



РУССКИЙ ЯЗЫК В ИНТЕРНЕТ-ПРОСТРАНСТВЕ

RUSSIAN ON THE INTERNET

DOI: 10.22363/2618-8163-2026-24-1-25-40

EDN: XARBFO

Научная статья

Потенциал нейросетевого моделирования семантического поля «Интернет» в русскоязычном дискурсе

А.А. Баркович  , Е.С. Астапкина 

Белорусский государственный университет, Минск, Республика Беларусь

 barkovichaa@gmail.com

Аннотация. Исследование посвящено лингвистическому анализу возможностей нейросетевого моделирования семантического поля «Интернет» на материале доступного в онлайн-формате русскоязычного контента. Актуальность исследования обеспечена качеством и количеством проанализированного в формате «больших данных» языкового материала и инновационным методологическим подходом к его метаописанию с применением нейросетевого инструментария. Цель исследования — лингвистическая характеристика потенциала нейросетевого моделирования семантического поля «Интернет» русскоязычного дискурса. Материалом послужил русскоязычный интернет-контент, объем которого для получения статистически репрезентативных метаданных не был искусственно регламентирован. Такой подход наилучшим образом соотносится с реалиями преимущественно декларативных ограничений функциональности интернет-дискурса. Ориентация на «интеллектуальность» алгоритмов обработки контента Интернета такими ресурсами, как базовый для нашего исследования проект OpenAI, обеспечила высокую референтность языковых данных. В исследовании задействован широкий спектр методов: от компонентного анализа до дискурс-анализа с использованием современного нейросетевого инструментария. Осуществлено двухаспектное нейросетевое моделирование посредством процедур кластерного и стратного анализа языковых единиц, ассоциирующихся с понятийным феноменом Интернет. В результате выявлен и описан потенциал применения методики нейросетевого моделирования для исследования семантического поля «Интернет». Проведенное моделирование способствовало выявлению и верификации широкого круга языковых единиц, речевая функциональность которых обусловлена ассоциированностью с понятийным феноменом Интернет, являющимся ядром соответствующего семантического поля. Полученные результаты перспективны, так как позволяют уверенно имплементировать апробированные в данном исследовании шаблоны нейросетевого моделирования в лингвистическую практику, что, в свою очередь, будет способствовать развитию парадигмы языкознания, модернизации методоло-

© Баркович А.А., Астапкина Е.С., 2026

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

гических подходов к изучению языкового функционирования, идентификации и квалификации речевых новаций.

Ключевые слова: нейросетевой инструментарий, кластерная модель, стратная модель, понятийный феномен, лингвистическая репрезентация, потенциал моделирования

Вклад авторов: авторы сделали эквивалентный вклад в разработку концепции исследования, написание и редактирование текста статьи.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

История статьи: поступила в редакцию 20.07.2025; принята к печати 18.10.2025.

Для цитирования: *Баркович А.А., Астапкина Е.С.* Потенциал нейросетевого моделирования семантического поля «Интернет» в русскоязычном дискурсе // *Русистика*. 2026. Т. 24. № 1. С. 25–40. <http://doi.org/10.22363/2618-8163-2026-24-1-25-40> EDN: XARBFO

Введение

Интернет как феномен является постоянно совершенствуемой технологической инновацией, глобально-значимой реалией и, естественно, — доминирующим в речевой практике понятием. При этом, по мере развития самого Интернета, связанное с ним семантическое поле неуклонно расширяется, вбирая в себя новые фрагменты, нередко дополняя или изменяя семантику уже присутствующих в языке единиц. Значимость «Интернет» выступает в роли «общего лингвистического интегрального признака», реализующего «способность притягивать к себе новые элементы, обладающие таким признаком» (Щур, 1974: 106). Семантическое поле «Интернет» — один из важнейших феноменологических фокусов коммуникации, а его изучение позволяет оценить реальную динамику современного языка, поскольку с момента появления в 1979 г. лексемы *Интернет* (изначально как *Internet*) прошло не так много времени и огромная масса интернет-референтной речевой практики создана, сохранена и доступна для изучения.

Так или иначе, динамичная экспансия информационно-компьютерных технологий, ориентированных на Интернет, затронула не только формат коммуникации, но и повлияла на ее содержание, семантику (Дубина, 2013; Кондрашева, Тинина, 2019; Sapone, 2021). Язык интернет-дискурса меняется и это особенно хорошо заметно в контексте высоких темпов развития социокультурной и производственной сфер в XXI в. Интенсивное развитие информационных технологий повлекло многоаспектное расширение коммуникационного инструментария (Aichner, Jacob, 2015). Базовое для Интернета компьютерное опосредование коммуникации стало основным форматом информационного обмена (Barkovich, 2019).

Характерно, что многие из задействованных в интернет-обусловленной коммуникации языковых средств изначально были сугубо терминологичны, однако высочайшая востребованность компьютерно-информационных технологий обеспечила массовый транзит соответствующих номинаций в узус, что подтверждается фиксацией интернет-лексики в толковых словарях. В научном контексте как никогда высокий уровень требований к объективности

и репрезентативности метаописаний языка предполагает максимально широкий охват речевой практики, ее статистически обоснованное описание и непротиворечивую интерпретацию: «Естественный язык — сложный объект для обработки компьютерами. Оставив в стороне философские споры, можно сказать, что в области обработки естественного языка наблюдается сдвиг парадигмы от методов, основанных на правилах, к статистическим подходам, которые доминируют с 1990-х годов» (Goldberg, 2017: 193). Конечно, современная лингвистика учитывает новые вызовы и тренды коммуникации, и это существенно приблизило горизонты языкознания, которые сегодня в интересах высокотехнологичной коммуникации необходимо описывать в режиме «реального времени».

В условиях стремительного развития цифровых технологий и повсеместного распространения интернет-коммуникации лингвистические исследования все чаще посвящаются анализу содержательной базы коммуникации (Дмитриева, Шаклеин, 2022; Разумовская, 2024). Многим традиционным методам, таким как компонентный и лексикографический анализ, при всей их актуальности, присущи свои ограничения при работе с экстенсивными и слабо структурированными массивами данных, так называемыми «большими данными» (англ. *Big Data*). Это обуславливает востребованность инновационных, в т.ч. нейросетевых исследовательских средств, способных мониторить динамику и многообразие значимости, функционирующей в онлайн-дискурсе (Азимов и др., 2023; Беляков, Максименко, 2024; Попова, Колесова, 2023). Так, архитектура трансформеров, например, *GPT-4*, в отличие от простых статистических моделей, не только фиксирует частотность слов, но и позволяет отражать контекстуальные связи и семантическую близость понятий. Это продуктивно и перспективно для моделирования понятийно-концептуальной структуры речевой практики. В частности, методы кластерной и стратной дистрибуции, реализованные на базе нейросетей, дают возможность выявить зачастую латентные и имплицитные для других исследовательских инструментов оттенки семантики. И, действительно, полученные метаданные убедительно свидетельствуют, что в коммуникации семантика не сводится к реализации ее словарного клише, а представляет собой сложную систему, которая постоянно изменяется. Нейросети позволяют объективно, на основе реального речевого материала, выявлять ее динамику и структуру.

Исследования в этой области имеют практическое значение. Апробированные и верифицированные модели в дальнейших изысканиях могут служить релевантной матрицей идентификации и квалификации речевых новаций (неологизмов, сленга) и понимания их роли в современной коммуникации. Например, такие модели могут помочь в автоматическом обнаружении и квалификации интернет-жаргонизмов или совершенствовании поисковых систем, учитывающих не только лексическую, но и понятийную семантику запросов. Они показывают, как семантика языковых единиц трансформируется в повседневной, динамичной речевой практике. Это позволяет оптимизировать исследовательские каноны, что крайне важно для ориентированной на многоаспектное сопровождение современной коммуникации лингвистики.

Цель исследования — лингвистическая характеристика потенциала нейросетевого моделирования семантического поля «Интернет» русскоязычного дискурса.

Методы и материалы

Материалом для решения поставленных задач послужил многообразный и уже чрезвычайно насыщенный речевой продукцией русскоязычный интернет-контент, что позволило обеспечить репрезентативное рассмотрение интернет-семантики, объективно в нем присутствующей, и уверенно фокусироваться на аспектах ее реализации в речевой практике. Объем задействованного в исследовании материала не был ограничен количественно, поскольку доступ широкого уже на сегодняшний день круга программ с элементами искусственного интеллекта (ИИ) к ресурсам Интернета фактически не имеет препятствий, обеспечивая высочайшую скорость и референтность обработки языковых данных (Blei, Ng, Jordan, 2003; Devlin et al., 2019; Mikolov et al., 2013). Собственно, одним из наиболее технически компетентных и авторитетных ресурсов такого рода является проект *OpenAI*, благодаря нейросетевым технологиям которого получены в целом непротиворечивые и репрезентативные метаданные для лингвистического анализа (более подробно описанные ниже)¹.

Методика, задействованная в данном исследовании, является комплексной и технологически обеспеченной. Она базируется на апробированных и результативных *аналитических* наборах инструментария: *дискурс-анализе*, практически безальтернативном для изучения больших объемов языкового материала, и *компонентном анализе*, позволяющем детализировать структуру языковых единиц и пути их модификации. В дискурсивной парадигме гармонично сосуществуют структуралистские подходы к описанию языковой системы с прикладными, ориентированными на коммуникацию методическими разработками, посвященными речевой практике (Пешё, 1999; Titscher et al., 2000; Widdowson, 2005). Дискурсивная методология «... органично совмещает преимущества когнитивно-семиологического и антропоцентрического подходов, что позволяет проводить значимые метаязыковые обобщения, востребованные в контексте совершенствования современной коммуникации, формализации естественных и создания искусственных языков» (Баркович, 2015: 37). Для решения задач исследования основные аналитические парадигмы были органично подкреплены аргументированным и целесообразным задействованием статистического, сравнительного и семантического анализа. Прагматизм в формировании исследовательской стратегии обосновывается многими специалистами (Виноградова, 2012; Тошович, 2015; Schiffrin, 2006). Полностью прагматически-оправданным и продуктивным в данной связи оказалось кластерное и стратное *моделирование* интернет-контента на базе нейросетевых технологий. Для этого были использованы возможности его глобального охвата посредством ресурса, включающего технологии так называемых «генеративных трансформеров» *GPT4*.

¹ OpenAI. URL: <https://openai.com/index/gpt-4/> (Accessed: 05.05.2025).

Результаты

Посредством алгоритмов на базе нейросетевых технологий создано детализированное лингвистическое представление о семантическом поле «Интернет». В частности, **кластерный** «скрининг» речевой практики выявил широкий круг задействованной в отражении интернет-проблематики семантики: нейросеть в 9 кластерах («связь», «технологии», «контент», «пользователи», «экономика», «культура», «проблемы», «будущее», «дополнительные слова») выявила 88 семантически релевантных понятийному феномену *Интернет* единиц. Выполненное **стратное** моделирование позволило отнести к ядерным элементам семантического поля «Интернет» следующие понятия: *связь* (коммуникация, общение, информация, данные); *сеть* (глобальная сеть, мировая паутина, инфраструктура); *технологии* (компьютеры, серверы, протоколы, сети, Wi-Fi, кабели) и *контент* (веб-сайты, видео, музыка, текст, изображения, приложения). Общее количество выявленных понятий составило 71 единицу.

Для перечней лексической реализации понятийной семантики на основе кластерной и стратной моделей выявлена их многоэлементная пересекающаяся **область**: *сеть, коммуникация, информация, данные, письмо, связь, компьютер, сервер, Wi-Fi, кабель, провайдер, IP-адрес, браузер, приложение, протокол, технология, видео, музыка, блог, социальная сеть, форум, игра, контент, хакер, киберспорт, онлайн-игра, искусственный интеллект, блокчейн, метавселенная, интернет вещей, глобальная сеть*. Тем самым по результатам кластерного и стратного моделирования определена представительная общая совокупность семантически близких лексем — их общий фонд составил 31 единицу из общего количества 154, или 20,13 %. Не менее ценен и перечень из идентифицированных 123 уникальных понятий, не имеющих пересечений в вышеописанных параллельных списках, — такой реестр также важен для создания развернутых лингвистических описаний.

В сопоставительном аспекте в исследовании оценены данные, доступные посредством традиционных лингвистических методов, а также данные репрезентации иной понятийной совокупности («Лес»). Благодаря задействованию нейросетевых технологий, рассмотрен широкий контекст взаимодействия лексемы Интернет с языковыми единицами, не связанными с ней напрямую (грамматически). В целом, выявлен и систематизирован значительный потенциал лингвистической репрезентации функциональности составляющих семантическое поле «Интернет» языковых единиц и структуры их семантически организованной совокупности.

Обсуждение

Нейросетевое моделирование способствует высоко результативной обработке больших массивов речевых данных, и, что не менее важно, данная обработка позволяет на основе их квантитативной агрегации формировать научно-значимые качественные обобщения. При этом нужно отметить, что,

подобно лингвистическим корпусам, нейросетевые инструменты не демонстрируют исследуемые тексты, но эффективно их обрабатывают. В частности, благодаря нейросетевому моделированию, могут быть получены убедительные данные о семантически-зависимых совокупностях языковых единиц, формирующих в конечном итоге единое семантическое поле языка. Современный синкретичный подход к интерпретации языковой значимости подтверждает важность учета всех доступных для анализа и опосредующих ту или иную значимость языковых единиц: «... из любого семантического поля, через более или менее длинную цепочку посредствующих звеньев, можно попасть в любое другое поле, так что семантическое пространство языка оказывается в этом смысле непрерывным» (Апресян, 1995: 252). В данной связи основанный на нейросетевом моделировании опыт характеристики семантического поля «Интернет» весьма информативен и существенно дополняет лингвистические представления о сущности и специфике интернет-дискурса.

Семантика — наиболее сложный для изучения компонент знаний о языке. Следствием этой сложности является и по-прежнему «открытый» метаязыковой статус семантического поля как средоточия обобщенных представлений о семантике: «... за термином „семантическое поле“ скрываются десятки различных его пониманий...» (Уфимцева, 1968: 263). С учетом логики нашего исследования, пожалуй, наиболее целесообразно определить *семантическое поле* как совокупность понятийно взаимосвязанных языковых единиц. В свою очередь, значимость той или иной языковой единицы — при ее задействовании в коммуникации — оказывается во многом условной и зависимой от достаточно широкого контекста, системный учет которого посредством компьютерных программ постепенно становится все более реалистичным. Находящиеся на острие современного научного прогресса «... нейронные сети — это фундаментальное компьютерное средство для обработки языка, хотя и очень старое» (Jurafsky, Martin, 2024: 119). Однако со времен «нейрона» МакКаллоха — Питтса компьютеры «научились» многому (McCulloch, Pitts, 1943).

Характерно, что использование современных технологий в лингвистике пока, как ни удивительно, заметно отстает от темпов задействования лингвистических «технологий» в сфере *IT*. Если после неудачных попыток создать в середине XX в. действительно автоматический «машинный перевод» математики не стремились выпячивать языковую составляющую своей работы, то в последнее время ситуация кардинально изменилась: технические специалисты, работающие в информационной сфере, все охотнее называют себя «компьютерными лингвистами». Объясняется данный тренд выдающимися успехами последних лет на пути создания ИИ, достигнутыми за счет полностью лингвистически-зависимых технологий машинного обучения больших языковых моделей (*LLM*) и нейросетевого моделирования. Не углубляясь в специфику данных инноваций, можно констатировать, что и для лингвистических исследований они оказались чрезвычайно полезными.

В контексте исследования ретроспективно обоснован интерес к фиксации элементов семантического поля «Интернета» в разного рода *словарях*.

Однако ввиду весьма скромной представленности связанных с понятийным феноменом *Интернет* языковых единиц в тезаурусах, моделирование на их базе семантического поля «Интернет» вряд ли целесообразно. Это реалии не только русскоязычной лингвистики: *Интернет* мало освоен, что показательно, и в англоязычных тезаурусах. При этом, выбор лексикографических ресурсов оказывается достаточно ограниченным в принципе: многие толковые словари, в частности, русского языка, по понятным причинам (они не обновляются) не содержат и самой лексемы *Интернет*². Впрочем, *Интернет* и 17 производных от него единиц уже зафиксированы в «Современном толковом словаре русского языка» Т.Ф. Ефремовой: *Интернет*; *интернет-издание*; *интернет-кафе* и др.³ Презентация понятий, связанных с Интернетом, намного продуктивнее оказалась в регулирующей речевую практику русского языка ресурсе «Русский орфографический словарь» под редакцией В.В. Лопатина и О.Е. Ивановой — здесь на 2024 г. зафиксировано 194 деривата от производящей основы «Интернет»⁴. Абсолютное большинство из них оказалось композитами: *интернет-автор*; *интернет-агентство* и др.⁵ Конечно, такого рода данные лингвистически ценны, однако, очевидно, они представляют, скорее, информацию о развитии самой лексемы «Интернет», нежели характеризуют всю связанную с ней понятийную совокупность (Астапкина, Баркович, 2025).

Для описания реалий речи достаточно апробированным уже лексикографическим потенциалом обладают *лингвистические корпусы*, или «корпусы текстов», позволяющие выполнить углубленную интерпретацию понятийной семантики того или иного рода. К примеру, на 2014 г. в «Национальном корпусе русского языка» уже присутствовало множество лексики, связанной с понятийным феноменом *Интернет*. Здесь по данным корпус-менеджера в «Основном корпусе» в 2305 «документах» было выявлено 6468 «вхождений» референтной лексики по самому простому запросу «-интернет-». И уже первые сто страниц конкорданса содержали 135 слов с основой *-интернет-*, многие из которых использовались неоднократно (Национальный корпус русского языка⁶). Однако, большинство из этих единиц до сих пор не зафиксировано ни в каких словарях. Попадут ли они в словари, насколько значимы они для узуса — вопросы, не имеющие однозначных ответов. Впрочем, корпусная интерпретация языкового потенциала — самодостаточная тема, предполагающая отдельное рассмотрение. При этом все еще не так значителен аккумулированный в корпусах объем текстов и далеко не все корпусы совместимы с нейро-

² Ожегов С.И., Шведова Н.Ю. Толковый словарь русского языка: 72500 слов и 7500 фразеологических выражений. М. : Азъ, 1994. 907 с. Ушаков Д.Н. Толковый словарь современного русского языка. М. : Альта-Принт : Дом. XXI век, 2008. 510 с. и др.

³ Ефремова Т.Ф. Современный толковый словарь русского языка : в 3 томах. М. : АСТ, 2006. 3312 с.

⁴ Русский орфографический словарь (ок. 200000 слов) / под ред. В.В. Лопатина, О.Е. Ивановой. М. : Аст-пресс книга, 2013. 896 с.

⁵ Там же.

⁶ Национальный корпус русского языка. URL: <http://www.ruscorpora.ru/> (дата доступа: 05.05.2025).

сетевым инструментарием обработки семантики. Уже некогда фантастические миллиарды словоупотреблений корпусов намного уступают объему, доступному для анализа посредством интернет-ориентированных нейросетевых инструментов.

Несомненным достоинством задействия **нейросетевого инструментария** *GPT4* (*Generative Pre-trained Transformer 4*) для анализа речевой практики является его фактически неограниченная масштабируемость — в фокус исследования попадает огромный массив речевых данных, находящихся в открытом доступе. Графическая идентичность русского языка позволяет несколько сузить рамки подобного рода масштабного анализа. При этом интерференция небольшого круга языков, использующих кириллическую графику на фоне доминирования среди них русского языка, несущественно искажает статистические результаты изучения русскоязычной специфики в контексте Интернета⁷.

Наиболее очевидным и, пожалуй, традиционным в данной связи шаблонном репрезентации является **кластерное моделирование**, ориентированное на описание языковых средств как групп и классов некоей общности. Именно такова сущность классической дифференциации языковых единиц на лексико-семантические группы. Сама по себе кластерная модель некоей понятийной общности отражает сегментное структурирование ее семантики, при котором выделяются и агрегируются по совокупностям относительно сопоставимые классы единиц — кластеры понятий. В частности, элементы семантического поля «Интернет» распределены по кластерам следующим образом:

- *сеть, коммуникация, информация, данные, передача, сигнал, телефон, почта, письмо* (кластер «связь»);
- *компьютер, сервер, сеть, Wi-Fi, кабель, оптоволокно, маршрутизатор, провайдер, домен, IP-адрес, браузер, сайт, приложение, протокол* (кластер «технологии»);
- *веб-страница, статья, видео, музыка, фото, новость, блог, социальные сети, форум, чат, игра, файл, поиск* (кластер «контент»);
- *человек, пользователь, юзер, компания, владелец, администратор, разработчик, хакер, бот* (кластер «пользователи»);
- *e-коммерция, онлайн-платежи, реклама, контент-маркетинг, бизнес, фриланс, криптовалюта* (кластер «экономика»);
- *виртуальная реальность, социальные сети, мемы, тренды, инфлюенсеры, киберспорт, онлайн-игры* (кластер «культура»);
- *киберпреступность, фейковые новости, зависимость, конфиденциальность, цензура, отключение* (кластер «проблемы»);
- *искусственный интеллект, блокчейн, метавселенная, 5G, интернет вещей* (кластер «будущее»);

⁷ W3Techs — extensive and reliable web technology surveys. URL: <http://w3techs.com/> (Accessed: 05.05.2025).

– *цифровой мир, глобальная сеть, виртуальное пространство, всемирная паутина, информационные технологии, цифровые технологии, веб, облако, информационная революция, сетевой этикет* (кластер «дополнительные слова»).

Характерно, что лексема *сеть* здесь реализуется в разных кластерах 6 раз — как отдельно, так и с атрибутами (*социальные сети, глобальная сеть, сетевой этикет*). Соответственно, после устранения дублетности в модели остается 83 единицы. Конечно, в лингвистическом контексте единицы данного перечня необходимо привести в форму единственного числа. Еще одной проблемой оказывается достаточно высокая степень идиоматизации большинства присутствующих в перечне словосочетаний. Именно таковы современные реалии речевой практики: возможно, статус лексических единиц в будущем мог бы быть (как в английском языке) распространен на такие идиоматически спаянные словосочетания как *социальная сеть, искусственный интеллект, информационные технологии* и многие им подобные. Они не являются фразеологизмами, но фактически выполняют функцию единой и семантически спаянной номинации.

В целом, кластерная модель семантического поля «Интернет» достаточно полно отразила его смысловую многоаспектность и насыщенность. Благодаря данной модели мы можем анализировать синтезированные ИИ метаданные о квантах (минимально значимых элементах континуума) интернет-семантики, являющихся по сути *концептами*. Являясь ментальным образованием, концепт воплощает представления о каком-либо значимом материальном или абстрактном феномене, представленном в языке. И в этом случае тот факт, что эти представления формируются алгоритмически (посредством нейросети) не уменьшает их ценности как одного из вариантов лингвистической репрезентации объекта. Может ли вышеописанная модель характеризовать отражаемую ей *концептосферу*? В определенной степени — да. При этом степень адекватности и определенности любой концептосферы зависит от ее объективной соотнесенности с представляемым (репрезентируемым) объектом. Такая семантическая соотнесенность формируется не столько предикативными или атрибутивными связями объекта, сколько его номинативными связями, весьма избирательно и зачастую завуалированно присутствующими в словарных определениях. И, конечно, 83 тесно взаимодействующие в речевой практике номинации отражают референтную понятийному феномену *Интернет* семантику более полно и содержательно, чем статья обычного толкового словаря.

Иной возможный метод нейросетевого моделирования семантического поля «Интернет» — по стратам, или сферам влияния его компонентов. **Стратная модель** соответствующей понятийной общности отражает ее уровневое структурирование, при котором выделяются иерархически связанные гипер- и суб-страты ядерного, приядерного и периферийного класса. Данный шаблон отражает один из приоритетов лингвистической репрезентации

структурирования семантического поля: «Семантическое поле — иерархическая структура множества языковых единиц, объединенных общим (инвариантным) значением и отражающих в языке определенную понятийную сферу, лексическая категория высшего порядка, синтезирующая в себе частные категории (синонимию, антонимию и др.)» (Новиков, 2001: 554). Получаемый в результате стратного моделирования материал может быть рассмотрен в сопоставительном аспекте — с привлечением данных кластерного анализа. Подобная репрезентация позволяет расширить фокус анализируемого материала и оценить характер и выраженность идентичности совокупности языковых единиц, отражающих интернет-семантику. Полученные в исследовании данные свидетельствуют, что такой подход к моделированию существенно дополняет результат кластерного моделирования и способствует созданию максимально объективных представлений о структуре и функциональности семантического поля «Интернет».

При этом, логично было бы ожидать, что все семантически связанные с ядерными единицами компоненты семантического поля «Интернет» окажутся в приядерной и периферийной зонах. Это было бы естественно для уже сформированных лексико-семантических систем. Так, при моделировании семантического поля «Лес» ядерное понятие *природа* четко соотносилось с входящими в его семантическую структуру компонентами значимости *животные* и *растения* уже в статусе приядерных и периферийных понятий. В свою очередь, компонентам приядерного понятия *животные* прямо соответствуют эквивалентные периферийные группы «птицы» (*совы, дятлы, певчие птицы*) и «звери» (*лисы, зайцы, волки, медведи*); а компонентам приядерного понятия *растения* — периферийные группы «грибы» (*съедобные, несъедобные, ядовитые*) и «ягоды» (*земляника, черника, брусника*). Таким образом, при стратном моделировании структура ядерного компонента семантического поля «Лес» *природа* прямо спроецировалась на суб-стратные приядерные и далее на периферийные зоны понятийной общности. В семантическом поле «Интернет», как можно видеть, подобные иерархичность и предсказуемость пока не сформировались.

Оказалась ли стратная модель семантического поля «Интернет» четко выраженной онтологией? Отнюдь: по крайней мере, ее иерархичность далека от идеала. Так, только понятие *веб-сайт* из разряда приядерных прямо связано с ядерным понятием *контент* и является компонентом его ядерной значимости. Все остальные приядерные понятия (*социальные сети, электронная почта, поисковые системы, онлайн-игры, электронная коммерция и браузеры*) не имеют как прямых компонентных проекций в гипер-страте семантики ядерного для них понятия, так и воплощенного продолжения среди периферийных суб-страт понятийной общности.

Далее, в моделируемой нейросетью на основании речевой практики приядерной зоне семантического поля «Интернет» лоцируется понятие *веб-сайт*, дифференцируемое на *страницы, блоги, порталы, форумы; социальные сети; электронная почта*, ассоциируемая с *письмами и сообщениями; поисковые системы: Google, Яндекс, Bing; онлайн-игры*, ассоциируемые с *играми и киберспортом; электронная коммерция*, которая связана с *онлайн-магазинами, по-*

купками, платежами; браузеры: *Chrome, Firefox, Safari, Edge*. Соответственно, к понятийной периферии относятся *IP-адреса* ('числовые идентификаторы') *доменные имена* ('адреса сайтов'); *протоколы (HTTP, HTTPS, TCP/IP)*; *кибербезопасность* (в т.ч. *защита данных, вирусы, хакеры*); *интернет вещей*: ('умные устройства'); *цифровой разрыв* ('неравный доступ к интернету'), *цифровая экономика* ('экономика, основанная на интернете'); *Big Data* ('большие данные'); *искусственный интеллект* ('искусственный интеллект в интернете'); *метавселенная* ('виртуальные миры'); *блокчейн* ('технология блокчейн').

Необходимо отметить достаточно большую совокупность именованных сущностей, выявленных при стратном моделировании: *Google, Яндекс, Bing, Chrome, Firefox, Safari, Edge, HTTP, HTTPS, TCP/IP*. Ценность данных о наличии тех или иных именованных сущностей в речевой практике сегодня достаточно высока и востребована.

Статистически доминирующий объем непересекающихся в данных совокупностях единиц (123 из 154) свидетельствует о динамике интернет-семантики и ее нерегламентированности. В целом же, агрегированный объем данных позволил составить аргументированное представление о сущности, специфике семантического поля «Интернет» и перспективности предложенной методики исследования. Ассоциированная с понятийным феноменом *Интернет* семантика языковых единиц продуктивно развивается, уверенно преодолевая изначальные терминологические рамки и формируя многоаспектное собирательное представление о специфике современной коммуникации.

Конечно, нейросетевое моделирование языкового материала не заменяет всех традиционных инструментов лингвистической практики, в т.ч. словарных. Ментальность исследователя-человека по-прежнему востребована для глубокого анализа языка и несомненно превосходит возможности компьютерных программ в систематизации языкового материала и его научном осмыслении. Однако сделанные наблюдения свидетельствуют, что сбор и предобработка материала по заданным параметрам может успешно выполняться нейросетевым инструментарием, что существенно расширяет диапазон и улучшает качество метаописаний.

Заключение

Полученные в процессе кластерного и стратного моделирования речевой практики релевантные совокупности лексических средств в целом объективно отразили структуру и функциональность семантического поля «Интернет». Семантическое поле «Интернет» представлено в исследовании не как статичная, а как динамичная, многоуровневая и эволюционирующая система. В целом, охарактеризованная референтная база интернет-семантики оказалась экстенсивной и разнообразной. Небольшая область пересечения данных, полученных по разным шаблонам моделирования (кластерному и стратному), может быть объяснена подвижностью и разнообразием интернет-семантики, что, в свою очередь, обусловлено высокими темпами развития интернет-технологий и спецификой их освоения социумом.

Таким образом, последовательная и обоснованная имплементация нейросетевого инструментария в лингвистическую практику, несомненно, способствует модернизации лингвистической методологии, обогащению нормативных и структурных подходов когнитивно-дискурсивными и прагматически ориентированными методиками. Полученные благодаря этому результаты могут быть использованы для улучшения поисковых систем и алгоритмов обработки естественного языка. Учет понятийной, а не только лексической семантики необходим для разработки инструментов цифровой лингводиagnostики и мониторинга языковых трендов. Нейросетевой инструментарий позволяет охватить огромные объемы речевой практики в режиме «реального времени», выявить контекстуальные связи и латентные семантические характеристики языковых единиц. Выполненное моделирование семантического поля «Интернет» продемонстрировало свою высокую эффективность, гибкость и лингвистический потенциал. Оно не только дополнило, но и в ряде аспектов превзошло традиционные лингвистические методы, позволив объективно и всесторонне отразить сложную, многомерную и динамичную семантику современной коммуникации. Данный подход актуален как для фундаментальных исследований, так и для решения широкого круга прикладных задач посредством цифровых технологий.

Список литературы

- Азимов Э.Г., Кулибина Н.В., Ван В.* Лингводидактический потенциал социальных сетей в обучении русскому языку как иностранному // *Русистика*. 2023. Т. 21. № 2. С. 133–147. <https://doi.org/10.22363/2618-8163-2023-21-2-133-147> EDN: VAUPZH
- Апресян Ю.Д.* Избранные труды: в 2 т. Т. I.: Лексическая семантика. Синонимические средства языка. М.: Языки русской культуры, 1995. 464 с.
- Астapkина Е.С., Баркович А.А.* Динамика русскоязычного интернет-дискурса: особенности словарной репрезентации лексики // *Русистика*. 2025. Т. 23. № 1. С. 97–116. <https://doi.org/10.22363/2618-8163-2025-23-1-97-116> EDN: VLFONT
- Баркович А.А.* Функциональность диады «коммуникационный – коммуникативный»: дискурсивный аспект // *Вестник Томского государственного университета. Филология*. 2015. № 5 (37). С. 37–52. <https://doi.org/10.17223/19986645/37/3> EDN: VCPSFD
- Беляков М.В., Максименко О.И.* Проблема семантических сдвигов в современном русском языке: корпусное исследование на примере новостного медиадискурса // *Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика*. 2024. Т. 15. № 1. С. 174–194. <https://doi.org/10.22363/2313-2299-2024-15-1-174-194> EDN: EJFXLT
- Виноградова В.Н.* Система русского словообразования в функционально-стилистическом аспекте. М.: Инфотех. 2012. 336 с.
- Дмитриева Н.М., Шаклеин В.М.* Семантические особенности функционирования доминантных вербализаторов этического концепта БЛАГО в средствах массовой информации // *Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика*. 2022. Т. 13. № 4. С. 956–971. <https://doi.org/10.22363/2313-2299-2022-13-4-956-971> EDN: SCBGJS
- Дубина Л.В.* Тенденции развития русского языка в условиях интернет-коммуникации // *Вестник Томского государственного педагогического университета*. 2013. № 2 (130). С. 177–181. EDN: PYABEF

- Кондрашева Е.В., Тинина А.О. Новая лексика языка интернета: способы образования, причины появления // Эпоха науки. 2019. № 18. С. 134–139. <https://doi.org/10.24411/2409-3203-2018-11830> EDN: HBPJAV
- Новиков Л.А. Эскиз семантического поля // Избранные труды. Т. II: Эстетические аспекты языка. Miscellanea. М. : Изд-во РУДН. 2001. С. 554–570.
- Пешё М. Контент-анализ и теория дискурса // Квадратура смысла: Французская школа анализа дискурса. М. : Прогресс, 1999. С. 302–337.
- Попова Т.И., Колесова Д.В. Обучающая модель заданий на взаимопроверку в онлайн-курсе русского языка как иностранного // Русистика. 2023. Т. 21. № 2. С. 181–195. <https://doi.org/10.22363/2618-8163-2023-21-2-181-195> EDN: HBPJAV
- Разумовская В.А. «Сверхтекст» как форма существования «сильного» художественного текста // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2024. Т. 15. № 1. С. 262–275. <https://doi.org/10.22363/2313-2299-2024-15-1-262-275> EDN: FOGHCU
- Тошович Б. Интернет-стилистика : монография. М. : Наука. 2015. 238 с.
- Уфимцева А.А. Слово в лексико-семантической системе языка. М. : Наука, 1968. 287 с.
- Щур Г.С. Теории поля в лингвистике. М. : Наука, 1974. 254 с.
- Aichner T., & Jacob F. Measuring the degree of corporate social media use // International Journal of Market Research. 2015. Vol. 2. No. 57. Pp. 257–275. <https://doi.org/10.2501/IJMR-2015-018>
- Barkovich A. Informational linguistics: computer, internet, artificial intelligence and language // 2019 IEEE 1st International Conference on Artificial Intelligence in Information and Communication (ICAIC 2019), Okinawa, Japan, February 11–13. 2019. Pp. 008–013. <https://doi.org/10.1109/ICAIC.2019.8668989> EDN: WJXAWT
- Blei D.M., Ng A.Y., & Jordan M.I. Latent dirichlet allocation // Journal of Machine Learning Research, Vol. 3. 2003. Pp. 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- Devlin J., Chang M.W., Lee K., & Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1. Pp. 4171–4186. <https://doi.org/10.48550/arXiv.1810.04805>
- Goldberg Y. Neural network methods for natural language processing. San Rafael: Morgan & Claypool Publishers. 2017. 285 p. <https://doi.org/10.2200/S00762ED1V01Y-201703HLT037> EDN: YEQZUL
- Jurafsky D., & Martin J.H. Speech and language processing : An introduction to natural language processing, computational linguistics, and speech recognition. Upper Saddle River, NJ : Pearson Education. 2024. 1024 p.
- McCulloch W.S., & Pitts W. A logical calculus of ideas immanent in nervous activity // Bulletin of Mathematical Biophysics. 1943. Vol. 5. P. 115–133.
- Mikolov T., Chen K., Corrado G., & Dean J. Efficient Estimation of Word Representations in Vector Space // 1st International Conference on Learning Representations (ICLR 2013), Scottsdale, Arizona, USA, May 2–4. 2013. arXiv preprint arXiv:1301.3781. <https://doi.org/10.48550/arXiv.1301.3781>.
- Sapone L. Moving fast and breaking things: an analysis of social media’s revolutionary effects on culture and its impending regulation // Duquesne Law Review. 2021. Vol. 59. Pp. 362–384.
- Schiffrin D. From linguistic reference to social reality // Discourse and identity. Cambridge : Cambridge University Press. 2006. Pp. 103–131.
- Titscher S., Meyer M., Vetter E., & Wodak R. Methods of text and discourse analysis. London : Sage, 2000. 278 p.
- Widdowson H. Applied linguistics, interdisciplinarity, and disparate realities // Directions in Applied Linguistics. Clevedon, UK : Multilingual Matters. 2005. 327 p.

Сведения об авторах:

Баркович Александр Аркадьевич, доктор филологических наук, доцент, заведующий кафедрой теоретического и славянского языкознания, Белорусский государственный университет, Республика Беларусь, 220004, Минск, пр. Независимости, д. 4. *Сфера научных интересов*: русский язык, белорусский язык, английский язык, компьютерная лингвистика, дискурс-анализ, коммуникация, информация. ORCID: 0000-0001-8469-8431. SPIN-код: 5171-1479. ResearcherID: W-2342-2018. ScopusID: 57208124708. E-mail: barkovichaa@gmail.com

Астapkина Екатерина Сергеевна, кандидат филологических наук, доцент кафедры германского языкознания, Белорусский государственный университет, Республика Беларусь, 220004, Минск, пр. Независимости, д. 4. *Сфера научных интересов*: русский язык, английский язык, лексикология, семантика, коммуникация. ORCID: 0009-0005-5941-1730. SPIN-код: 3271-8936. E-mail: astapkina@gmail.com

DOI: 10.22363/2618-8163-2026-24-1-25-40

EDN: XARBFO

Research article

Neural network modeling of the semantic field “Internet” in Russian-language discourse

Alexander A. Barkovich  , Ekaterina S. Astapkina 

Belarusian State University, *Minsk, Republic of Belarus*

 barkovichaa@gmail.com

Abstract. The authors perform the linguistic analysis of neural network modeling of the semantic field “Internet” on the material of available online Russian-language content. The relevance of the study is ensured by the quality and quantity of the linguistic material in the “big data” format and by an innovative methodological approach to its meta-description with neural network instruments. The study is aimed at giving a linguistic characteristic of neural network modeling of the semantic field “Internet” in Russian-language discourse. The material was Russian-language Internet content. The volume of the content had not been limited to obtain statistically representative metadata. This approach corresponds to the mainly declarative limitations of the Internet discourse functionality. Due to the focus on the “intelligent” algorithms for processing Internet content, such as basic for our research OpenAI project, the high referentiality of language data was ensured. The authors used a wide range of methods, from component analysis to discourse analysis, with modern neural network instruments. A two-dimensional neural network modeling was carried out with cluster and stratum analysis of language units associated with the conceptual phenomenon *Internet*. The conducted research demonstrated the potential of neural network modeling techniques to study the semantic field “Internet”. The modeling identified and verified a wide range of language units whose speech functionality was associated with the conceptual phenomenon *Internet* as the core of the corresponding semantic field. The results obtained are promising; we can confidently implement the neural network modeling patterns tested in this study into linguistic practice. This, in turn, will develop the paradigm of linguistics, modernize methodological approaches to language functioning, and identify and qualify speech innovations.

Keywords: neural network instruments, cluster model, stratum model, conceptual phenomenon, linguistic representation, modeling potential

Contribution of the authors: the authors made an equal contribution to developing the concept of the research, writing and editing the article.

Conflict of interest. The authors declare that there is no conflict of interest.

Article history: received 20.07.2025; accepted 18.10.2025.

For citation: Barkovich, A.A., & Astapkina, E.S. (2025). Neural network modeling of the semantic field “Internet” in Russian-language discourse. *Russian Language Studies*, 24(1), 25–40. (In Russ.). <http://doi.org/10.22363/2618-8163-2026-24-1-25-40> EDN: XARBFO

References

- Aichner, T., & Jacob, F. (2015). Measuring the degree of corporate social media use. *International Journal of Market Research*, 2(57), 257–275. <https://doi.org/10.2501/IJMR-2015-018>
- Apresyan, Yu. D. (1995). *Selected works: in 2 vol. V. I.: Lexical semantics. Synonymous means of the language*. Moscow: Languages of Russian culture Publ. (In Russ.).
- Astapkina, E. S., & Barkovich, A. A. (2025). Dynamics of Russian language internet discourse: peculiar representation of lexis in dictionaries. *Russian Language Studies*, 23(1), 97–116. (In Russ.). <http://doi.org/10.22363/2618-8163-2025-23-1-97-116> EDN: VLFOHT
- Azimov, E. G., Kulibina, N. V., & Van, V. (2023). Linguistic and didactic potential of social networks in teaching Russian as a foreign language. *Russian Language Studies*, 21(2), 133–147. (In Russ.). <http://doi.org/10.22363/2618-8163-2023-21-2-133-147> EDN: BAUPZH
- Barkovich, A. A. (2015). Functionality of the “communicational-communicative” dyad: A discursive aspect. *Tomsk State University Journal of Philology*, (5), 37–52. (In Russ.). <https://doi.org/10.17223/19986645/37/3> EDN: VCPSFD
- Barkovich, A. (2019). Informational linguistics: Computer, internet, artificial intelligence and language. In: *2019 IEEE 1st International Conference on Artificial Intelligence in Information and Communication (ICAIIIC 2019) (Okinawa, Japan, February 11–13)* (pp. 8–13). IEEE Publ. <https://doi.org/10.1109/ICAIIIC.2019.8668989> EDN: WJXAWT
- Belyakov, M. V., & Maksimenko, O. I. (2024). The problem of semantic shift in modern Russian: Corpus research on the example of a news media discourse. *RUDN Journal of Language Studies, Semiotics and Semantics*, 15(1), 174–194. (In Russ.). <https://doi.org/10.22363/2313-2299-2024-15-1-174-194> EDN: EJFXLT
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 4171–4186. <https://doi.org/10.48550/arXiv.1810.04805>
- Dmitrieva, N. M., & Shaklein, V. M. (2022). Semantic features of the functioning of dominant verbalizers of the ethical concept BOON in the mass media. *RUDN Journal of Language Studies, Semiotics and Semantics*, 13(4), 956–971. (In Russ.). <https://doi.org/10.22363/2313-2299-2022-13-4-956-971> EDN: SCBGJS
- Dubina, L. V. (2013). Trends of developments of the Russian language in the context of internet-communication. *Tomsk State Pedagogical University Bulletin*, (2), 177–181. EDN: PYABEF. (In Russ.).
- Goldberg, Y. (2017). *Neural Network Methods for Natural Language Processing*. San Rafael: Morgan & Claypool Publishers. <https://doi.org/10.2200/S00762ED1V01Y201703HLT037> EDN: YEQZUL

- Kondrasheva, E. V., & Tinina, A. O. (2019). New lexicon of the Internet language: methods of formation, reasons for its emergence. *Era of Science*, (18), 134–139. (In Russ.). <https://doi.org/10.24411/2409-3203-2018-11830> EDN: HBPJAV
- Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Pearson Education.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In: *1st International Conference on Learning Representations (ICLR 2013), Scottsdale, Arizona, USA, May 2–4*. arXiv preprint arXiv:1301.3781. <https://doi.org/10.48550/arXiv.1301.3781>
- Novikov, L. A. (2001). Sketch of the semantic field. In: *Selected Works. Vol. II: Aesthetic Aspects of Language. Miscellanea* (pp. 554–570). Moscow: RUDN University Press. (In Russ.).
- Pêcheux, M. (1999). Content analysis and discourse theory. In: *Quadrature of meaning: The French school of discourse analysis* (pp. 302–337). Moscow: Progress Publ. (In Russ.).
- Popova, T. I., & Kolesova, D. V. (2023). The teaching model of mutual checking exercises in online course on Russian as a foreign language. *Russian Language Studies*, 21(2), 181–195. (In Russ.). <https://doi.org/10.22363/2618-8163-2023-21-2-181-195> EDN: HBPJAV
- Razumovskaya, V. A. (2024). “Supertext” as a form of a “Strong” fiction text existence. *RUDN Journal of Language Studies, Semiotics and Semantics*, 15(1), 262–275. (In Russ.). <https://doi.org/10.22363/2313-2299-2024-15-1-262-275> EDN: FOGHCU
- Sapone, L. (2021). Moving fast and breaking things: an analysis of social media’s revolutionary effects on culture and its impending regulation. *Duquesne Law Review*, 59, 362–384.
- Schiffrin, D. (2006). From linguistic reference to social reality. In: *Discourse and identity* (pp. 103–131). Cambridge: Cambridge University Press.
- Shehur, G. S. (1974). *Field Theories in Linguistics*. Moscow: Nauka Publ. (In Russ.).
- Titscher, S., Meyer, M., Vetter, E., & Wodak, R. (2000). *Methods of Text and Discourse Analysis*. London: Sage.
- Tošović, B. (2015). *Internet stylistics*. Moscow: Nauka Publ. (In Russ.).
- Ufimtseva, A. A. (1968). *Word in the lexical-semantic system of language*. Moscow: Nauka Publ. (In Russ.).
- Vinogradova, V. N. (2012). *The system of Russian word formation in the functional-stylistic aspect*. Moscow: LLC Infotech Publ. (In Russ.).
- Widdowson, H. (2005). Applied linguistics, interdisciplinarity, and disparate realities. In: *Directions in Applied Linguistics*. Clevedon, UK: Multilingual Matters.

Bio notes:

Ekaterina S. Astapkina, Candidate of Philology, Associate Professor at the Department of Germanic Linguistics, Belarusian State University, 4 Nezavisimosti Ave., Minsk, 220004, Republic of Belarus. *Research interests*: Russian language, English language, lexicology, semantics, communication. ORCID: 0009-0005-5941-1730. SPIN-code: 3271-8936. E-mail: astapkina@gmail.com

Alexander A. Barkovich, Doctor of Philology, Associate Professor, Head of the Department of Theoretical and Slavic Linguistics, Belarusian State University, 4 Nezavisimosti Ave., Minsk, 220004, Republic of Belarus. *Research interests*: Russian language, Belarusian language, English language, computational linguistics, discourse analysis, communication, information. ORCID: 0000-0001-8469-8431. SPIN-code: 5171-1479. ResearcherID: W-2342-2018. ScopusID: 57208124708. E-mail: barkovichaa@gmail.com