

**ЛИНГВОДИДАКТИКА И НОВАЦИИ. ПСИХОЛОГИЧЕСКИЕ ОСНОВЫ ИЗУЧЕНИЯ
ЯЗЫКОВ И КУЛЬТУР | LINGUODIDACTICS AND INNOVATIONS.
PSYCHOLOGICAL BASIS OF LEARNING LANGUAGES AND CULTURES**

Научная статья | Original paper

**Компьютерная лингвистика в психологии: ключ к пониманию
языка и поведения человека**

К.С. Ивашко✉, С.А. Изосимова, В.Н. Пигуз, К.В. Ковалева

Федеральное государственное бюджетное научное учреждение «Институт проблем
искусственного интеллекта», Донецк, Российская Федерация

✉ kristi_8@mail.ru

Резюме

Актуальность определяется экспоненциальным ростом текстовых данных в цифровой среде, требующим автоматизированных методов анализа для выявления скрытых закономерностей. **Цель работы** – анализ возможностей и ограничений методов компьютерной лингвистики в исследованиях личности. **Гипотеза** заключается в том, что автоматизированный анализ текста позволяет выявлять психологические особенности авторов. **Методы и материалы:** использованы Национальный корпус русского языка (НКРЯ) для лингвистического анализа частотности, методы NLP (тональный анализ, выделение ключевых слов) для определения эмоциональной окраски и выявления ключевых тем. Проведен анализ синтаксических конструкций для выявления особенностей стиля письма. **Результаты:** анализ НКРЯ позволяет выявлять преобладающие эмоциональные состояния и когнитивные стили авторов. Частое использование местоимений первого лица единственного числа может указывать на эгоцентричность. Тональный анализ и выделение ключевых слов позволяют определять эмоциональную окраску текста. Анализ синтаксических конструкций, таких как предпочтение пассивного залога, может свидетельствовать об избегании ответственности. Сложные предложения указывают на стремление к детализации. **Новизна исследования** – систематизация и адаптация методов компьютерной лингвистики для проектирования динамических когнитивных систем диагностики и немедикаментозной коррекции. **Выводы:** компьютерная лингвистика предоставляет мощные инструменты для исследования личности, требующие осторожного и осознанного применения с учетом этических аспектов.

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

Ключевые слова: компьютерная лингвистика, машинное обучение, анализ текста, частотность, социолингвистика, когнитивные системы, диагностика, коррекция, стрессогенные факторы, интеллектуальные системы

Для цитирования: Ивашко, К.С., Изосимова, С.А., Пигуз, В.Н., Ковалева, К.В. (2025). Компьютерная лингвистика в психологии: ключ к пониманию языка и поведения человека. *Язык и текст*, 12(2), 140—162. <https://doi.org/10.17759/langt.2025120213>

Computational linguistics in psychology: a key to understanding language and human behavior

K.S. Ivashko✉, S.A. Izosimova, V.N. Piguz, K.V. Kovaleva

Federal State Budgetary Scientific Institution «Institute of Artificial Intelligence Problems»,
Donetsk, Russian Federation

✉ kristi_8@mail.ru

Abstract

The relevance is determined by the exponential growth of textual data in the digital environment, requiring automated analysis methods to reveal hidden patterns. **The aim** of the paper is to analyze the possibilities and limitations of computational linguistics methods in personality research. **The hypothesis** is that automated text analysis can reveal psychological characteristics of authors. **Methods and materials:** we used the National Corpus of the Russian Language (NCLR) for linguistic frequency analysis, NLP methods (tone analysis, keyword extraction) to determine emotional coloring and identify key themes. Syntactic constructions were analyzed to identify peculiarities of writing style. **Results:** the analysis of NLP allows to identify the prevailing emotional states and cognitive styles of the authors. Frequent use of first person singular pronouns may indicate egocentricity. Tonal analysis and key word selection allow us to determine the emotional coloring of the text. Analysis of syntactic constructions, such as the preference for passive voice, may indicate avoidance of responsibility. Complex sentences indicate a desire for detail. **The novelty** of the study is the systematization and adaptation of computational linguistics methods for designing dynamic cognitive systems for diagnosis and non-medical correction. **Conclusions:** computational linguistics provides powerful tools for personality research, requiring careful and informed application with ethical considerations.

Keywords: computational linguistics, machine learning, text analysis, frequency, sociolinguistics, cognitive systems, diagnosis, correction, stressors, intelligent systems

For citation: Ivashko, K.S., Izosimova, S.A., Piguz, V.N., Kovaleva, K.V. (2025). Computational linguistics in psychology: a key to understanding language and human behavior. *Language and Text*, 12(2), 140—162. (In Russ.). <https://doi.org/10.17759/langt.2025120213>

Введение

Специализированные методы обработки информации в интеллектуальных системах стали неотъемлемой частью исследовательского ландшафта, позволяют исследователям автоматизировать рутинные операции, выявлять скрытые закономерности и повышать объективность интерпретаций. Эти методы, опирающиеся на достижения компьютерной лингвистики и машинного обучения, открывают новые горизонты для изучения человеческого мышления, эмоций и поведения. Важнейшая задача лингвистического корпуса – предоставление лингвостатистической информации о частоте употребления различных грамматических форм, отдельных слов или устойчивых словосочетаний, то есть вычисление их частотности. Эта возможность помогает выявлять нюансы в семантике синонимов, устанавливать типичные контексты для синонимичных слов, а также различать жанровые и стилистические особенности и оттенки значений лексических единиц. В социолингвистических исследованиях частотность позволяет установить, насколько часто определенная языковая единица используется конкретной социальной, возрастной или гендерной группой. В историческом плане эта функция дает возможность проследить этимологию и эволюцию значения языковой единицы. Результатом частотного анализа текста является получение статистических данных. Следовательно, в исследованиях, использующих корпусный анализ, фундаментом научной гипотезы являются точные и конкретные эмпирические данные. Одним из ключевых преимуществ использования специализированных методов анализа текста является возможность обработки больших объемов информации. Нейросети и алгоритмы машинного обучения способны анализировать тысячи текстовых документов за считанные минуты, выявляя в них ключевые темы, настроения и взаимосвязи. Это позволяет исследователям получать более полную и объективную картину изучаемого явления. Кроме того, специализированные методы анализа текста позволяют минимизировать субъективность интерпретаций. Алгоритмы машинного обучения, обученные на больших объемах данных, способны выявлять закономерности, которые могут быть незаметны для человеческого глаза. Это повышает надежность и валидность результатов исследования. Наконец, специализированные методы анализа текста позволяют исследователям задавать более сложные и тонкие вопросы. Например, можно изучать, как меняется эмоциональная окраска текста в зависимости от контекста, или выявлять скрытые связи между различными понятиями. Это открывает новые возможности для углубленного понимания человеческой психики (Бах, 2009), (Кузьмина, Лифшиц, Костенко, 2022). Статья Толдова С.Ю. и Ляшевской О.Н. "Современные проблемы и тенденции компьютерной лингвистики" представляет собой обзор актуальных направлений в области, находящейся на стыке лингвистики и информационных технологий. Авторы рассматривают широкий спектр задач, стоящих перед компьютерной лингвистикой, включая автоматическую обработку текста, машинный перевод, создание лингвистических ресурсов и разработку алгоритмов для анализа и синтеза языка. Особое внимание уделяется проблемам, связанным с неоднозначностью языка, необходимостью учета контекста и знаний о мире при обработке текста. В статье анализируются различные подходы к решению этих проблем, такие как использование статистических методов, нейронных сетей и лингвистических правил. Авторы подчеркивают важность создания больших лингвистических корпусов и

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

разработки эффективных методов их разметки для обучения компьютерных систем (Толдова, Ляшевская, 2014).

Новизна исследования – систематизация и адаптация методов компьютерной лингвистики и обработки естественного языка для решения специфических задач проектирования динамических когнитивных систем диагностики и немедикаментозной коррекции в реальном времени с учетом стрессогенных факторов, создание методов генерирования таксономии онтологий, эмоциональных паттернов, формирование экспериментальной базы данных мультимодальных возможностей, обработки информации нейросетевыми технологиями.

Цель работы – проанализировать карту возможностей и ограничений, возникающих при применении методов компьютерной лингвистики, а также рассмотреть основные методы, доказавшие свою эффективность в теоретических, методических и эмпирических исследованиях личности.

Материалы и методы. В работе рассмотрены методы компьютерной лингвистики и обработки естественного языка (NLP) для анализа текстовых данных, такие как Национальный корпус русского языка (НКРЯ) для лингвистического анализа частотности употребления слов и грамматических конструкций, а также методы NLP, такие как тональный анализ и выделение ключевых слов, использовались для определения эмоциональной окраски текста и выявления ключевых тем, применяемых для использования систем автоматической диагностики и прогнозирования психологического состояния индивида. Исследование включало анализ синтаксических конструкций для выявления особенностей стиля письма и склонностей к избеганию ответственности или подчеркиванию важности действий.

Актуальность

Цифровой мир сталкивается с беспрецедентным потоком текстовой информации, генерируемой людьми в социальных сетях, блогах и онлайн-коммуникациях. Методы компьютерной лингвистики и обработки естественного языка открывают новые горизонты в анализе этих данных, предоставляя инструменты для выявления закономерностей и инсайтов, которые были бы недоступны при ручном анализе. Обработка больших данных сопряжена с морфологической сложностью и синтаксической вариативностью, разработкой специализированных словарей и алгоритмов, учитывающих специфику лингвистики, стремлением максимально повысить точность и надежность анализа текстовых данных. При разработке исследований важно учитывать этические аспекты, связанные с использованием текстовых данных, особенно если они содержат личную информацию. Необходимо обеспечить конфиденциальность и анонимность участников исследования, а также получить их согласие на использование их текстов. В целом, компьютерная лингвистика предоставляет мощный инструмент для исследования личности, но требует осторожного и осознанного применения, с учетом возможностей и ограничений, а также этических аспектов.

Результаты

Национальный корпус русского языка (НКРЯ) представляет собой ценнейший ресурс для психологов, изучающих личность. Благодаря методам компьютерной лингвистики и обработки естественного языка, становится возможным автоматизировать анализ больших объемов текстовых данных, извлекая значимую информацию о психологических особенностях авторов и персонажей (Плунгян, 2005). В частности, анализ частотности употребления определенных слов и грамматических конструкций позволяет выявлять преобладающие эмоциональные состояния, когнитивные стили и ценностные ориентации. Например, частое использование местоимений первого лица единственного числа может свидетельствовать об эгоцентричности, а преобладание слов, описывающих чувства и эмоции, – об эмоциональной лабильности. Методы NLP, такие как тональный анализ и выделение ключевых слов, дают возможность определять эмоциональную окраску текста и выявлять темы, которые наиболее важны для автора. Это, в свою очередь, позволяет строить гипотезы о его личностных чертах и мотивациях. Использование лингвистических особенностей текста для психологического анализа открывает новые горизонты в понимании личности.

Анализ синтаксических конструкций, например, предпочтение пассивного залога активному, может указывать на склонность к избеганию ответственности или на стремление дистанцироваться от действий, например, в официальных отчетах или новостных сводках, где важна объективность. Активный залог, напротив, демонстрирует прямоту и четкое указание на субъект действия, что может свидетельствовать об уверенности и готовности брать на себя ответственность, например в таком высказывании, как "Я разработал этот проект". Здесь фокус смещен на исполнителя, а не на сам проект. В деловой переписке или официальных документах активный залог придает заявлениям вес и убедительность. Он делает текст более динамичным и легким для восприятия, поскольку позволяет читателю сразу понять, кто и что делает. В отличие от страдательного залога, где акцент смещается на объект действия, активный залог подчеркивает роль исполнителя. Это особенно важно в ситуациях, когда необходимо выделить вклад конкретного человека или группы в достижение результата, например в таком предложении как: "Команда инженеров разработала инновационную технологию". Здесь активный залог акцентирует внимание на команде и их непосредственном участии в разработке. Однако стоит помнить, что выбор между активным и страдательным залогом зависит от конкретной цели и контекста. В некоторых случаях использование страдательного залога может быть более уместным, например, когда необходимо скрыть исполнителя или подчеркнуть важность самого действия, а не того, кто его совершил. Использование сложных, многосоставных предложений с обилием придаточных частей может указывать на стремление к детализации и всестороннему охвату темы, в таком предложении как, например, "Поскольку экономические показатели демонстрируют устойчивый рост, что, в свою очередь, благоприятно сказывается на инвестиционном климате, который определяет приток капитала в развивающиеся отрасли, правительство планирует разработать дополнительные меры поддержки, направленные на стимулирование инноваций, чтобы обеспечить дальнейшее развитие и конкурентоспособность страны на мировом рынке". Подобные конструкции позволяют автору выразить сложные взаимосвязи между явлениями и аргументировать свою

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

точку зрения с максимальной точностью, поскольку каждое придаточное предложение добавляет новый слой информации, углубляя понимание читателем рассматриваемого вопроса. Однако, злоупотребление сложными предложениями может привести к перегруженности текста и затруднить его восприятие. В то же время, лаконичные и простые предложения, напротив, способствуют быстрому усвоению информации и создают ощущение динамичности. Они идеально подходят для передачи ключевых тезисов и формирования четкого, запоминающегося образа, например, в таком предложении как: "Солнце село". Сила коротких фраз заключается в их непосредственности и способности мгновенно привлечь внимание. Короткие фразы действуют как катализатор, усиливающий эмоциональный отклик и заставляющий задуматься, в таком высказывании как: "Вот он, момент истины".

Таким образом, выбор между сложными и простыми предложениями зависит от поставленных целей и характера текста, а умелое сочетание разных стилистических приемов позволяет добиться максимальной выразительности и эффективности коммуникации. Однако, злоупотребление подобными конструкциями может затруднять понимание текста и свидетельствовать о желании усложнить информацию. Простые, лаконичные предложения, напротив, облегчают восприятие и подчеркивают ясность мысли. Их использование часто наблюдается в текстах, ориентированных на широкую аудиторию, или в материалах, требующих быстрого понимания, например, в рекламных слоганах или инструкциях. Синтаксические повторы, анафоры и эпифоры могут усиливать эмоциональное воздействие текста и использоваться для акцентирования ключевых идей. Они часто встречаются в художественной литературе и публицистике, где важна выразительность и убедительность, например в таком предложении как "Свобода, свобода, свобода!" - восклицает герой, подчеркивая свою жажду освобождения от угнетения. Синтаксический повтор, в частности, создает ритмичность и музыкальность, что способствует лучшему запоминанию. Фразы или предложения, построенные по одной и той же модели, словно вбивают в сознание слушателя или читателя основную мысль. Анафора, повторение слова или фразы в начале нескольких предложений или абзацев, акцентирует внимание на определенном аспекте повествования, создавая ощущение нарастающего напряжения или усиления чувства.

Выбор синтаксических конструкций, таким образом, является мощным инструментом для формирования смысла и воздействия на читателя, отражая авторские намерения и цели. Разнообразие словарного запаса и частота использования сложных грамматических конструкций могут свидетельствовать об уровне интеллектуального развития и когнитивной гибкости. Применение NLP в психолингвистике позволяет автоматизировать процесс анализа больших объемов текста, что особенно актуально в эпоху цифровых коммуникаций. Анализ текстов в социальных сетях, блогах и форумах позволяет выявлять закономерности в проявлении психологических характеристик у различных групп людей.

Однако, стоит помнить об ограничениях такого подхода. Интерпретация результатов анализа требует осторожности и учета контекста. Лингвистические особенности могут быть обусловлены не только психологическими факторами, но и стилистическими предпочтениями автора, жанром текста и культурными нормами. Несмотря на ограничения, психолингвистический анализ текста с использованием NLP является мощным инструментом для исследования личности и позволяет получить ценную информацию о

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

когнитивных, эмоциональных и поведенческих особенностях человека. Кроме того, НКРЯ предоставляет возможность сравнивать тексты разных авторов или персонажей, выявляя различия в их языковых паттернах и, следовательно, в их психологических характеристиках. Это открывает перспективы для изучения влияния личностных особенностей на стиль письма и речевое поведение.

Безусловно, в качестве источников данных для анализа текстов в психологических исследованиях могут использоваться публикации и сообщения в социальных сетях, а также дневниковые записи, личные письма, протоколы психотерапевтических сессий, расшифровки интервью, результаты контент-анализа художественной литературы и киносценариев также представляют ценный материал для изучения. Особую нишу занимают данные, полученные в ходе экспериментов. В частности, это могут быть письменные ответы участников на различные стимулы, эссе на заданную тему или тексты, сгенерированные в процессе выполнения когнитивных задач.

Методы анализа таких текстов варьируются в зависимости от целей исследования. Это может быть как качественный анализ, направленный на выявление общих тем и закономерностей, так и количественный, основанный на подсчете частоты употребления определенных слов и фраз. При этом важно учитывать контекст создания текста, личностные особенности автора и возможные искажения, связанные с субъективностью восприятия. Тщательный учет этих факторов позволит получить более точные и надежные результаты. В конечном итоге, выбор источника данных и метода анализа определяется конкретной исследовательской задачей и доступными ресурсами. Главное – обеспечить репрезентативность выборки и валидность полученных результатов.

В своем труде "Экспериментальная психология" Роберт Вудвортс, выступая в роли стороннего наблюдателя, тщательно документирует различные исследования, стремясь максимально точно отразить процесс изучения конкретных психологических вопросов. Книга Вудвортса охватывает темы, связанные с психологией ощущений, памяти, мышления, обучения, эмоций, внимания и другие аспекты. В ней широко используется психофизиологический материал, особенно при изучении функций органов чувств и органических основ эмоций. Отличительной чертой данной работы является стремление к систематизации обширного объема эмпирических данных, накопленных в психологии к тому времени. Вудвортс акцентирует внимание на критическом разборе экспериментальных методик, подчеркивая необходимость строгого контроля переменных и применения статистической обработки данных для обеспечения надежности результатов. Его труд значительно повлиял на формирование методологических принципов в экспериментальной психологии (Вудвортс, 1950).

В процессе разработки любой лингвистической модели крайне полезно создать специализированный алгоритм, способный определять и сохранять оптимальную последовательность шагов, рекомендованную исследователем, что позволит достичь наиболее эффективных результатов. Важнейшим шагом в любом психологическом исследовании является обработка и интерпретация собранных данных, как в количественном, так и в качественном аспектах. Именно содержательный анализ итогов работы представляет собой наиболее трудоемкий, ответственный и креативный процесс. Статистические методы в психологии – неотъемлемая часть обработки информации,

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

предоставляющая числовые данные, требующие глубокой интерпретации и обоснования с точки зрения психологии. Применение статистического инструментария психологом-практиком неразрывно связано с многообразием его деятельности: от тонкой диагностики и бережной развивающей работы до чуткого консультирования, целительной психотерапии и кропотливых научных изысканий. Рассмотрим же пристальнее, как статистика становится незаменимым союзником психолога в каждом из этих направлений. В диагностике статистика помогает выявлять отклонения от нормы, сравнивать индивидуальные показатели с групповыми, определять структуру личности и прогнозировать поведение. Стандартизированные тесты и опросники, основанные на статистических данных о валидности и надежности, позволяют с большей уверенностью судить о наличии тех или иных психологических особенностей. Еще Голдберг Л.Р. в своей работе "Язык и индивидуальные различия: Поиск универсалий в лексиконе личности" акцентирует внимание на важности лексического подхода к изучению структуры личности. Он утверждает, что наиболее значимые индивидуальные различия, которые люди используют в повседневной жизни для описания друг друга, закодированы в языке. Голдберг подчеркивает, что анализ словарного запаса, используемого для характеристики личности, позволяет выделить основные измерения личностных черт. Этот подход, по его мнению, является эмпирически обоснованным и позволяет создать надежную и валидную модель личности.

В своей работе Голдберг исследует, как лексические единицы, описывающие личностные качества, организованы в различные категории и как эти категории соотносятся друг с другом. Он стремится выявить универсальные структуры личности, которые проявляются в разных культурах и языках. Он также обсуждает проблемы, связанные с лексическим подходом, такие как синонимия, полисемия и влияние контекста на значение слов. Несмотря на эти трудности, Голдберг убежден, что лексический подход является ценным инструментом для понимания структуры личности и индивидуальных различий (Goldberg, 1981). В развивающей работе и консультировании статистический анализ может использоваться для оценки эффективности применяемых методов и программ. Сравнивая результаты до и после вмешательства, психолог получает объективное подтверждение позитивных изменений или, напротив, выявляет необходимость корректировки подхода.

В психотерапии статистика помогает отслеживать динамику состояния клиента, оценивать эффективность различных терапевтических техник и выявлять факторы, способствующие или препятствующие прогрессу. Анализ больших массивов данных позволяет исследователям разрабатывать более эффективные протоколы лечения. В научных изысканиях статистические методы являются основой для проверки гипотез, выявления закономерностей и установления причинно-следственных связей. Без статистического анализа невозможно представить себе ни одно серьезное психологическое исследование, претендующее на объективность и научную обоснованность. В арсенале компьютерной лингвистики и обработки естественного языка статистические методы играют ключевую роль в анализе больших текстовых массивов. Они позволяют извлекать ценную информацию, автоматизировать процессы и решать сложные лингвистические задачи (Витко, Калугин, Щебетенко, 2022). Существующая пятифакторная модель личности, так называемая «Большая пятерка», которая получила широкое распространение и при этом продемонстрировала высокую воспроизводимость в различных культурах. Однако, данные,

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

полученные в разных странах и культурах, все чаще выявляют несоответствия и противоречия. В результате, появляются модели, предлагающие иное количество факторов. К ним относятся шестифакторные и семифакторные подходы, а также модели, стремящиеся к межкультурной универсальности, такие как "пан-культурная тройка", двухфакторная и однофакторная модели. Несмотря на популярность "Большой Пятерки", критики отмечают ее ограничения в объяснении всего спектра личностных проявлений. Шестифакторные модели, например, HEXACO, добавляют фактор "Честность-Скромность", что позволяет более точно описывать моральные качества. Семифакторные модели, в свою очередь, стремятся дифференцировать отдельные аспекты, недостаточно представленные в "Большой Пятерке". "Пан-культурная тройка" фокусируется на факторах, наиболее устойчивых к культурным различиям, предлагая упрощенную, но универсальную модель. Двухфакторные и однофакторные модели, как правило, выделяют общие диспозиционные тенденции, такие как "позитивность" или "стабильность", стремясь к максимальной обобщенности. Выбор оптимальной модели зависит от целей исследования и контекста. "Большая Пятерка" остается полезной для широких популяционных исследований, в то время как альтернативные модели могут быть более подходящими для изучения специфических аспектов личности или для работы с определенными культурными группами¹. Дальнейшие исследования необходимы для уточнения структуры личности и разработки моделей, учитывающих как универсальные, так и культурно-специфичные аспекты.

Одним из наиболее распространенных методов является частотный анализ, который позволяет выявлять наиболее употребляемые слова и словосочетания в тексте. Это может быть полезно для определения тематики текста, выявления ключевых понятий и оценки стилистических особенностей. Например, в исследовательской работе А.Ю. Тычкова и П.П. Чуракова, рассматривающей "Использование частотно-временных алгоритмов анализа при решении задач обработки ЭЭГ", в контексте методов компьютерной лингвистики и обработки текста на естественном языке, указывается следующее: Применение частотно-временных алгоритмов (ЧВА) для анализа электроэнцефалограмм (ЭЭГ) открывает новые перспективы в выявлении скрытых паттернов мозговой активности. Эти алгоритмы, такие как вейвлет-преобразование и кратковременное преобразование Фурье, позволяют одновременно анализировать частотные и временные характеристики сигнала, что критически важно для понимания динамических процессов в мозге. В частности, авторы подчеркивают эффективность ЧВА в задачах классификации ЭЭГ-сигналов, связанных с различными когнитивными состояниями или патологиями. Использование этих методов позволяет выделить информативные признаки, устойчивые к шумам и артефактам, что повышает точность диагностики и мониторинга состояния пациентов. Таким образом, интеграция ЧВА в анализ ЭЭГ представляет собой перспективное направление, способствующее развитию как клинической практики, так и фундаментальных исследований в области нейронаук.

¹ Что Такое "Большая Пятерка" Личностных Качеств? [Электронный ресурс]. URL: <https://www.b17.ru/article/617847/>

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

Другой важный метод – статистическое моделирование языка, которое позволяет строить вероятностные модели последовательностей слов. Эти модели используются в задачах машинного перевода, распознавания речи и генерации текста. В своей работе "Статистическое моделирование качества взаимодействия в речевых диалоговых системах: сравнение классификаторов, основанных на (условных) скрытых марковских моделях, и машин опорных векторов" С. Ультес, А. Шмитт, Р. Эль Хаб, В. Минкер обращают внимание на проблему оценки качества взаимодействия в речевых диалоговых системах, предлагая сравнение двух подходов к классификации. Авторы исследуют применение скрытых марковских моделей (НММ) и машин опорных векторов (SVM) для автоматической оценки качества диалога, учитывая такие факторы, как успешность выполнения задачи, эффективность диалога и удовлетворенность пользователя. В статье детально рассматриваются особенности построения классификаторов на основе НММ и SVM, а также проводится экспериментальное сравнение их производительности на реальных данных речевых диалоговых систем. Особое внимание уделяется анализу влияния различных параметров моделей и характеристик данных на точность классификации. Результаты исследования показывают, что оба подхода обладают своими преимуществами и недостатками, а выбор оптимального метода зависит от конкретных требований к точности и вычислительной сложности. Авторы подчеркивают важность дальнейших исследований в области статистического моделирования качества взаимодействия для улучшения речевых диалоговых систем (Ультес и др., 2012).

Кластеризация текстов – еще один пример применения статистических методов в NLP. Она позволяет автоматически группировать тексты по темам, стилям или авторам, что полезно для организации больших текстовых коллекций и поиска информации. Исследования, представленные в Springer, Web of Science и Scopus, демонстрируют возможности автоматизированного выявления личностных черт, эмоциональных состояний и когнитивных особенностей на основе анализа текстов, в таких журналах как *Journal of Personality and Social Psychology*, *Psychological Science*, *Journal of Research in Personality*, регулярно публикуются статьи, посвященные применению методов компьютерной лингвистики и обработки естественного языка (NLP) в психологии личности, а также специализированных изданиях, посвященных обработке естественного языка, например, *Computational Linguistics* и *Natural Language Engineering*. Примером может служить исследование, опубликованное в *Journal of Personality and Social Psychology*, где авторы анализируют взаимосвязь между лингвистическими паттернами в текстах и определенными чертами личности. В статье «Automatic personality assessment through social media language» Парк Г. и соавторы исследуют возможность автоматической оценки личности на основе анализа текстов из социальных сетей. Авторы подчеркивают, что лингвистические паттерны, проявляющиеся в письменной речи, могут служить надежным индикатором психологических характеристик индивида. Исследование фокусируется на выявлении корреляций между использованием определенных слов, фраз и стилистических приемов и пятью основными чертами личности (Big Five): открытость опыту, добросовестность, экстраверсия, доброжелательность и нейротизм. Авторы применяют методы машинного обучения для построения моделей, способных прогнозировать личностные характеристики на основе анализа текстовых данных. Результаты показывают, что модели, обученные на

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

языке социальных сетей, могут достаточно точно оценивать личностные черты. Это открывает перспективы для использования автоматической оценки личности в различных областях, включая HR, маркетинг и психологические исследования (Park et al., 2015).

Другая статья, представленная в *Psychological Science*, демонстрирует возможности NLP для выявления скрытых эмоциональных состояний на основе анализа письменной речи. В исследовании Ениколопова С.Н. и коллег, опубликованном в журнале «Психологические исследования» (2018), анализируются лингвистические характеристики текстов, созданных психически больными и здоровыми людьми. Целью работы являлось выявление паттернов в письменной речи, которые могли бы служить индикаторами психического состояния. Авторы использовали методы NLP для анализа таких параметров, как частота употребления определенных слов, структура предложений и стилистические особенности. Результаты показали, что существуют значимые различия в лингвистических характеристиках текстов, созданных разными группами испытуемых. Эти различия могут быть использованы для разработки инструментов автоматической диагностики и мониторинга психического здоровья. Исследование подчеркивает потенциал NLP в области психиатрии и психологии, открывая новые перспективы для раннего выявления и лечения психических расстройств (Ениколопов и др., 2018). Круг исследователей, активно работающих в данной области, постоянно расширяется. Их работы открывают новые перспективы для изучения личности, предлагая объективные и масштабируемые методы анализа текстовых данных, что позволяет получать более глубокое понимание индивидуальных различий и психологических процессов. Среди современных исследователей, активно работающих в этой области, можно выделить Jamie Pennebaker, Yla Tausczik, James W. Pennebaker, который разрабатывает методы автоматического анализа текстов для изучения психологических процессов. Эти и другие ученые вносят значительный вклад в понимание того, как язык отражает и формирует личность, и какие возможности и ограничения существуют при использовании методов NLP для решения задач психологии.

В своей работе Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015). *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates (www.LIWC.net) (Kraus, Feuerriegel, 2019), (Pennebaker et al., 2015) подробно описывают разработанную ими программу LIWC2015, ставшую общепризнанным инструментом для анализа текстов. Эта программа позволяет автоматически анализировать большие объемы текстовых данных и выявлять частоту использования различных категорий слов, отражающих психологические и эмоциональные состояния автора. LIWC2015 анализирует тексты, подсчитывая процентное содержание слов, относящихся к предопределенным категориям, таким как эмоции (позитивные, негативные), когнитивные процессы, социальные процессы и другие. Результаты анализа могут быть использованы для исследования широкого спектра явлений, включая психологическое состояние людей, межличностные отношения, культурные различия и эффективность коммуникации. LIWC2015 находит применение в психологии, социологии, лингвистике, маркетинге и других областях, где требуется анализ текстовых данных для выявления скрытых закономерностей и тенденций. Кроме того, LIWC2015 активно используется для изучения влияния COVID-19 на психическое здоровье населения. Исследования анализируют тексты новостных статей, сообщений в социальных сетях и личных дневников, чтобы оценить уровень тревожности, страха и других негативных

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

эмоций, связанных с пандемией. Методология LIWC2015 также совершенствуется. Разрабатываются новые категории и словари, адаптированные для конкретных контекстов и языков, что позволяет повысить точность и релевантность анализа.

Современные исследования подчеркивают важность учета культурных и контекстуальных особенностей при интерпретации результатов LIWC2015. Подчеркивается необходимость калибровки и адаптации программы для различных языковых и культурных групп, чтобы избежать искажений и неверных выводов.

Также Pennebaker и Boyd акцентируют внимание на использовании лингвистического анализа для определения психологического состояния авторов текстов. Они разрабатывают алгоритмы, способные выявлять закономерности в использовании определенных слов и фраз, которые коррелируют с конкретными эмоциональными переживаниями. Например, в работе *The Narrative Arc: Revealing Core Narrative Structures Through Text Analysis* Ryan L. Boyd, Kate G. Blackburn, James W. Pennebaker подробно описывается возможность автоматического анализа текстовых данных для выявления ключевых элементов нарративной структуры. Авторы предлагают использовать методы компьютерной лингвистики и статистического анализа для определения эмоциональной окраски текста, выявления персонажей и их взаимоотношений, а также отслеживания динамики развития сюжета. Особое внимание уделяется концепции "нарративной дуги", представляющей собой графическое отображение изменения эмоционального тона текста на протяжении повествования. Анализ формы этой дуги позволяет исследователям выявлять типичные паттерны развития сюжета, такие как взлет и падение, триумф над невзгодами или трагическое поражение. Применение этих методов возможно в различных областях, от литературоведения и журналистики до маркетинга и политической коммуникации. Автоматический анализ нарративной структуры позволяет быстро обрабатывать большие объемы текстовых данных и выявлять скрытые закономерности в повествованиях.

В работах Т. Yarkoni, Д. Ванг, Дж. Д. Вайс, М. Мюллер, П. Рам, У. Гейер, К. Дуган, Й. Таучик описываются методы компьютерной лингвистики и обработки естественного языка: возможности и ограничения для задач психологии личности. В статье "Personality Judgments Based on Physical Appearance" (Yarkoni, 2010) (Yarkoni, Westfall, 2017) исследуется, как внешность влияет на восприятие личности. "Words as Social Signals: Automatic Language Use Measures Predict Real-World Social Integration" (Borghi et al., 2018) демонстрирует, как анализ языка может предсказать социальную интеграцию. "Ecological Momentary Assessment Using Text Messages" описывает использование текстовых сообщений для оценки эмоционального состояния в реальном времени. Эти работы подчеркивают растущую роль вычислительных методов в изучении человеческой личности. Анализ больших текстовых массивов, машинное обучение и автоматизированный искусственный интеллект открывают новые горизонты для понимания психологических особенностей. В частности, исследования показывают, как лексический анализ текста может выявить корреляции между языковыми паттернами и чертами личности. Вместе с тем, авторы акцентируют внимание на необходимости критического осмысления возможностей и ограничений этих методов. Важно учитывать, что предсказательная сила моделей машинного обучения не всегда подразумевает глубокое понимание механизмов, лежащих в основе наблюдаемых связей. Необходимо избегать упрощенных интерпретаций и стремиться

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

к интеграции вычислительных подходов с традиционными методами психологических исследований. Сотрудничество между специалистами в области психологии и компьютерных наук становится ключевым фактором успешного применения этих технологий (Ван, 2019).

В дальнейшем, подобные исследования могут привести к созданию автоматизированных систем для анализа и оценки текстов, а также к разработке новых методов повествования, основанных на эмпирических данных о восприятии аудитории. Разработанные алгоритмы и методы компьютерной лингвистики и обработки естественного языка, по сути, становятся динамической дорожной картой, адаптирующейся к результатам каждого этапа. Они анализируют промежуточные данные, оценивают эффективность примененных методов и на основе этого корректирует дальнейшую последовательность действий. Ключевым преимуществом является возможность избежать повторения неэффективных шагов и фокусироваться на наиболее перспективных направлениях. Алгоритм может учитывать различные факторы, такие как объем данных, вычислительные ресурсы и специфические требования к конечной модели. Такой адаптивный подход позволяет не только оптимизировать процесс обработки текста, но и существенно повысить точность и релевантность конечных результатов. Например, при анализе тональности текста алгоритм может изначально опираться на словарь ключевых слов, но в процессе обучения адаптироваться к контексту и учитывать более сложные лингвистические конструкции, определяющие эмоциональную окраску высказывания. Важную роль в этом процессе играют методы машинного обучения, позволяющие алгоритмам самостоятельно извлекать знания из данных и строить модели, предсказывающие результаты с высокой точностью. Эти модели могут учитывать различные факторы, такие как частота употребления слов, их сочетаемость, синтаксическая структура предложений и даже культурный контекст. В своей работе Kraus M., Feuerriegel S. предлагают модель глубокой нейронной сети, которая учитывает иерархическую структуру текста, выявляя связи между различными его частями. Это позволяет более точно определять общее эмоциональное содержание, особенно в сложных и неоднозначных текстах, где традиционные методы могут давать сбои. В частности, модель использует рекурсивные нейронные сети для обработки деревьев RST, что позволяет ей эффективно улавливать контекст и взаимосвязи между предложениями. Авторы демонстрируют, что их подход превосходит существующие методы анализа тональности на нескольких наборах данных, показывая, что учет структуры дискурса может значительно повысить точность. Однако, они также отмечают, что построение деревьев RST является сложной задачей, требующей значительных вычислительных ресурсов и лингвистической экспертизы (Kraus, Feuerriegel, 2019). Кроме того, качество анализа тональности напрямую зависит от точности построения этих деревьев. Таким образом, работа подчеркивает потенциал использования RST в анализе тональности, но также указывает на существующие ограничения и направления для дальнейших исследований. Более того, подобные инструменты позволяют систематизировать процесс исследований, документируя каждый шаг и его обоснование. Это облегчает воспроизводимость результатов, обмен опытом между исследователями и ускоряет процесс обучения новых специалистов. В перспективе, такие алгоритмы и методы могут быть интегрированы в автоматизированные системы разработки лингвистических моделей, позволяя значительно сократить время и ресурсы, затрачиваемые

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

на создание высококачественных и эффективных инструментов обработки естественного языка.

Однако исследователи также указывают на ограничения, связанные с контекстуальной зависимостью языка и сложностью интерпретации результатов. Автоматический анализ может упускать нюансы, которые важны для понимания индивидуальных особенностей личности, в частности скрытых мотивов. Поэтому, несмотря на возможности автоматизированных методов, участие квалифицированного специалиста в интерпретации результатов остается ключевым элементом. Компьютерная лингвистика и обработка естественного языка (NLP) предлагают мощные инструменты для анализа текстов, потенциально раскрывая закономерности и характеристики, связанные с психологическими особенностями личности. Графовые методы, применяемые в NLP, позволяют визуализировать и анализировать структуру текста, выявляя ключевые темы, отношения между понятиями и паттерны коммуникации. Эти методы особенно полезны для анализа больших объемов текста, позволяя выявлять доминирующие нарративы и скрытые связи. Однако, необходимо учитывать ограничения. Автоматический анализ опирается на алгоритмы, которые, несмотря на свою сложность, не всегда способны уловить нюансы человеческого языка, такие как ирония, сарказм или метафоры. Кроме того, интерпретация результатов требует глубокого понимания как лингвистических, так и психологических принципов².

Компьютерная лингвистика позволяет автоматизировать процессы контент-анализа, тональности текста и семантического поиска, что особенно ценно при работе с масштабными опросами общественного мнения, как демонстрируют данные ВЦИОМ и РАНХиГС³. Одним из ключевых преимуществ является возможность выявления паттернов в языковых данных, которые могут указывать на определенные психологические характеристики личности или

² Тестирование лингвистических технологий: соревнования по автоматическому разрешению кореферентности и анафоры. АБВУ [Электронный ресурс]. URL: <https://habr.com/ru/company/abbyy/blog/229515/>

AI Powered Text Analysis Apis. ParallelDots AI APIs [Электронный ресурс]. URL: https://apis.paralldots.com/text_docs/index.html

NTA Графовые методы анализа в Gephi / NTA. vc.ru : [Электронный ресурс]. URL: <https://vc.ru/ml/216547-grafovye-metody-analiza-v-gephi>

³ WCIOM. SPUTNIK Daily All-Russian Poll. WCIOM News, 2022. URL: <https://ok.wciom.ru/research/wciom-sputnik>

RANEPА (2020) Eurobarometer Methodology. Russian Presidential Academy of National Economy and Public Administration. URL: <https://www.ranepa.ru/nauka-i-konsalting/strategii-i-doklady/evrobarometr/metodologiya-evrobarometra/> (дата обращения: 27.03.2025).

Unsupervised Cross-lingual Representation Learning at Scale / A. Conneau [et al.]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2020, 8440-8451. <https://doi.org/10.18653/v1/2020.acl-main.747>

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

группы. Например, анализ частоты употребления определенных слов, стилистических особенностей или грамматических конструкций может быть использован для оценки уровня тревожности, уверенности или социальной направленности. Однако необходимо учитывать ограничения, связанные с интерпретацией результатов. Компьютерные методы не всегда способны улавливать контекст и нюансы человеческого общения, что может приводить к искажениям в анализе. В психологии личности требуется комплексный подход, сочетающий автоматизированный анализ данных с экспертной оценкой и учетом социокультурного контекста (Лукияненко, 2025).

В Федеральном государственном научном бюджетном учреждении «Институт проблем искусственного интеллекта» (ФГНБУ ИПИИ, ДНР, г. Донецк) группа исследователей: Дорохина Г.В., проф. Шелепов В.Ю., Ниценко А.В., Пикалев Я.С. и др. активно занимаются изысканиями в сфере компьютерной лингвистики и обработки естественного языка (Natural Language Processing, NLP). Данные исследования ориентированы на разработку решений для комплексных задач, в том числе затрагивающих аспекты психологии личности ведут интенсивные исследования в области компьютерной лингвистики и обработки естественного языка, направленные на решение сложных задач, в том числе и в психологии личности. Эти разработки открывают новые перспективы, в которой одной из ключевых областей является разработка методов автоматического анализа текстов для выявления эмоциональной окраски, психологических характеристик и скрытых намерений авторов. Эти методы базируются на сложных алгоритмах машинного обучения, способных анализировать синтаксические структуры, семантические связи и прагматические особенности языка. В рамках этих исследований особое внимание уделяется разработке и совершенствованию методов автоматической обработки речи, включая распознавание, синтез и анализ. Акцент делается на создании эффективных алгоритмов и моделей, способных адаптироваться к различным акцентам (Бурибаева, 2013), (Дорохина, 2024), (Ниценко, Шелепов, Большакова, 2021), Пикалев, Ермоленко, 2023), (Пигуз, Изосимова, Ивашко, 2022). Применяются современные подходы машинного обучения, нейронные сети и методы статистического моделирования. Одним из ключевых направлений является разработка систем, способных понимать запросы пользователей на естественном языке и предоставлять точные и релевантные ответы. Эти системы находят применение в различных областях, таких как образование, здравоохранение, обслуживание клиентов и информационный поиск. Особое внимание уделяется интеграции лингвистических знаний и семантического анализа для улучшения понимания смысла запросов. В области психологии личности исследования направлены на разработку методов автоматического анализа текстов и речи для выявления психологических характеристик и эмоционального состояния индивида. Используются методы машинного обучения и анализа тональности для определения настроения, уровня стресса и других психологических параметров (Ивашко, Изосимова, Пигуз, 2022). Эти методы могут быть использованы для разработки персонализированных рекомендаций и поддержки в области психического здоровья. Продолжая исследования в области автоматического анализа текстов и речи, ученые ФГНБУ «Института проблем искусственного интеллекта» активно изучают возможности для использования систем автоматической диагностики и прогнозирования психологического состояния индивида на основе анализа текстов и речи. Эти системы могут быть использованы для раннего выявления депрессии, тревожности и других психических

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

расстройств. Разрабатываются алгоритмы, способные учитывать индивидуальные особенности и контекстуальные факторы, что повышает точность и надежность диагностики. Во взаимодействии науки и технологий ученые ФГНБУ «ИПИИ» видят не только возможность усовершенствовать методы диагностики и прогнозирования, но и стремятся расширить границы понимания самой человеческой природы. Изучение взаимосвязи между языком, речью и эмоциональным состоянием открывает новые перспективы для углубленного исследования когнитивных процессов и механизмов, лежащих в основе нашего мышления и поведения (Ивашко, Изосимова, Пигуз, 2025).

Перспективы, практическое применение

Фундаментальная задача исследований в области автоматического анализа текстов заключается в разработке систем, способных к глубокому пониманию естественного языка и предоставлению релевантных ответов. В данном контексте интеграция лингвистических знаний и методов семантического анализа существенно повышает точность интерпретации пользовательских запросов, обеспечивая более эффективное взаимодействие человека с машиной. В области психологии личности методы машинного обучения и анализа тональности находят применение для автоматического выявления психологических характеристик и эмоционального состояния индивида, включая оценку уровня стресса и преобладающего настроения. Полученные результаты открывают перспективы для разработки персонализированных рекомендаций в области психического здоровья, что способствует более эффективной поддержке и улучшению благополучия личности.

В ФГБНУ «Институт проблем искусственного интеллекта» проводятся передовые исследования в области автоматической диагностики и прогнозирования психологического состояния на основе анализа текстовых и речевых данных. Данный подход позволяет выявлять признаки депрессии и тревожности на ранних стадиях, что имеет важное значение для своевременного оказания помощи. Для повышения точности диагностики учитываются индивидуальные и контекстуальные факторы, влияющие на психическое состояние. Изучение сложной взаимосвязи между языком, речью и эмоциональным состоянием открывает новые перспективы для углубленного исследования когнитивных процессов, лежащих в основе человеческого мышления и поведения. Дальнейшее развитие этих технологий предполагает создание интерактивных систем, способных не только анализировать, но и генерировать тексты и речь, адаптированные к конкретным психологическим профилям и эмоциональным потребностям пользователей. Это открывает возможности для разработки инновационных инструментов в области психотерапии. Важным аспектом является разработка методов защиты от манипуляций и дезинформации, основанных на анализе лингвистических особенностей текстов и речи. Эти методы позволяют выявлять скрытые намерения и предвзятость в сообщениях, что способствует повышению критического мышления и осознанности у пользователей. В целом, исследования, проводимые в ФГНБУ «ИПИИ», открывают новые возможности для использования компьютерной лингвистики и обработки естественного языка в психологии личности. Эти разработки могут помочь в решении широкого круга задач, от ранней диагностики психических заболеваний до выявления лжи и обмана. Необходим баланс между возможностями технологий и ответственностью за их применение.

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

Заключение

Для русского языка особенно важна адаптация существующих методов компьютерной лингвистики, учитывая его морфологическую сложность и синтаксическую вариативность. Разработка специализированных словарей и алгоритмов, учитывающих специфику русского языка, позволит повысить точность и надежность анализа текстовых данных. При разработке исследований важно учитывать этические аспекты, связанные с использованием текстовых данных, особенно если они содержат личную информацию. Необходимо обеспечить конфиденциальность и анонимность участников исследования, а также получить их согласие на использование их текстов. В целом, компьютерная лингвистика предоставляет мощный инструмент для исследования личности, но требует осторожного и осознанного применения, с учетом возможностей и ограничений, а также этических аспектов.

В заключение, необходимо подчеркнуть, что для успешного применения методов компьютерной лингвистики в психологии личности необходимо учитывать как возможности, так и ограничения этих методов, а также разрабатывать методологии, обеспечивающие валидность и надежность результатов.

Список источников / References

1. Бах, Э. (2009). Неформальные лекции по формальной семантике. М.: Книжный дом «Либроком».
Bach, E. (2009). Informal lectures on formal semantics. Moscow: Librocom Book House (In Russ.).
2. Бурибаева, А.К., Дорохина, Г.В., Ниценко, А.В., Шелепов, В.Ю. (2013). Сегментация и дифонное распознавание речевых сигналов, Тр. СПИИРАН, 31, 20-42.
Buribaeva, A.K., Dorokhina, G.V., Nitsenko, A.V., Shelepov, V.Y. (2013). Segmentation and diphone recognition of speech signals, Proc. SPIIRAN, 31, 20-42 (In Russ.).
3. Ван, Д., Вайс, Дж.Д., Мюллер, М., Рам, П., Гейер, П., Дуган, К., Таучик, Ю.Р., Самуловиц, Х., Грей, А.Г. (2019). Сотрудничество человека и ИИ в науке о данных. Материалы конференции ACM по взаимодействию человека с компьютером, 3 (CSCW), 1-24. <https://doi.org/10.1145/3359313>
Wang, D., Weiss, J.D., Mueller, M., Ram, P., Geyer, P., Dugan, K., Tauchik, J.R., Samulowitz, H., Gray, A.G. (2019). Human-AI collaboration in data science. Proceedings of the ACM conference on Human-Computer Interaction, 3 (CSCW), 1-24. <https://doi.org/10.1145/3359313> (In Russ.).
4. Витко, Ю.С., Калугин, А.Ю., Щебетенко, С.А. (2022). Многообразие структур черт личности. Социальные и гуманитарные науки: теория и практика.
Vitko, Y.S., Kalugin, A.Y., Shchebetenko, S.A. (2022). Variety of structures of personality traits. Social and Humanities: Theory and Practice (In Russ.).
5. Вудвортс, Р. (1950). Экспериментальная психология. Издательство иностранной литературы.

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В. (2025)
Компьютерная лингвистика в психологии: ключ к пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N., Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to understanding language and human behavior
Language and Text, 12(2), 140—162.

Woodworths, R. (1950). *Experimental psychology*. Foreign Literature Publishing House (In Russ.).

6. Дорохина, Г.В., Полоус, А.Г. (2024). Методы организации вопросно-ответных систем. *Известия Кабардино-Балкарского научного центра РАН*, 26 (6), 175-187.

Dorokhina, G.V., Polous, A.G. (2024). Methods of organization of question-answer systems. *Proceedings of Kabardino-Balkarian Center of RAS*, 26 (6), 175-187 (In Russ.).

7. Дорохина, Г.В. (2024). Пофонемное распознавание как задача классификации рядов на множестве последовательностей элементов сложных объектов с применением усовершенствованного trie-дерева. *Информатика и автоматизация*, 23 (6), 1784-1822.

Dorokhina, G.V. (2024). Phoneme recognition as a problem of row classification on the set of sequences of complex object elements using an improved trie-tree. *Informatics and Automation*, 23 (6), 1784-1822 (In Russ.).

8. Ениколопов, С.Н., Медведева, Т.И., Воронцова, О.Ю., Чудова, Н.В., Кузнецова, Ю.М., Пенкина, М.Ю., Минин, А.Н., Станкевич, М.А., Смирнов, И.В., Любавская, А.А. (2018). Лингвистические характеристики текстов психически больных и здоровых людей. *Психологические исследования*, 11 (61), 1.

Enikolopov, S.N., Medvedeva, T.I., Vorontsova, O.Y., Chudova, N.V., Kuznetsova, Y.M., Penkina, M.Y., Minin, A.N., Stankevich, M.A., Smirnov, I.V., Lubavskaya, A.A. (2018). Linguistic characteristics of texts of mentally ill and healthy people. *Psychological Research*, 11 (61), 1 (In Russ.).

9. Ивашко, К.С., Изосимова, С.А., Пигуз, В.Н., Ковалева, К.В. (2025). Теоретические аспекты диагностики и саморегуляции психоэмоциональных состояний личности. *Проблемы искусственного интеллекта*, 1 (36), 58-67.

Ivashko, K.S., Izosimova, S.A., Piguz, V.N., Kovaleva, K.V. (2025). Theoretical aspects of diagnostics and self-regulation of psychoemotional conditions. *Problems of Artificial Intelligence*, 1 (36), 58-67 (In Russ.).

10. Ивашко, К.С., Изосимова, С.А., Пигуз, В.Н. (2022). Анализ возможностей интеллектуально-духовных средств при использовании их для саморегуляции психоэмоциональных состояний личности. *Искусственный интеллект: теоретические аспекты, практическое применение: Материалы Донецкого международного научного круглого стола* (с. 60-72). Донецк.

Ivashko, K.S., Izosimova, S.A., Piguz, V.N. (2022). Analysis of possibilities of intellectual-spiritual means when using them for self-regulation of psychoemotional states of personality. *Artificial intelligence: theoretical aspects, practical application: Proceedings of Donetsk international scientific round table* (pp. 60-72). Donetsk (In Russ.).

11. Кузьмина, А.А., Лифшиц, М.А., Костенко, В.Ю. (2022). Методы компьютерной лингвистики и обработки естественного языка: возможности и ограничения для задач

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В. (2025)
Компьютерная лингвистика в психологии: ключ к пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N., Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to understanding language and human behavior
Language and Text, 12(2), 140—162.

- психологии личности. Современная зарубежная психология, 11 (1), 104-115. <https://doi.org/10.17759/jmfp.2022110110>
- Kuzmina, A.A., Lifshits, M.A., Kostenko, V.Y. (2022). Methods of computational linguistics and natural language processing: possibilities and limitations for tasks of personality psychology. *Modern foreign psychology*, 11 (1), 104-115. <https://doi.org/10.17759/jmfp.2022110110> (In Russ.).
12. Лукьяненко, Роман (2 января 2025 г.). Генеративный искусственный интеллект: развивающиеся технологии, растущее влияние на общество и возможности для исследований в области информационных систем. Границы информационных систем. <http://dx.doi.org/10.2139/ssrn.5153894>
- Lukyanenko, Roman (January 2, 2025). Generative Artificial Intelligence: emerging technologies, growing societal impact, and opportunities for information systems research. *Frontiers in Information Systems*. <http://dx.doi.org/10.2139/ssrn.5153894> (In Russ.).
13. Ниценко, А.В., Шелепов, В.Ю., Большакова, С.А. (2021). О снятии омонимии словосочетаний, которые могут быть предикативами. Проблемы искусственного интеллекта, 1 (20), 53-63.
- Nitsenko, A.V., Sheleпов, V.Yu., Bolshakova, S.A. (2021). On the removal of homonymy of word combinations that can be predicatives. *Problems of Artificial Intelligence*, 1 (20), 53-63 (In Russ.).
14. Олпорт, Г. (1998). Личность в психологии. СПб.: Ювента.
Allport, G. (1998). *Personality in Psychology*. Saint Petersburg: Juventa (In Russ.).
15. Пигуз, В.Н., Изосимова, С.А., Ивашко, К.С. (2022). Человек и искусственный интеллект: связь нервной системы человека с его психоэмоциональным состоянием. Электронный научно-публицистический журнал «Номо Cyberus», 1(12).
- Piguz, V.N., Izosimova, S.A., Ivashko, K.S. (2022). Man and artificial intelligence: the connection of the human nervous system with its psychoemotional state. *Electronic scientific-publicistic journal «Nomo Cyberus»*, 1(12) (In Russ.).
16. Пикалев, Я.С., Ермоленко, Т.В. (2023). О нейронных архитектурах извлечения признаков для задачи распознавания объектов на устройствах с ограниченной вычислительной мощностью. Проблемы искусственного интеллекта, 3 (30), 44-54. <https://doi.org/10.34757/2413-7383.2023.30.3.004>
- Pikalev, Ya.S., Ermolenko, T.V. (2023). About neural architectures of feature extraction for the task of object recognition on the devices with limited computing power. *Problems of Artificial Intelligence*, 3 (30), 44-54. <https://doi.org/10.34757/2413-7383.2023.30.3.004> (In Russ.).
17. Плунгян, В.А. (2005). Зачем нужен Национальный корпус русского языка? Неформальное введение [Электронный ресурс]. Национальный корпус русского языка:

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В. (2025)
Компьютерная лингвистика в психологии: ключ к пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N., Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to understanding language and human behavior
Language and Text, 12(2), 140—162.

- 2003—2005. Результаты и перспективы. М.: Индрик, 6-20. URL: <https://ruscorpora.ru/new/sbornik2005/02plu.pdf> (дата обращения: 10.03.2025)
- Plungian, V.A. (2023). Why do we need the National Corpus of the Russian Language? Informal introduction [Electronic resource]. National Corpus of the Russian Language: 2003-2005. Results and Prospects. Moscow: Indrik, 2005, 6-20. Available at: <https://ruscorpora.ru/new/sbornik2005/02plu.pdf> (viewed: 10.03.2025) (In Russ.).
18. Рэймонд Кеттел: структурная теория черт личности. PSYCHOJOURNAL.RU. Раздел: Психологи. Дата публикации: 13-10-2014. <https://psychojournal.ru/psychologists/95-reymond-kettel-strukturnaya-teoriya-chert-lichnosti.html>
- Raymond Kettel: structural theory of personality traits. PSYCHOJOURNAL.RU. Section: Psychologists. Date of publication: 13-10-2014. <https://psychojournal.ru/psychologists/95-reymond-kettel-strukturnaya-teoriya-chert-lichnosti.html> (In Russ.).
19. Тестирование лингвистических технологий: соревнования по автоматическому разрешению кореферентности и анафоры. АБВУ [Электронный ресурс]. URL: <https://habr.com/ru/company/abbyy/blog/229515/>
- Testing linguistic technologies: competitions in automatic co-referentiality and anaphora resolution. АБВУ [Electronic resource]. Available at: <https://habr.com/ru/company/abbyy/blog/229515/> (In Russ.)
20. Толдова, С.Ю., Ляшевская, О.Н. (2014). Современные проблемы и тенденции компьютерной лингвистики. Вопросы языкознания, 1, 120-145.
- Toldova, S.Y., Lyashevskaya, O.N. (2014). Modern problems and trends in computational linguistics. Questions of Linguistics, 1, 120-145 (In Russ.).
21. Ультес, С., Шмитт, А., Эль Хаб, Р., Минкер, В. (2012). Статистическое моделирование качества взаимодействия в речевых диалоговых системах: сравнение классификаторов, основанных на (условных) скрытых марковских моделях, и машин опорных векторов. Вестник Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева, 115-119.
- Ultes, S., Schmitt, A., El Hab, R., Minker, W. (2012). Statistical modeling of interaction quality in speech dialog systems: a comparison of classifiers based on (conditional) hidden Markov models and support vector machines. Bulletin of the Siberian State Aerospace University named after Academician M.F. Reshetnev, 115-119 (In Russ.)
22. Что Такое «Большая Пятерка» Личностных Качеств? [Электронный ресурс]. URL: <https://www.b17.ru/article/617847/>
- What is the “Big Five” Personality Qualities? [Electronic resource]. Available at: <https://www.b17.ru/article/617847/> (In Russ.)
23. AI Powered Text Analysis Apis. ParallelDots AI APIs [Электронный ресурс]. URL: https://apis.paralldots.com/text_docs/index.html

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В. (2025)
Компьютерная лингвистика в психологии: ключ к пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N., Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to understanding language and human behavior
Language and Text, 12(2), 140—162.

24. Borghi AM, et al. (2018). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Phys Life Rev.* <https://doi.org/10.1016/j.plrev.2018.12.001>
25. WCIOM. SPUTNIK Daily All-Russian Poll. VCIOM News, 2022. URL: <https://ok.wciom.ru/research/vciom-sputnik>
26. Goldberg, L.R. (1981). Language and individual differences: The search for universals in personality lexicons. *Review of personality and social psychology*, 2 (1), 141-165.
27. Kraus, M., Feuerriegel, S. (2019). Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees. *Expert Systems with Applications*, 118, 65-79.
28. Park, G., Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D.J., Ungar, L.H., Seligman, M.E.P. (2015). Automatic personality assessment through social media language (автоматическая оценка личности на основе языка социальных сетей). *Journal of Personality and Social Psychology*, 108 (6), 934-952.
29. Pennebaker, J.W., Booth, R.J., Boyd, R.L., Francis, M.E. (2015). *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates.
30. RANEPА (2020) Eurobarometer Methodology. Russian Presidential Academy of National Economy and Public Administration. URL: <https://www.ranepa.ru/nauka-i-konsalting/strategii-i-doklady/evrobarometr/metodologiya-evrobarometra/> (дата обращения: 27.03.2025).
31. The Narrative Arc: Revealing Core Narrative Structures Through Text Analysis / Ryan L. Boyd, Kate G. Blackburn, James W. Pennebaker. <https://doi.org/10.29025/2079-6021-2022-4-17-34>
32. Unsupervised Cross-lingual Representation Learning at Scale / A. Conneau [et al.]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2020, 8440-8451. <https://doi.org/10.18653/v1/2020.acl-main.747>
33. Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, 44 (3), 363-373.
34. Yarkoni, T., Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12 (6), 1100-1122.

Информация об авторах

Кристина Сергеевна Ивашко, младший научный сотрудник, ФГНБУ «Институт проблем искусственного интеллекта», Донецк, Российская Федерация, ORCID: <https://orcid.org/0000-0002-0195-2279>, e-mail: kristi_8@mail.ru

Снежана Александровна Изосимова, младший научный сотрудник, ФГНБУ «Институт проблем искусственного интеллекта», Донецк, Российская Федерация, ORCID: <https://orcid.org/0000-0002-1521-5579>, e-mail: izosimova.snezhana@mail.ru

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

Валентина Николаевна Пигуз, заведующий отделом компьютерно-информационных технологий, ФГНБУ «Институт проблем искусственного интеллекта», Донецк, Российская Федерация, ORCID: <https://orcid.org/0009-0003-3968-3144>, e-mail: maxpvn77@gmail.com

Ксения Витальевна Ковалева, младший научный сотрудник, ФГНБУ «Институт проблем искусственного интеллекта»; врач-психиатр ГБУ ДНР «РКПБ-ПТЦ», Донецк, Российская Федерация, ORCID: <https://orcid.org/0009-0002-8218-1898>, e-mail: kv_kovaleva@mail.ru

Information about the authors

Kristina S. Ivashko, Junior Researcher, Institute of Artificial Intelligence Problems, Donetsk, Russian Federation, ORCID: <https://orcid.org/0000-0002-0195-2279>, e-mail: kristi_8@mail.ru

Snezhana A. Izosimova, Junior Researcher, Institute of Artificial Intelligence Problems, Donetsk, Russian Federation, ORCID: <https://orcid.org/0000-0002-1521-5579>, e-mail: izosimova.snezhana@mail.ru

Valentina N. Piguz, Head of the Department of Computer-Information Technologies, Institute of Artificial Intelligence Problems, Donetsk, Russian Federation, ORCID: <https://orcid.org/0009-0003-3968-3144>, e-mail: maxpvn77@gmail.com

Ksenia V. Kovaleva, Junior Researcher, Institute of Artificial Intelligence Problems; Psychiatrist, SBI DPR RCPH-PTC, Donetsk, Russian Federation, ORCID: <https://orcid.org/0009-0002-8218-1898>, e-mail: kv_kovaleva@mail.ru

Вклад авторов

Пигуз В.Н., Ковалева К.В. — идеи исследования; аннотирование, написание и оформление рукописи; планирование исследования; контроль за проведением исследования.

Ивашко К.С., Изосимова С.А. — сбор и анализ данных; визуализация результатов исследования.

Все авторы приняли участие в обсуждении результатов и согласовали окончательный текст рукописи.

Contribution of the authors

Valentina N. Piguz, Ksenia V. Kovaleva — ideas; annotation, writing and design of the manuscript; planning of the research; control over the research.

Kristina S. Ivashko, Snezhana A. Izosimova — conducting the experiment; data collection and analysis; visualization of research results.

All authors participated in the discussion of the results and approved the final text of the manuscript.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Ивашко К.С., Изосимова С.А., Пигуз В.Н., Ковалева К.В.
(2025)
Компьютерная лингвистика в психологии: ключ к
пониманию языка и поведения человека
Язык и текст, 12(2), 140—162.

Ivashko K.S., Izosimova S.A., Piguz V.N.,
Kovaleva K.V. (2025)
Computational linguistics in psychology: a key to
understanding language and human behavior
Language and Text, 12(2), 140—162.

Conflict of interest

The authors declare no conflict of interest.

Поступила в редакцию 15.05.2025
Поступила после рецензирования 28.05.2025
Принята к публикации 10.06.2025
Опубликована 23.06.2025

Received 2025.05.15
Revised 2025.05.28
Accepted 2025.06.10
Published 2025.06.23