

КЛАВИАТУРНО-ОПОСРЕДОВАННАЯ РЕЧЬ КАК ОДНА ИЗ ОСНОВНЫХ СФЕР РЕАЛИЗАЦИИ ИДИОЛЕКТА НОСИТЕЛЯ СОВРЕМЕННОГО РУССКОГО ЯЗЫКА: ПРОБЛЕМЫ ИССЛЕДОВАНИЯ ПРОЦЕССА ПОРОЖДЕНИЯ И ПЕРВЫЕ РЕЗУЛЬТАТЫ

Татьяна Александровна Литвинова¹, Ольга Владимировна Загоровская²

Воронежский государственный педагогический университет^{1, 2}
Воронеж, Россия

¹Кандидат филологических наук, заведующий научно-исследовательской лабораторией корпусной идиолектологии,
e-mail: centr_rus_yaz@mail.ru

²Доктор филологических наук, профессор кафедры русского языка, современной русской и зарубежной литературы,
e-mail: olzagor@yandex.ru

Аннотация. В настоящее время одной из основных форм реализации идиолекта (под которым мы понимаем продуктивную индивидуально-речевую деятельность) носителя русского языка является производство т.н. клавиатурных текстов, то есть текстов, созданных при помощи клавиатуры электронного устройства. Лингвисты исследуют различные характеристики подобных текстов как продуктов, между тем как процессы порождения таких текстов остаются неизученными. Современные программные и технические средства позволяют ученым в определенной степени визуализировать процесс порождения текстов и тем самым расширить научные представления о механизмах текстопорождения. Подобные исследования активно проводятся в мировой науке в рамках направления, носящего название *writing research*. Русский язык находится на периферии подобных исследований. В статье представлен краткий обзор состояния исследований процесса текстопорождения в современной мировой науке, выделены основные теоретико-методологические и собственно методологические проблемы в данной области и представлены результаты пилотных экспериментальных исследований процесса порождения текстов на русском языке. Показано, что решение одного из фундаментальных вопросов языкознания – проблемы порождения текста, в том числе проблемы определения минимальных единиц текстопорождения, невозможно без учета идиолектного фактора.

Ключевые слова: идиолект, корпусная лингвистика, идиолектология, порождение текста, психолингвистика, клавиатурный текст.

Для цитирования: Литвинова Т. А., Загоровская О. В. Клавиатурно-опосредованная речь как одна из основных сфер реализации идиолекта носителя современного русского языка: проблемы исследования процесса порождения и первые результаты // Известия Воронежского государственного педагогического университета. 2021. № 3. С. 127–133. DOI 10.47438/2309-7078_2021_3_127

Введение

В настоящее время двучленная парадигма онтологических форм существования русского языка (устная речь – письменная речь) стремительно расширяется за счет появления нового важного основания классификации, которым является способ продуцирования текста (традиционный/клавиатурно-опосредованный) [3]. Выбор названного способа является одним из важнейших выборов, осуществляемых субъектом идиолекта как продуктивной индивидуально-речевой деятельности. В отличие от традиционного письма от руки, электронное устройство предлагает субъекту идиолекта различные опции, существенно влияющие на характеристики его речевой продукции. Принципи-

альная возможность подобного взаимодействия с электронным устройством является основанием для выделения клавиатурно-речевого поведения как особого объекта исследований (см. подробнее: [2]).

Безусловно, продукты речевой деятельности, порожденные с использованием клавиатуры электронного устройства, активно исследуются отечественными учеными в различных аспектах (см. об этом: [3]), однако проблема порождения подобных текстов, и в частности выделения минимальных единиц текстопорождения, остается в российской лингвистике малоисследованной, между тем как современные технические и программные средства позволяют в определенной степени визуализировать процесс создания текста носителем языка.

Как известно, проблема порождения текста является из одной из ключевых проблем психолинг-

вистики. В настоящее время в психолингвистике существует «несколько десятков» [1] моделей речепорождения, что свидетельствует о том, что названная весьма проблема далека от своего разрешения. Вне всяких сомнений, отечественные психолингвисты внесли значительный вклад в решение указанной проблемы, однако работы, в которых процесс порождения текста на русском языке рассматривался бы с использованием экспериментальных методов (в первую очередь – с использованием программ для анализа клавиатурного поведения – килоггеров), отсутствуют. Применительно же к другим мировым языкам указанные новые экспериментальные методы широко используются для извлечения процессуальных характеристик текстопорождения: локаций и длительности пауз, самокоррекций и т.д.

Как известно, проблема порождения текста имеет не только собственно научную, но и социально-культурную значимость. Новые научные данные об особенностях порождения текста пишущим (в том числе о затруднениях, которые он испытывает на различных стадиях процесса письма), влияющих на лингвистические характеристики текста, определяющие его качество, полученные с учетом жанра и способа продуцирования текста (рукописный/клавиатурный), а также индивидуальных особенностей пишущего, могут быть положены в основу разработки рекомендаций по улучшению навыков связной письменной речи и повышению качества создаваемых обучающимися текстов, что является одной из ключевых задач современного образования.

1. Современное состояние исследований в области изучения процесса текстопорождения с использованием экспериментальных методов

Исследования, посвященные изучению процесса порождения текста, ведутся психолингвистами на протяжении длительного времени. Процесс порождения текста, в отличие от его результата (то есть текста как продукта), имеет два измерения: пространственное и временное. Несмотря на очевидную взаимосвязь между процессом и продуктом, невозможно реконструировать первое на основе второго. Следовательно, необходимо искать способы исследования процесса текстопорождения, при этом принимая во внимание существование различных способов порождения текста (рукописный и клавиатурный).

Первоначально исследования процесса порождения текста основывались на анализе различных версий одного и того же текста, но позже в подобных работах стал использоваться метод видеозаписи процесса письма, и таким образом фокус исследования с пространственного измерения сместился на темпоральный. Еще в работе 1987 года *Writing in real time: modelling production processes* [23], в которой был выделен концептуальный уровень письма (семантика, грамматика и правописание) и уровень последовательных действий (планирование и формулирование), были заложены основы экспериментального подхода к исследованию процесса письма, при котором основное внимание уделяется его временным характеристикам.

В настоящее время благодаря появлению новых технических средств, позволяющих наблюдать процесс текстопорождения, подобные исследования приобрели новую доказательную базу. К указанным техническим средствам прежде всего относятся программы, фиксирующие длительность нажатия клавиш, а также тип действия пишущего (например, удаление, вставка и т.д.) (килоггеры, см. обзор в работе [15], а также [20]). Названные средства позволяют получать данные об общих характеристиках процесса текстопорождения – местоположении и длительности пауз, самокоррекций и т.д. Внимание исследователей к паузам объясняется тем, что они составляют в среднем 50 % времени продуцирования текста и представляют собой фундаментальные моменты концептуализации, формулирования или контроля над сообщением [8].

Результаты подобных исследований широко используются для моделирования процесса письма [12], изучения когнитивных процессов, связанных с ним [11] и т.д.

Широкое распространение названных выше методов привело к появлению призывов перейти от лингвистики письменного текста к «лингвистике письма» (*from written linguistic stowriting linguistics*), от лингвистики продукта к лингвистике процесса [19].

В современной мировой лингвистике исследования процессов текстопорождения имеют прикладную направленность, связанную с разработкой методов, помогающих улучшить навыки письменной речи путем анализа участков текстов, на которых пишущие испытывают затруднения. Так, авторы работы [24], проанализировав 29 признаков, описывающих процесс текстопорождения, извлеченных на основе данных килоггера, обнаружили, что подобные признаки могут быть разделены на 4 группы («Общие характеристики текстопорождения», «Редактирование на уровне слов и границ слов», «Редактирование на уровне фраз и синтагм», «Обдумывание»), и связали выделенные группы признаков с компонентами теоретической модели письма J. Hayes [13]. Было установлено, что студенты могут испытывать трудности с одним или несколькими названными процессами, связанными с письмом. В указанной работе был сделан вывод о том, что подобная информация, извлеченная из данных килоггера, позволит преподавателям давать учащимся более конкретные инструкции по улучшению навыков их связной письменной речи.

Несмотря на то, что экспериментальный анализ текстопорождения является активно развивающимся направлением в исследованиях письма (*writing research*), подобные работы на материале текстов на русском языке единичны. К ним, в частности, относится диссертационное исследование Е.М. Чухарева «Лингвостатистические корреляты спонтанности в компьютерно-опосредованном дискурсе: на материале русскоязычного чата» [7], посвященное анализу письменной диалогической речи на русском языке 36 респондентов – участников чата с использованием данных о длительности пауз между нажатиями клавиш. В названной работе исследователем было выделено два типа пауз хезитации: лексические,

связанные с орфографическими словами, и сегментные, связанные с речевыми отрезками большей протяженности. В работе, выполненной на том же эмпирическом материале [9], названный автор исследовал характеристики длительности пауз между нажатиями и пришел к выводу о том, что паузы между нажатиями клавиш свыше 500 мс могут рассматриваться как маркер хезитации, свыше 1,2 с – как маркер планирования.

Следует отметить, что большая часть работ, посвященных исследованию процесса письма с использованием экспериментальных методов, ограничивается анализом общих признаков текстопорождения – длительностью и местоположением пауз и самокоррекций, между тем как дополнение подобных признаков собственно лингвистической информации может дать новые, интересные результаты. Килоггеры предоставляют информацию на уровне знака, и хотя она может быть очень интересной, традиция исследования текста как продукта убеждает в необходимости применения методов анализа единиц более высоких уровней, нежели знак. Крайне важно связать когнитивные процессы, связанные с процессом письма, с лингвистическими характеристиками текста. В таком случае пауза выступает слишком обобщенным понятием, даже если разделить ее на уровни и разряды (пауза на уровне знака – слова – предложения). Для понимания механизмов текстопорождения принципиально важным является решение вопроса о его минимальных единицах и их лингвистической природе, однако, как справедливо отмечается французскими исследователями G. Cislaru и Th. Olive в единственном на сегодняшний день монографическом исследовании, специально посвященном выделению минимальных единиц текстопорождения (в англоязычной литературе – bursts) [10], указанный вопрос, являющийся «наиболее интересным для психолингвистов», остается практически не изученным. В названной работе проверялась гипотеза о том, что bursts представляют собой коллокации (Multiword expression), но она не подтвердилась.

Вопрос о лингвистической природе минимальных единиц текстопорождения ставился и в упомянутой выше работе Е.М. Чухарева 2014 года [9], при этом в названном исследовании использовался подход от гипотезы к данным: сначала текст размечался на слова и «пропозициональные выражения», затем с границами этих выражений сопоставлялись паузы. И хотя в целом паузы были связаны с выделенными единицами, ответить однозначно на вопрос о лингвистической природе минимальных единиц текстопорождения оказалось невозможным, что, на наш взгляд, связано с самой методологией исследования: в связи с разными моторными навыками и особенностями когнитивной сферы авторов следует определять порог пауз индивидуально для каждого автора и затем анализировать лингвистическую природу вычлеченных единиц.

В русской лингвистической традиции неоднократно высказывались предположения о том, что единицей кодирования в модели речепорождения должно являться не слово, а совокупность слов (фраза, синтагма): «Каждая речь, являющаяся

средством общения, является не столько комплексом лексических единиц (слов), сколько системой синтагм (целых высказываний)» [5, с. 37]. Подход к синтагме как к единице линейного прогнозирования и моторного программирования высказывания характерен и для работ А.А. Леонтьева. Вопрос о формальных характеристиках синтагмы, однако, является сложным, особенно применительно к письменной речи (в устной речи одним из основных критериев выступает фонетический).

В качестве минимальной единицы порождения устного текста («кванта дискурса») исследователями (см., например, [6]) выделяется элементарная дискурсивная единица (ЭДЕ). Предполагается, что дискурс порождается толчками и пульсациями, его течение не похоже на течение реки [6] (отметим, что англоязычный термин burst в переводе с английского означает «взрыв»).

Согласно выводам, изложенным в работе [6], типичные ЭДЕ отличаются координацией между различными аспектами процесса порождения дискурса: с физиологической точки зрения ЭДЕ произносятся на одном выдохе; с когнитивной точки зрения – вербализует один «фокус сознания»; с синтаксической представляет собой одну предикацию (клаузу). Просодически ЭДЕ организована как один производительный контур с точки зрения движения тона (частота), наличия основного акцентного центра (интенсивность), темпа (ускорение – замедление) и громкости (затухание к концу). Предлагается возможным предположить существование аналога ЭДЕ и в письменной речи.

Как видится, основной теоретико-методологической проблемой исследований в области выявления и описания минимальных единиц порождения письменного текста является отсутствие учета идиолектного фактора: необходимо исследовать процессы порождения текстов одними и теми же субъектами в разных условиях осуществления идиолектной деятельности (в том числе с использованием разных способов продуцирования текстов), с тем чтобы выявить индивидуальные различия и установить общие закономерности. Для решения данной задачи весьма перспективно использовать методы, специально предназначенные для выделения источников вариации в многомерных данных.

Следует отметить, что отечественными психолингвистами неоднократно отмечалась необходимость учета индивидуальных различий пишущих при моделировании процессов текстопорождения (работы Н. И. Жинкина, А. А. Леонтьева, И. А. Зимней и др.).

Упомянутое выше отсутствие исследований лингвистической природы минимальных единиц текстопорождения в идиолектном аспекте объясняется прежде всего обстоятельствами методологического характера, и прежде всего отсутствием соответствующих корпусов текстов. Еще одной методологической проблемой является отсутствие специальных программ, позволяющих трансформировать данные килоггеров в лингвистически значимую информацию, для большинства языков, за исключением английского и голландского [14]. Кроме того, в большинстве работ процессуальные и лингвисти-

ческие характеристики текстопорождения рассматриваются обособленно.

2. Результаты экспериментальных исследований процесса порождения клавиатурной речи

Нами были проведены эксперименты, направленные на выявление степени идиолектной обусловленности процессуально-лингвистических признаков текстопорождения. В качестве эмпирической базы исследования послужил созданный нами корпус текстов MULTIFACTOR, содержащий как традиционные, так и клавиатурно-опосредованные устные и письменные тексты, созданные одними и теми же авторами. Названный корпус входит в состав БД *RusIdiolect* – уникального ресурса, специально направленного на исследование идиолекта [17]. Помимо собственно текстов, корпус содержит видеозаписи процесса порождения текста (3 диалогических и 3 монологических текста каждого автора: диалог-беседа на свободную тему с разными собеседниками, монолог-жалоба; эссе; пересказ одного и того же продемонстрированного респондентам фильма). Для нивелирования различий в объеме для каждого текста (в данном пилотном исследовании анализировались тексты 3 случайно отобранных авторов) были выбраны для анализа первые 400 событий (под событием подразумевается токен – словоформа, знак препинания и т.д., различные виды редактирования и форматирования).

Видеозаписи текстов были размечены в программе ELAN [22], получена информация о длительности тех или иных событий текста. На основании полученных данных были сконструированы признаки, комбинирующие лингвистическую и темпоральную информацию этих двух уровней (например, общее время, затраченное на редактирование; средняя длительность набора строевых слов и т.д.). Анализировались признаки двух групп: 1) общие характеристики текстопорождения как процесса (например, *Pure_Words* = общее число слов / (продолжительность письма – продолжительность пауз *pause time* – продолжительность исправлений (*part+del+corr*); *PM_DEL* – отношение числа пунктуационных знаков к числу удаленных слов и т.д.; всего 20 признаков); 2) частоты биграмм, то есть последовательностей двух событий, одним из которых является пауза (164 признака).

В качестве метода анализа мы использовали один из методов многомерного анализа данных – метод главных компонент (PCA), ранее успешно применявшийся нами для решения различных задач, связанных с анализом идиолекта [4; 16], с визуализацией полученных результатов с использованием пакета *mixOmics* [21]. Для нивелирования эффекта типа текста мы использовали метод разделения источников вариации в данных (функция *within Variation* пакета *mixOmics*, см. подробнее об этом в работе [21]). Пакет *mixOmics* объединяет методы поиска структуры в данных и метод классификации *Partial Least Squares Discriminant Analysis (PLS-DA)* – популярный в области биохимических исследований классификатор, учитывающий мультиколлинеарность признаков и характеризующийся интерпретируемостью (признаки

напрямую связаны с откликом; можно выделить признаки, вносящие наибольший вклад в модель).

Нами тестировался структурно-разнородный сценарий, то есть в тестовую и тренировочную выборку входили как диалогические, так и монологические тексты. Целью классификации было отнесение текста к автору. Был применен метод нивелирования влияния типа текста.

Ошибка классификации PLS-DA составила 0.175 для первой группы признаков, 0.08 – для второй группы признаков (пермутационные тесты показывают, что разделение по классам не является случайным, $p < 0.05$).

Нами были проведены дополнительные исследования, которые показали, что стилеметрические признаки, традиционно используемые в идентификационных идиолектных исследованиях и не учитывающие информацию о темпоральных характеристиках процесса порождения текста (см., например, [18]), не позволяют получить статистически значимые результаты классификации в данном сценарии.

Результаты проведенного нами пилотного исследования показывают высокую степень идиолектоспецифичности характеристик текстопорождения, что убедительно подтверждает высказанный нами ранее тезис о необходимости учета идиолектного фактора при моделировании процессов создания текста.

Выводы

Таким образом, как показал проведенный нами анализ научной литературы, несмотря на то, что в настоящее время активно развиваются исследования, направленные на изучение процесса порождения письменных (клавиатурных) текстов с использованием современных экспериментальных методов, позволяющих визуализировать названный процесс, в данной области остается множеством нерешенных проблем, в частности, мало изучен вопрос о лингвистической природе минимальных структурных единиц текстопорождения. Прогрессу исследований в данной области, по нашему мнению, препятствует отсутствие учета идиолектоспецифичности особенностей процессов создания письменных текстов в целом и клавиатурно-опосредованных текстов в частности. Именно отсутствие учета названного фактора является, по нашему мнению, ключевой теоретико-методологической проблемой в рассматриваемой предметной области. К основным методологическим проблемам названной области относится прежде всего отсутствие специальных корпусов текстов, содержащих данные о процессах порождения текстов одними и теми же авторами в разных условиях осуществления идиолектной деятельности; использование признаков, несущих темпоральную и лингвистическую информацию о процессах текстопорождения, в отдельности; невнимание к возможностям, предоставляемым современными методами анализа многомерных данных, в том числе направленных на разделение источников вариации в данных.

Результаты проведенного нами пилотного экспериментального исследования на материале уникального корпуса текстов на русском языке, содержащего видеозаписи процессов их порождения, по-

казали высокую степень идиолектоспецифичности характеристик процессов текстопорождения, а также перспективность использования новейших методов многомерного анализа данных.

Дальнейшие работы предполагают расширение круга респондентов, а также арсенала методов исследования путем добавления современных методов, направленных на анализ временных рядов.

Источник финансирования

Статья подготовлена при поддержке гранта Российского научного фонда № 18-78-10081.

Конфликт интересов

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

Библиографический список

1. Глухов В. П. Основы психолингвистики: учеб. пособие для студентов педвузов. М. : АСТ : Астрель, 2005. 351 с.
2. Загоровская О. В., Литвинова Т. А. Клавиатурно-речевое поведение носителя русского языка как объект экспериментальной идиолектной психолингвистики : к становлению нового научного направления // Методология современного языкознания – 3 : сб. ст. в честь юбилея В.А. Пищальниковой. М. : Р. Валент, 2021. С. 72–81.
3. Загоровская О. В., Литвинова Т. А. Новые формы речи в медиапространстве современной России как сфера реализации идиолекта и идиостиля носителя русского языка // Известия Воронежского государственного педагогического университета. 2021. № 1 (290). С. 157–162.
4. Литвинова Т. А. Стилеметрическая идентификация идиолекта на основе принципов корпусной идиолектологии // Вопросы криминологии, криминалистики и судебной экспертизы. 2021. № 1 (49). С. 126–132.
5. Лурия А. Р. Основные проблемы нейролингвистики. М. : Изд-во МГУ, 1975. 256 с.
6. Рассказы о сновидениях. Корпусное исследование устного русского дискурса / под ред. А. А. Кибрика, В. И. Подлесской. М. : Языки славянских культур, 2009. 736 с.
7. Чухарев Е. М. Лингвостатистические корреляты спонтанности в компьютерно-опосредованном дискурсе : на материале русскоязычного чата : дис. ... канд. филол. наук. СПб., 2009. 211 с.
8. Chanquoy L., Foulin J.-N., Fayol M. Writing in adults: A real-time approach // Theories, models and methodology in writing research / G. Rijlaarsdam, H. van den Bergh, M. Couzijn (eds.). Amsterdam : Amsterdam University Press, 1996. P. 36–43.
9. Chukharev-Hudilainen E. Pauses in spontaneous written communication: A key-stroke logging study // Journal of Writing Research. 2014. Vol. 6, no. 1. P. 61–84.
10. Cislaru G., Olive T. Le Processus de textualisation. Analyse des unités linguistiques de performance. Louvain-la-Neuve : De Boeck, 2018. 272 p.
11. Galbraith D., Baaijen V. M. Aligning keystrokes with cognitive processes in writing // Observing writing: Insights from keystroke logging and handwriting / E. Lindgren and K. P.H. Sullivan (eds.). Leiden : Brill Academic Publishers, 2019. P. 306–325.
12. Modeling basic writing processes from keystroke logs / H. Guo, P. Deane, P. W. van Rijn [et al.] // Journal of Educational Measurement. 2018. Vol. 55. P. 194–216.
13. Hayes J. R. Modeling and remodeling writing. Written Communication. 2012. Vol. 29. P. 369–388.
14. Leijten M., Van Waes L. Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes // Written Communication. 2013. Vol. 30, no. 3. P. 358–392. DOI: 10.1177/0741088313491692
15. Lindgren E., Knospe Y., Sullivan K. P. H. Researching Writing with Observational Logging Tools from 2006 to the Present // Observing writing: insights from keystroke logging and handwriting / E. Lindgren and K. P. H. Sullivan (eds.). Leiden : Brill Academic Publishers, 2019. P. 1–29.
16. Litvinova T. Process-oriented characteristics of an idiolect for authorship attribution of heterogeneous texts: a pilot study // CEUR Workshop Proceedings. Сер. "CMLS 2020 – Proceedings of the Computational Models in Language and Speech Workshop, co-located with 16th International Conference on Computational and Cognitive Linguistics, TEL 2020", 2020. P. 3–16.
17. Litvinova T. RusIdiolect: A New Resource for Authorship Studies // Lecture Notes in Networks and Systems. 2021. Vol. 186. P. 14–23. DOI: 10.1007/978-3-030-66093-2_2
18. Overview of the Rusprofiling PAN at FIRE track on cross-genre gender identification in Russian / T. Litvinova, F. Rangel, P. Rosso [et al.] // CEUR Workshop Proceedings. Сер. "FIRE 2017 – Working Notes of FIRE 2017 – Forum for Information Retrieval Evaluation", 2017. P. 1–7.
19. Editorial genesis: From comparing texts (product) to interpreting rewritings (process) / Mahrer R., De Angelis R., Del Lungo A. [et al.] // Writing(s) at the Crossroads / G. Cislaru (ed.). Amsterdam : John Benjamins, 2015. P. 151–170.
20. Olive T. Methods, tools and techniques for the on-line study of the writing process // Writing: Processes, Tools and Techniques / N. L. Mertens (ed.). NY : Nova Publisher. P. 1–18.

21. K-A mixomics: An R package for 'omics feature selection and multiple data integration / F. Rohart, B. Gautier, A. Singh, K-A. Lk C. // *PLoSComput Biol*. 2017. Vol. 13, no. 11. e1005752. DOI: 10.1371/journal.pcbi.1005752
22. ELAN: a Professional Framework for Multimodality Research / P. Wittenburg, H. Brugman, A. Russel [et al.] // *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*. 2006. P. 1556–1559.
23. Writing in real time: modelling production processes / A. Matsuhashi (ed.). Norwood, NJ : Ablex Pub. Corp., 1987. 299 p.
24. Zhang M., Deane P. Process features in writing: Internal structure and incremental value over product features // *ETS Research Report Series*. 2015. Vol. 2015, no. 2. P. 1–12.

References

1. Glukhov V. P. *Osnovy psikholingvistiki* [Fundamentals of psycholinguistics]. Moscow, ACT, Astrel' Publ., 2005. 351 p.
2. Zagorovskaya O. V., Litvinova T. A. [Keyboard-speech behavior of a native speaker of the Russian language as an object of experimental idiolect psycholinguistics : towards the formation of a new scientific direction]. *Metodologiya sovremennogo yazykoznaniya – 3 : sb. st. v chest' yubileya V. A. Pishchal'nikovoi* [Methodology of modern linguistics-3: collection of articles in honor of the anniversary of V. A. Pishchalnikova]. Moscow, R. Valent Publ., 2021, pp. 72–81.
3. Zagorovskaya O. V., Litvinova T. A. Novye formy rechi v mediaprostranstve sovremennoi Rossii kak sfera realizatsii idiolekta i idiostilya nositelya russkogo yazyka [New forms of speech in the media space of modern Russia as a sphere of realization of the idiolect and idiostyle of a native Russian speaker]. *Izvestiya Voronezhskogo gosudarstvennogo pedagogicheskogo universiteta*, 2021, no. 1 (290), pp. 157–162.
4. Litvinova T. A. Stilemetricheskaya identifikatsiya idiolekta na osnove printsipov korpusnoi idiolektologii [Stylometric identification of an idiolect based on the principles of corpus idiolectology]. *Voprosy kriminologii, kriminalistiki i sudebnoi ekspertizy*, 2021, no. 1 (49), pp. 126–132.
5. Luriya A. R. *Osnovnye problemy neirolingvistiki* [Basic problems of neuro-linguistics]. Moscow, Izd-vo MGU, 1975. 256 p.
6. Kibrik A. A., Podlesskaya V. I. *Rasskazy o snovidenyakh. Korpusnoe issledovanie ustnogo russkogo diskursa* [Stories about dreams. Corpus research of oral Russian discourse]. Moscow, Yazyki slavyanskikh kul'tur Publ., 2009. 736 p.
7. Chukharev E. M. *Lingvostatisticheskie korrelyaty spontannosti v komp'yuterno-oposredovannom diskurse : na materiale russkoyazychnogo chata*. Diss. kand. filol. nauk [Linguostatistical correlates of spontaneity in computer-mediated discourse: on the material of the Russian-language chat. Cand. philol. sci. diss.]. St. Petersburg, 2009. 211 p.
8. Chanquoy L., Foulain J.-N., Fayol M. Writing in adults: A real-time approach. *Theories, models and methodology in writing research*. Amsterdam, Amsterdam University Press, 1996, pp. 36–43.
9. Chukharev-Hudilainen E. Pauses in spontaneous written communication: A key-stroke logging study. *Journal of Writing Research*, 2014, vol. 6, no. 1, pp. 61–84.
10. Cisarlu G., Olive T. *Le Processus de textualisation. Analyse des unit slinguistiques de performance*. Louvain-la-Neuve, De Boeck, 2018. 272 p.
11. Galbraith D., Baaijen V. M. Aligning keystrokes with cognitive processes in writing. *Observing writing: Insights from keystroke logging and handwriting*. Leiden, Brill Academic Publishers, 2019, pp. 306–325.
12. H. Guo, P. Deane, P. W. van Rijn e. a. Modeling basic writing processes from keystroke logs. *Journal of Educational Measurement*, 2018, vol. 55, pp. 194–216.
13. Hayes J. R. Modeling and remodeling writing. *Written Communication*, 2012, vol. 29, pp. 369–388.
14. Leijten M., Van Waes L. Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes. *Written Communication*, 2013, vol. 30, no. 3, pp. 358–392. doi: 10.1177/0741088313491692
15. Lindgren E., Knospe Y., Sullivan K. P. H. Researching Writing with Observational Logging Tools from 2006 to the Present. *Observing writing: insights from keystroke logging and handwriting*. Leidenm Brill Academic Publishers, 2019, pp. 1–29.
16. Litvinova T. Process-oriented characteristics of an idiolect for authorship attribution of heterogeneous texts: a pilot study. *CEUR Workshop Proceedings. Cep. "CMLS 2020 – Proceedings of the Computational Models in Language and Speech Workshop, co-located with 16th International Conference on Computational and Cognitive Linguistics, TEL 2020"*, 2020, pp. 3–16.
17. Litvinova T. RusIdiolect: A New Resource for Authorship Studies. *Lecture Notes in Networks and Systems*, 2021, vol. 186, pp. 14–23. doi: 10.1007/978-3-030-66093-2_2
18. Litvinova T., Rangel F., Rosso P. e. a. Overview of the Rusprofiling PAN at FIRE track on cross-genre gender identification in Russian. *CEUR Workshop Proceedings. Cep. "FIRE 2017 – Working Notes of FIRE 2017 – Forum for Information Retrieval Evaluation"*, 2017, pp. 1–7.
19. Mahrer R., De Angelis R., Del Lungo A. e. a. Editorial genesis: From comparing texts (product) to interpreting rewritings (process). *Writing(s) at the Crossroads*. Amsterdam, John Benjamins, 2015, pp. 151–170.

20. Olive T. Methods, tools and techniques for the on-line study of the writing process. *Writing: Processes, Tools and Techniques*. NY, Nova Publisher, pp. 1–18.
21. Rohart F., Gautier B., Singh A., Lk C. K-A mixOmics: An R package for ‘omics feature selection and multiple data integration. *PLoSComput Biol*, 2017, vol. 13, no. 11. e1005752. doi: 10.1371/journal.pcbi.1005752
22. Wittenburg P., Brugman H., Russel A. e. a. ELAN: a Professional Framework for Multimodality Research. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, 2006, pp. 1556–1559.
23. Matsuhashi A., ed. *Writing in real time: modelling production processes*. Norwood, NJ, Ablex Pub. Corp., 1987. 299 p.
24. Zhang M., Deane P. Process features in writing: Internal structure and incremental value over product features. *ETS Research Report Series*, 2015, vol. 2015, no. 2, pp. 1–12.

Поступила в редакцию 12.08.2021

Подписана в печать 25.09.2021

**KEYBOARD-MEDIATED SPEECH AS ONE OF THE MAIN AREAS
OF IMPLEMENTATION OF THE IDIOLECT OF THE MODERN RUSSIAN LANGUAGE SPEAKER:
PROBLEMS OF THE TEXT GENERATION STUDY AND FIRST RESULTS**

Tatiana A. Litvinova¹, Olga V. Zagorovskaya²

Voronezh State Pedagogical University^{1, 2}
Voronezh, Russia

¹*Cand. Philolog. Sci., Head of Research Laboratory of Corpus Idiolectology, e-mail: centr_rus_yaz@mail.ru*

²*Dr. Philolog. Sci., Professor of the Department of Russian Language,
Modern Russian and Foreign Literature, e-mail: olzagor@yandex.ru*

Abstract. Currently, one of the main forms of the implementation of the idiolect (by which we mean the productive individual speech activity of the subject) of the Russian speaker is the production of the so-called keyboard texts, that is, texts created using the keyboard of an electronic device. Linguists explore the various characteristics of such texts as products, while the processes of generation of such texts remain unexplored. Modern software and hardware tools allow scientists to visualize the process of generating texts to a certain extent and thereby expand the scientific understanding of the mechanisms of text generation. Such research is actively carried out in world science in the framework of the direction called writing research. The Russian language is on the periphery of such studies. The article provides a brief overview of the state-of-the-art in the field of text generation in modern world science, highlights the main theoretical and methodological problems in this area, and presents the results of pilot experimental studies of the processes of text generation in Russian. It is shown that the solution of one of the fundamental issues of linguistics, which is the problem of text generation, including the problem of determining the minimum units of text generation is impossible without consideration of the idiolect factor.

Key words: idiolect, corpus linguistics, idiolectology, text generation, psycholinguistics, keyboard text.

Cite as: Litvinova T. A., Zagorovskaya O. V. Keyboard-mediated speech as one of the main areas of implementation of the idiolect of the modern Russian language speaker: problems of the text generation study and first results. *Izvestiya Voronezhskogo gosudarstvennogo pedagogicheskogo universiteta* [Izvestia Voronezh State Pedagogical University], 2021, no. 3, pp. 127–133. (in Russian). DOI 10.47438/2309-7078_2021_3_127

Received 12.08.2021

Accepted 25.09.2021