8912-2024-10-3-0-5. EDN RCQMSX.

7. Litvinova T.A., Molchanova V.A. Investigating between-word pause duration in Russian typed texts using mixture modeling based on keystroke data // Research Result. Theoretical and Applied Linguistics. 2024. Vol. 10, no. 4. P. 147-166. DOI: 10.18413/2313-8912-2024-10-4-0-8. EDN SAFDNA.

8. Литвинова Т.А., Молчанова В.А., Любова С.Г. Минимальные структурные единицы порождения текста на русском языке: анализ на основе данных кейлоггера и семантической модели // Известия Воронежского государственного педагогического университета. 2025. № 1(306). С. 195-203. DOI 10.47438/2309-7078_2025_1_195. EDN HDJOHH.

9. Bennett R.E., Zhang M., Deane P., Rijn P.W. van. How do proficient and less proficient students differ in their composition processes? // Educational Assessment.2020. No. 25 (3). P. 198–217.

10. Guo H., Deane P.D., Rijn P.W. van, Zhang M., Bennett R.E. Modeling basic writing processes from keystroke logs // Journal of Educational Measurement. 2018. No. 55 (2). P. 194–216.

11. Kim M., Tian Y., Crossley S.A. Exploring the relationships among cognitive and linguistic resources, writing processes, and written products in second language writing // Journal of Second Language Writing. 2021. 53. DOI: 10.1016/j.jslw.2021.100824.

12. Weijen D. van, Bergh H. van den, Rijlaarsdam G., Sanders T. Differences in process and process-product relations in L2 Writing ITL // International Journal of Applied Linguistics. 2008. No. 156. P. 203–226.

13. Medimorec S., Risko E.F. Pauses in written composition: On the importance of where writers pause // Reading and Writing. 2017. No. 30 (6). P. 1276-1285. DOI: 10.1007/s11145-017-9723-7

14. Barkaoui K. What can L2 writers' pausing behavior tell us about their L2 writing processes? // Studies in Second Language Acquisition. 2019. No. 41 (3). DOI: 10.1017/S027226311900010X

15. C. Xu Understanding online revisions in L2 writing: A computer keystroke-log perspective // System. 2018. No. 78. P. 104–114.

References

1. Baaijen, V. M., & Galbraith, D. (2018). Discovery Through Writing: Relationships with Writing Processes and Text Quality. *Cognition and Instruction*. 36(3), 199–223. Available from: https://doi.org/10.1080/07370008.2018.1456431 (In English)

2. Vandermeulen, N., Lindgren, E., Waldmann C., 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), 101113. Available from: https://doi.org/10.1016/j.jslw.2024.101113. (In English)

3. Xinyun, He, Qi, Shu, Mo, Zhang, Wei, Huang, Han, Zhao, and Mengxiao, Zhu. (2025) Beyond Final Products: Multi-Dimensional Essay Scoring Using Keystroke Logs and Deep Learning. *In: Proceedings of the 15th International Learning Analytics and Knowledge Conference (LAK '25). Association for Computing Machinery*, 601–610. Available from: https://doi.org/10.1145/3706468.3706548 (In English)

4. Haake, L. (2025). Does the relationship between executive functions and L2 writing depend on language proficiency?. *Journal of Writing Research*. 17(1), 115–153. Available from: https://doi.org/10.17239/jowr-2025.17.01.05 (In English)

5. Sehlstrцm, P., Waldmann, C., Levlin, M. (2023) Self-efficacy for writing and written text quality of upper secondary students with and without reading difficulties. *Front Psychol*. Sep 22;14:1231817. Available from: doi:10.3389/fpsyg.2023.1231817. (In English)

6. Dekhnich, O.V., Litvinova, T.A. (2024) A keystroke analysis to study writing: a bibliometric review using R and VOSviewer. *Research Result. Theoretical and Applied Linguistics*. 10 (3), 91-115. Available from: doi: 10.18413/2313-8912-2024-10-3-0-5. EDN RCQMSX. (In English)

7. 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. Available from: doi: 10.18413/2313-8912-2024-10-4-0-8. EDN SAFDNA. (In English)

8. Litvinova, T.A., Molchanova, V.A., Lyubova, S.G. (2025) Minimal'nye strukturnye edinitsy porozhdeniya teksta na russkom yazyke: analiz na osnove dannykh keiloggera i semanticheskoi modeli [Minimal structural units of text generation in Russian: analysis based on keylogger data and semantic model]. *Izvestiya Voronezhskogo gosudarstvennogo pedagogicheskogo universiteta*. 1(306), 195-203. Available from: doi: 10.47438/2309-7078_2025_1_195. EDN HDJOHH (In Russian)

9. Bennett, R.E., Zhang, M., Deane, P., Rijn, P.W. van. (2020) How do proficient and less proficient students differ in their composition processes? *Educational Assessment*. 25 (3), 198–217. (In English)

10. Guo, H., Deane, P.D., Rijn, P.W. van, Zhang, M., Bennett R.E. (2018) Modeling basic writing processes from keystroke logs. *Journal of Educational Measurement*. 55 (2), pp. 194–216. (In English)

11. Kim, M., Tian, Y., Crossley, S.A. (2021) Exploring the relationships among cognitive and linguistic resources, writing processes, and written products in second language writing. Journal of Second Language Writing, (53). Available from: doi: 10.1016/j.jslw.2021.100824 (In English)

12. Weijen, D. van, Bergh, H. van den, Rijlaarsdam, G., Sanders, T. (2008) Differences in process and process-product relations in L2 Writing ITL. *International Journal of Applied Linguistics*. (156), 203–226. (In English)

13. Medimorec, S., Risko, E.F. (2017) Pauses in written composition: On the importance of where writers pause. *Reading and Writing*. 30 (6), 1276-1285. Available from: doi: 10.1007/s11145-017-9723-7 (In English)

14. Barkaoui, K. (2019) What can L2 writers' pausing behavior tell us about their L2 writing processes?. *Studies in Second Language Acquisition*. 41 (3). Available from: doi: 10.1017/S027226311900010X (In English)

15. Xu, C. (2018) Understanding online revisions in L2 writing: A computer keystroke-log perspective. System. (78), 104–114. (In English)

Поступила в редакцию 16.10.2025
Подписана в печать 30.12.2025

Original article
UDC 81-112.4
DOI: 10.47438/2309-7078_2025_4_225

PERCEPTION OF WRITTEN TEXT AND THE PROCESS OF ITS GENERATION: ANALYSIS BASED ON KEYLOGGER DATA

Tatiana A. Litvinova[1], Elena S. Kotlyarova[2]

Voronezh State Pedagogical University[1, 2]
Voronezh, Russia

[1]*Dr. Philol. Sci, Professor of the Department of Russian language, Modern Russian and Foreign Literature, ORCID ID: 0000-0002-6019-3700, tel.: +7-473-264-44-17, e-mail: centr_rus_yaz@mail.ru*
[2]*Postgraduate student of the Department of Russian Language, Modern Russian and Foreign Literature, ORCID ID: 0000-0002-4162-5002, e-mail: kray-voronezh@.ru*

Abstract. In modern research of written speech, more and more attention is being paid not only to the analysis of text as the final product of speech generation, but also to the study – using special programs – academic keyloggers that register keystrokes and all related temporal data – of the very process of its generation, identifying various characteristics of this process and building strategies for text generation, including in links to various characteristics of the final text. Especially relevant are studies of the relationship between the procedural characteristics of the text and subjective assessments of its quality. Such studies have not only theoretical, but also pronounced applied significance, since their results are used to improve teaching methods for coherent written speech. However, to date, the complex relationships between the procedural characteristics of the text and the text as a product have not been fully investigated, including using multidimensional analysis methods. In addition, most of the research is based on the English language. The article presents the results of the first pilot study based on texts in Russian, aimed at identifying the relationship between the features of the generation process, calculated on the basis of logs obtained using an academic keylogger, and expert assessments of the quality of texts, as well as the number of linguistic errors of various types. A correlation selection of procedural features was performed, followed by a principal component analysis (PCA) aimed at establishing the relationship between procedural features and text quality assessment, while error rates and overall quality assessment were used as additional variables for interpretation. It was revealed that the writing process is characterized by a stable multidimensional structure reflecting individual differences in writing rhythm and text editing strategies. The general expert assessment of quality is moderately correlated with the measurement related to the rhythm and continuity of writing, but it is not linearly determined by it. The results obtained confirm the heuristic value of process-oriented analysis for the study of written speech and the development of effective teaching methods for creating effective texts.

Key words: text generation, written speech, academic keylogger, procedural characteristics of the text, text quality, text perception, mathematical methods in linguistics, analysis of the main components, text generation strategies, pausing, text editing.

Cite as: Litvinova, T.A., Kotlyarova, E.S. (2025) Perception of written text and the process of its generation: analysis based on keylogger data. *Izvestia Voronezh State Pedagogical University.* (4), 225–232. (In Russ., abstract in Eng.). DOI: 10.47438/2309-7078_2025_4_225

Received 16.10.2025
Accepted 30.12.2025