

Научная статья

УДК 81'322

DOI: <https://doi.org/10.18721/JHSS.14409>



## ИНДИВИДУАЛЬНАЯ СЕМАНТИКА Л. Н. ТОЛСТОГО В СВЕТЕ ВЕКТОРНЫХ МОДЕЛЕЙ

**Б.В. Орехов** 

Национальный исследовательский университет «Высшая школа экономики»,  
Москва, Российская Федерация;

Институт русской литературы (Пушкинский Дом) РАН,  
Санкт-Петербург, Российская Федерация

✉ [nevmenandr@gmail.com](mailto:nevmenandr@gmail.com)

**Аннотация.** Статья представляет цифровой подход к установлению особенностей индивидуальной семантики Л.Н. Толстого с помощью векторно-семантических моделей. Сравнение данных таких моделей позволяет отразить особенности индивидуальной авторской семантики. На лемматизированных текстах Толстого с помощью программного модуля gensim было построено две модели. Одна из них сравнивалась с уже имеющейся моделью с сайта rusvectors. Особое внимание в статье уделено лексемам «любовь» и «поле». Для первого установлена значимая в контексте мировоззрения писателя оппозиция истинной и неистинной любви, последняя выражается нейтральными для русского языка (но не для языка Толстого) словами «боготворить» и «обожать». Лексема «поле» против ожиданий выражает не батальный, а аграрный и пейзажный смысл. При сравнении векторов слов идиостиля Толстого и «Национального корпуса русского языка» оказалось возможным выявить те лексемы, которые в наименьшей степени отличаются по семантике в языке и идиостиле. К таким принадлежат, например, слова «социалист» и «таинство». Метод оказывается перспективным для дополнения традиционных лексикографических практик.

**Ключевые слова:** векторные модели, семантика, Толстой, идиостиль, значение.

**Финансирование:** Работа выполнена при финансовой поддержке Российского научного фонда в рамках программы «Множественный и повторный поэтический перевод в России: лингвостатистический и количественный подход» (грант № 23-28-01201).

**Для цитирования:** Орехов Б.В. Индивидуальная семантика Л. Н. Толстого в свете векторных моделей // Terra Linguistica. 2023. Т. 14. № 4. С. 119–129. DOI: 10.18721/JHSS.14409



## INDIVIDUAL SEMANTICS OF L. N. TOLSTOY IN THE CONTEXT OF VECTOR MODELS

**B.V. Orekhov** 

National Research University Higher School of Economics,  
Moscow, Russian Federation;  
Institute of Russian Literature (Pushkin House) of the RAS,  
St. Petersburg, Russian Federation

✉ [nevmenandr@gmail.com](mailto:nevmenandr@gmail.com)

**Abstract.** The paper presents a digital approach to establishing the features of L.N. Tolstoy's individual semantics by means of vector space models. Two models were built on Tolstoy's texts. One of them was compared with an already available model from the *rusvectores* website. Special attention in the paper is paid to lexemes “любовь” (“love”) and “поле” (“field”). For the first one, the opposition of true and non-true love, significant in the context of the writer's worldview, is established; the latter is expressed by the words “боготворить” (“idolize”) and “обожать” (“adore”), which are neutral for the Russian language (but not for Tolstoy's language). The lexeme “поле” (“field”) against expectation expresses not battle, but agrarian and landscape meaning. When comparing the word vectors of Tolstoy's idiosyncrasy and the Russian National Corpus, it was possible to identify those lexemes that differ least in semantics in the language and idiosyncrasy. For example, the words “социалист” (“socialist”) and “таинство” (“sacrament”) belong to such lexemes. The method turns out to be promising for supplementing traditional lexicographic practices.

**Keywords:** word embeddings, semantics, Tolstoy, style, meaning.

**Acknowledgements:** The reported study was funded by Russian Science Foundation (grant No. 23-28-01201) “Multiple and repeated poetic translation in Russia: linguostatistical and quantitative approach”.

**Citation:** Orekhov B.V., Individual semantics of L. N. Tolstoy in the context of vector models, *Terra Linguistica*, 14 (4) (2023) 119–129. DOI: 10.18721/JHSS.14409

### Введение

Идиостиль писателя проявляет себя на разных языковых уровнях от грамматики до фонетики. Особенно заметны индивидуальные особенности словоупотребления, конструкций и словообразования в стихе, поэтому именно стихотворная речь с этой точки зрения изучена лучше всего. Выгодным свойством поэтических произведений является их сравнительно небольшой объем, делающий возможным ручной анализ [1].

В случае, когда речь идет о значительных текстовых массивах, исследователь, стремящийся ввести в оборот максимум данных, вынужден прибегать к компьютерной обработке материала. Если программные библиотеки для работы с русскоязычной грамматикой существуют уже довольно давно [2], то семантический уровень текста попал в поле зрения исследователей, применяющих компьютерные методы, только в последнее десятилетие.

Самая актуальная технология, позволяющая изучать семантику в контексте с применением машинных методов, это векторно-семантическое представление слов. Теоретическим основанием для него служит дистрибутивная гипотеза, которую можно сформулировать как сопоставимость сходства контекстов употребления слов со сходством их семантики. В середине XX века эта идея звучала в трудах лингвистов в сугубо теоретическом ключе: «Вы узнаете слово по той компании, в которой оно ходит» (“You shall know a word by the company it keeps”) [6, с. 11], но в 2010-х годах появились эффективные способы компьютерного моделирования контекстов в векторном пространстве. В рамках такого подхода каждое слово представляет собой вектор в многомерном пространстве. Координатами вектора служит количественно выраженная встре-



чаемость соответствующего вектору слова с другими словами в текстах некоторого текстового корпуса. Таким образом, мы можем строить векторные модели дистрибуции слов (а, следовательно, и их семантики) на разных текстовых выборках. Это могут быть корпуса, состоящие из произведений одного жанра, одного автора, одной эпохи или, наоборот, вбирающие в себя максимум доступных документов. В зависимости от особенностей исходного корпуса будут находиться и векторные представления слов, отражающих специфику контекстов в корпусе.

Существует несколько алгоритмов векторизации контекста, самым популярным из которых является `word2vec` [14].

Удобство векторного представления состоит в том, что исследователю оказываются доступны те же математические операции, которые в принципе совершаются с векторами, но самой привлекательной из них становится нахождение ближайшего вектора к данному путем вычисления косинусного расстояния, что в проекции на лингвистический материал означает нахождение слова, семантически наиболее близкого к заданному.

Векторное представление слов активно используется для установления специфики различных жанров [11], для описания исторических изменений в семантике слов. Получил этот инструмент и некоторое распространение в сфере исследования художественного текста.

На англоязычном материале векторные модели позволили сделать вывод о тенденциях в изображении персонажей художественной прозы [6]. Но чаще эта технология использовалась при изучении поэтического языка на английском [7], русском [8], башкирском [9, 10] и китайском материале [11, 12].

Качество векторной модели заметно зависит от размера корпуса. Чем больше корпус, тем лучше построенная на нем векторная модель отражает специфику семантики отдельных слов. С этой точки зрения очевидно, что наиболее выгодным для исследования корпусом текстов одного автора становится наследие Л. Н. Толстого, полное собрание сочинений которого включает 90 томов, что соответствует приблизительно 7,8 млн словоупотреблений [13]. Кроме того, собрание сочинений классика было качественно оцифровано и представлено в виде семантического издания [14], что также облегчает предварительную подготовку материала для анализа.

### **Постановка задачи**

Исследовательские задачи, которые можно поставить в связи с анализом текстов Толстого с помощью векторных моделей, лежат в плоскости сравнения семантики отдельных лексем у автора «Войны и мира» с семантикой тех же лексем в других речевых произведениях на русском языке, а также в плоскости установления идиостилевых сходств и различий между полными словами в авторском употреблении.

Перспективными эти уточнения являются в частности для авторской лексикографии [14], которая получает возможность фиксировать тонкие семантические различия в контексте идиостиля.

В конечном счете речь идет о моделировании семантики слов при помощи алгоритмов, апробированных и успешно используемых в индустрии информационного поиска.

### **Методы и источники данных**

Исследование, которое в данном случае проводится на материале текстов Л.Н. Толстого, повторяет дизайн уже осуществленных научных работ [8–11] и др.

Общий характер исследовательских операций таков. Слова в текстах автоматическими средствами приводятся к своей словарной форме. С помощью специального программного обеспечения, рассчитанного на работу с векторно-семантическими моделями (в нашем случае это пакет `gensim` для языка программирования Python) строится векторно-семантическая модель, отражающая функционирование конкретного слова-вектора в контексте. Затем вектора слов сравниваются либо с другими векторами в той же модели, либо с теми же векторами в других



моделях. Сравнение обычно производится по списку ближайших соседей-векторов к данному. На основании сравнения делается вывод об особенностях контекстов, в которых встречается интересующее нас слово, а, следовательно, об особенностях его семантики. Так, разница между векторами некоторого слова в модели, построенной на текстах Толстого, и в модели, построенной на текстах Национального корпуса русского языка, будет демонстрировать специфику индивидуальной семантики Толстого.

Мы построили две векторных семантических модели на наличных текстах Л.Н. Толстого (7,8 млн словоупотреблений), предварительно лемматизировав их и удалив неполнозначные слова. Первая модель обучена с помощью алгоритма Skip-gram, включает около 45 тыс. слов-векторов с неединичным входением, длина вектора 500, размер окна: 2. Выбор этих параметров был продиктован наличием уже готовой аналогичной модели<sup>1</sup>, построенной на материале текстов Национального корпуса русского языка (НКРЯ), и перспективой сравнения двух векторных представлений с помощью того же модуля gensim. Вторая модель обучена на том же корпусе, но при помощи алгоритма CBOW (Continuous Bag of Words) в 300-мерном пространстве; размер окна контекста для каждого слова был равен 10. Эта модель будет использоваться для установления отношений между словами внутри идиостиля Толстого. Выбор параметров был обусловлен общими соображениями, подсказываемыми опытом работы с таким материалом: широкий размер окна, обычно позволяет лучше отразить в векторном пространстве связи между словами на уровне ассоциаций.

### Результаты

На первом этапе обратимся к сравнению представлений в двух моделях. В модели Толстого использовались все жанры, отраженные в 90-томном собрании сочинений, включая художественные, публицистические, эго-документы и эпистолярный. В модели НКРЯ также нашли отражения тексты разных жанров, присутствующих в корпусе. Но гораздо более значимо и может вызывать потенциальные возражения то, что эта модель построена на текстах, не только современных Л.Н. Толстому, но и на текстах XX и XXI века. Сравнение семантики, характерной для идиостиля Толстого, с семантикой, усредняющей эволюцию языка в течение нескольких веков, может показаться неоправданной. Однако, на наш взгляд, именно такое сравнение способно заострить особенности, значимые для современного носителя языка, пусть и не всегда отражающие именно индивидуальную авторскую стилистику, а восходящие к языку эпохи. Модель НКРЯ фиксирует синхронную для нас палитру значений. Модель Толстого в сравнении с ней будет отражать и те семантические сдвиги, которые «мало или вовсе незаметны взгляду без проверки узусом эпохи – нынешней и тогдашней. Попав в ловушку, мы склонны подставлять "свое" значение. Механизм такой подмены Анна А. Зализняк удачно называет эффектом ближней семантической эволюции, определяя его как "неправильное понимание текста, возникающее из-за того, что читатель не распознает семантический сдвиг, произошедший в слове или конструкции родного языка, и понимает их в современном ему значении"» [15, с. 25].

Последовательное сравнение десятков тысяч лексем, отраженных в модели, построенной на текстах Л.Н. Толстого, с теми же лексемами в модели, построенной на текстах Национального корпуса русского языка, – объемная задача, поэтому мы сосредоточимся на отдельных случаях, демонстрирующих возможности метода, и представляющих интерес для описания идиостиля писателя.

Одним из ключевых понятий для прозы и философии Толстого является понятие любви, которому посвящен целый ряд специальных работ.

О.В. Ланская рассматривает «любовь» в романе «Война и мир», пользуясь при этом концептуальным аппаратом когнитивной лингвистики конца 2000-х годов [16]. Объем анализируемого

<sup>1</sup> <https://rusvectors.org/ru/models/>



понятия составлен из нескольких лексем (например, «улыбка»), что несовместимо с нашим подходом, опирающимся на репрезентацию в тексте конкретных слов. Автор делает вывод о включенности концепта «любовь» у Толстого в ряд оппозиций: «любовь – бедность», «любовь – жертва», «любовь – долг».

А.К. Базилевская обращается к теме любви в ранних художественных текстах Толстого, пересказывая их содержание, и включая рассматриваемую тему в контекст предствлений писателя о других ментальных объектах – мыслях и чувствах [17]. Одним из аспектов их описания, по замечанию исследователя, становятся искания «истинной любви». Это имплицитное указание на оппозицию истинного и ложного у Толстого нам представляется чрезвычайно важным.

Работа М.Л. Гельфонд представляет не филологический, а философский анализ темы любви у Толстого. Автор указывает на религиозную составляющую этого понятия и на его эволюционировавший со временем характер: «Только спустя годы Толстой полностью откажется от всех ограничений своей этикоцентрической трактовки заповеди любви к врагам, окончательно распространив предметность ее обязывающей нормативности не только на чужеземцев, но на всех людей и, даже, на все живое и все существующее в целом, и окончательно провозгласив требование любви ко всем, включая своих личных врагов». М.Л. Гельфонд также обращает внимание на «истинность» любви как на важнейшую характеристику этого понятия у Толстого: «В этой монолитной целостности нет места какому бы то ни было онтологическому сепаратизму или отпадению от единой основы всего сущего, коей и является истинная любовь. <...> Для полноты квалификации толстовской модели истинной любви следует указать на то, что закон любви не есть внешнее дисциплинарное правило, механическое следование предписаниям которого является определяющим критерием оценки нравственного статуса личности исполнителя. Реализация закона любви состоит в поступательном движении человека по направлению к недостижимому в условиях его наличного существования идеалу всеобщего любовного единения со всеми и всем без исключения, т. е. в неограниченной перспективе – с бесконечной универсальностью Всего» [18].

Близкую нам методологически лексикографическую трактовку термина «любовь» в текстах Толстого дает Е.В. Ефимова [19]. Автор проводит сравнение лексем «любовь» и «привязанность», констатируя их семантическое сходство. Важной для нас идеей в работе является очередное обращение к оппозиции истинной и не-истинной любви: «Выявленные в процессе анализа семантические особенности лексемы любовь вполне соответствуют взглядам Л.Н. Толстого: естественная плотская любовь между мужчиной и женщиной противопоставлена любви истинной, она не имеет ничего общего с любовью к ближнему, заповеданной христом Спасителем в Его Евангелии. Плотская любовь стремится доставить любимому земное благополучие. В ней нет постоянства, она может разгореться очень сильно, но нередко после короткого времени легко обращается в отвращение и даже ненависть, что характерно для отношений Позднышева с женой» [19].

А.И. Вялов приходит к выводу о сходстве в понимании любви у Л.Н. Толстого и К.Д. Кавелина, подчеркивая при этом христианские корни этого понимания [20].

Религиозно-философскому анализу концепции любви у Толстого посвящена статья Е.Д. Мелешко. Автор обращает внимание на сложность в установлении истоков этой концепции в мировоззрении писателя, но все же констатирует, что «толстовская метафизика любви, несмотря на ее очевидный платонизм, вполне соответствует духу христианско-богословской традиции, особенно в ее патриотическом выражении» [21].

Сравнение векторного представления лексемы «любовь» в моделях, построенных на текстах Толстого и НКРЯ, должно дать представление о специфике этого слова в творчестве писателя, проявленной на общезыковом фоне.

Метод сравнения в таких случаях обычно представляет собой нахождение 10 ближайших векторов к данному путем вычисления косинусного расстояния между ними и последующее



сравнение двух списков. В случае, если ближайшие вектора в обеих моделях полностью совпадают, следует сделать вывод о тождестве семантики анализируемых слов. Существенные различия в составе списков будут говорить о разнице в семантике.

Для слова «любовь» в модели Толстого получен следующий список 10 ближайших соседей:

1. полюбить
2. уважать
3. ненавидеть
4. страстно
5. любящий
6. ценить
7. презирать
8. дорожить
9. ближний
10. жалеть

Обратим внимание, что ближайшие соседи не всегда являются членами синонимического ряда: общие контексты употребления, на обобщении которых строится модель, могут объединять и антонимы. Так, не следует удивляться появлению в этом ряду слов «ненавидеть» и «презирать». Именно поэтому в терминах анализа векторных представлений следует говорить не о синонимах, а о квази-синонимах, в число которых могут войти антонимы, гиперонимы, гипонимы и т. д.

Для модели НКРЯ аналогичный список выглядит следующим образом:

1. обожать
2. полюбить
3. любить NOUN
4. уважать
5. нравиться
6. ненавидеть
7. любимый
8. любить ADJ
9. боготворить
10. презирать

В этом списке обращает на себя внимание присутствие слов «любить» в грамматической интерпретации существительного (NOUN) и прилагательного (ADJ). Очевидно, речь идет об ошибках автоматического морфологического анализа при обработке текстов корпуса.

В остальном списке весьма похожи за двумя значимыми исключениями.

Для установления специфики семантики слова «любовь» в идиостиле Толстого значимым представляется наличие в списке из модели НКРЯ слов «обожать» и «боготворить», отсутствующих среди ближайших квази-синонимов в модели Толстого. Данные модели свидетельствуют, что для контекстов, порождаемых с участием этих слов на русском языке, в целом характерна высокая степень сходства с контекстами употребления слова «любовь». Для контекстов Толстого это не так. Иными словами, для носителей русского языка естественно представление об обожании и боготворении как аналогах любви, а в представлении Толстого это более далекие друг от друга понятия.

Обратимся к соответствующим контекстам.

«— Ничего нет ужасного, — сказал Новодворов, прислушивавшийся к разговору. — Массы всегда обожают только власть, — сказал он своим трещащим голосом. — Правительство властвует — они обожают его и ненавидят нас; завтра мы будем во власти — они будут обожать нас...»<sup>2</sup>

<sup>2</sup> Толстой Л. Н. Полное собрание сочинений в 90 тт. М.: Государственное издательство «Художественная литература». Т. 32. С. 398.



Персонаж «Воскресения» подчеркивает зыбкость обожания, его изменчивость, нестабильность. Сходный акцент мы находим и в цитате из «Юности»: «...второй, Сергей, которого я обожал некогда, был тоже в Петербурге большим толстым кадетом в Пажеском Корпусе»<sup>3</sup>. Такое настойчивое указание на преходящесть обожания позволяет установить, что для Толстого это понятие противопоставлено любви как искреннему, целостному чувству. «Обожать» в идиостиле Толстого в отличие от общеязыковых данных не синонимично «любить», лексема представляет другое понятие, лишенное важного для писателя представление об истинности.

Аналогичная функция проявляет себя и в лексеме «боготворить». Одно из свидетельств этого — явно иронический контекст в повести «Хаджи-Мурат»: «Девушка эта рассказала Николаю, как она с детства еще, по портретам, влюбилась в него, боготворила его и решила во что бы то ни стало добиться его внимания»<sup>4</sup>.

Внутреннее противоречие понятия подчеркивается и в контексте из повести «Крейцера соната»: «И заметьте хитрость дьявола: ну, наслаждение, удовольствие, так так бы и знать, что удовольствие, что женщина сладкий кусок. Нет, сначала рыцари уверяли, что они боготворят женщину (боготворят, а всё-таки смотрят на нее как на орудие наслаждения). Теперь уже уверяют, что уважают женщину. Одни уступают ей место, поднимают ей платки; другие признают ее права на занятие всех должностей, на участие в правлении и т. д.»<sup>5</sup>.

Как мы видим, слова, не входящие в перечень ближайших квази-синонимов «любви» в модели Толстого, включаются в важную для писателя оппозицию истинного и ложного. Если «любовь» для автора «Войны и мира» — это прежде всего истинная любовь, то ее ложные варианты репрезентируются в тексте с помощью лексем, которые в текстах других носителей языка не обладают этой спецификой. Толстой видоизменяет семантику привычных слов, придавая им индивидуально авторские оттенки.

Данные двух моделей (Толстого и НКРЯ) позволяют с помощью аналогичных методов установить и зону наименьшего проявления индивидуальности в семантике, то есть выявить список тех слов, которые обладают приблизительно общим значением у писателя и в языке вообще.

Метод установления этой зоны — уже описанное попарное сравнение списков квази-синонимов для конкретных слов. Те слова, для которых списки квази-синонимов не отличаются, можно считать сходными по их семантике в идиостиле и в языке. Мету отличия обычно вычисляют с помощью коэффициента Жаккара [11].

В обеих моделях находится 34 тыс. общих лексем, для которых можно составить списки квази-синонимов. При этом только 8,5 тыс. слов имеют хотя бы единичные пересечения в списках квази-синонимов из 30 позиций. Остальные слова при такой исследовательской оптике не демонстрируют семантического сходства у Толстого и в текстах НКРЯ.

В случае Толстого наибольшим сходством обладают глаголы (значение коэффициента Жаккара дано в скобках, оно находится в диапазоне от 0, что соответствует отсутствию совпадений, до 1, что означает полное совпадение списков) *просыпаться* (0.428), *ехать* (0.379), *выпивать* (0.333), *поехать* (0.333), *грабить* (0.290), *молиться* (0.290), *подвигать* (0.290), *сидеть* (0.25) и существительные *таинство* (0.428), *радость* (0.379), *беспокойство* (0.333), *овца* (0.333), *ненависть* (0.290), *Италия* (0.290), *социалист* (0.290), *священник* (0.290).

Если церковный термин «таинство» в силу ограниченности бытования действительно должен обладать высокой мерой сходства в употреблении в текстах Толстого и НКРЯ, слово присутствие в этом перечне слова «социалист», значение и контексты которого должны были существенно поменяться в XX веке, привлекает гораздо больше внимания.

Вероятно, причина в том, что контексты слова «социалист» у Толстого предполагают перечисления, включающие близкий к фиксированному набор политических терминов, фигурирующий

<sup>3</sup> Толстой Л. Н. Полное собрание сочинений в 90 тт. М.: Государственное издательство «Художественная литература». Т. 2. С. 124.

<sup>4</sup> Толстой Л. Н. Полное собрание сочинений в 90 тт. М.: Государственное издательство «Художественная литература». Т. 35. С. 68.

<sup>5</sup> Толстой Л. Н. Полное собрание сочинений в 90 тт. М.: Государственное издательство «Художественная литература». Т. 27. С. 37.

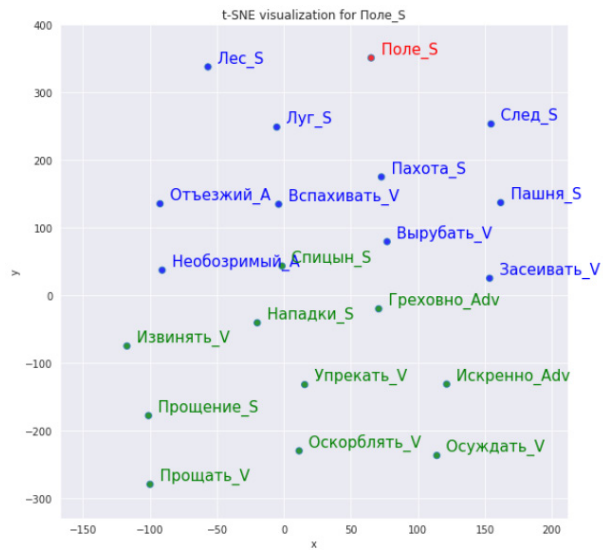


Рис. 1. Результаты исследования

Fig. 1. Research results

и в текстах других авторов: «Она будет разрушена не потому, что ее разрушат революционеры, анархисты, рабочие, государственные социалисты, Японцы или Китайцы, а она будет разрушена потому, что она уже разрушена на главную половину — она разрушена в сознании людей» (“Carthago delenda est”)<sup>6</sup>.

Вторая модель позволяет исследователю установить особенности взаимоотношения слов-векторов внутри семантического пространства Толстого. В качестве метода здесь также выступает поиск ближайших соседей и одновременно нахождение наиболее удаленных по косинусному расстоянию слов от данного.

Одним из любопытных исследовательских вопросов, которые можно задать при таком подходе, будет вопрос о месте лексемы «поле» в идиостилистической семантике Толстого. С одной стороны, писатель известен как баталист, его умение изображать масштабные сражения, «поле боя», не вызывает у читателя сомнений. С другой стороны, важным для биографии и мировоззрения Толстого является концепция аграрного труда, «работы в поле». Запрос к векторно-семантической модели даст понимание того, какое из этих значений в текстах Толстого превалирует. Если в списке квази-синонимов будет доминировать военная лексика, это будет означать, что батальные компоненты значения «поля» в целом для языка Толстого важнее, и наоборот, отнесенность ближайших соседей к семантическим классам, связанным с аграрной деятельностью, продемонстрирует, что этот компонент значения в идиостиле Толстого занимает большее место.

### Выводы

Результаты исследования отражены на рис. 1. Он представляет векторно-семантическое пространство, сведенное до двух измерений с помощью метода t-SNE. Ближайшими соседями к слову «поле» оказываются, с одной стороны, аграрные термины «пахота», «пашня», «засеивать», «вспахивать», с другой стороны, компоненты пейзажа «лес» и «луг». Таким образом, батальная семантика слова «поле» оказывается для идиостиля Толстого не ведущей. Любопытно в этом смысле, что даже в описании разгара сражения при Бородине Толстой специально обращает

<sup>6</sup> Толстой Л. Н. Полное собрание сочинений в 90 тт. М.: Государственное издательство «Художественная литература». Т. 27. С. 534.





внимание на аграрно значимый элемент окружающей обстановки: князь Андрей находится рядом с овсяным полем: «Князь Андрей точно так же, как и все люди полка, нахмуренный и бледный, ходил взад и вперед по лугу подле овсяного поля от одной межи до другой, заложив назад руки и опустив голову. Делать и приказывать ему нечего было. Всё делалось само собою. Убитых оттаскивали за фронт, раненых относили, ряды смыкались. Ежели отбегали солдаты, то они тотчас же поспешно возвращались».

Векторно-семантические модели показали свою перспективность для изучения идиостилевой семантики писателя. Этот количественный метод приложим не к любому текстовому корпусу, так как чувствителен к объему исходных данных. Наследие Толстого как раз позволяет строить репрезентативные модели. То, что в докомпьютерную эпоху было проблемой (невероятный объем исходного материала), для применения машинных методов оказывается благом.

### СПИСОК ИСТОЧНИКОВ

1. **Тарасова И.А.** Эволюция идиостиля Георгия Иванова в зеркале его словаря // Авторская лексикография и история слов. М.: Азбуковник, 2013. С. 313–318.
2. **Зобнин А.И., Носырев Г.В.** Морфологический анализатор MyStem 3.0 // Труды Института русского языка им. В.В. Виноградова. 2015. Т. 6. С. 300–310.
3. **Bonch-Osmolovskaya A., Orekhov B., Tolstaya F.** Digital Edition of Complete Tolstoy's Heritage: OCR Crowd Sourcing Initiative, Literary Scholarship and User Scenarios // Digital Humanities 2023: Book of Abstracts: Zentrum für Informationsmodellierung - Austrian Centre for Digital Humanities, University of Graz, 2023. Pp. 288–289.
4. **Шестакова Л.Л.** Русская авторская лексикография: Теория, история, современность. М.: Языки славянских культур, 2011.
5. **Herbelot A.** The semantics of poetry: A distributional reading // Digital Scholarship in the Humanities. 2015. Т. 30. №. 4. С. 516–531.
6. **Firth J.R.** Papers in Linguistics 1934–1951. London: Oxford, 1957. 233 p.
7. **Grayson S. et al.** Novel2Vec: Characterising 19th Century Fiction via Word Embeddings // AICS. 2016. Pp. 68–79.
8. **Орехов Б.В.** Стихи и проза через призму дистрибутивной семантики // Острова любви Бор-Феда: Сборник к 90-летию Бориса Федоровича Егорова. СПб.: Росток, 2016. С. 652–655.
9. **Гречачин В.А.** Язык русской и башкирской поэзии XX века в свете дистрибутивного подхода. Автореферат ... кандидата филологических наук. Уфа, 2018. 26 с.
10. **Орехов Б.В.** Башкирский стих XX века. Корпусное исследование. СПб.: Алетейя, 2019. 335 с.
11. **Kutuzov A., Kuzmenko E.** Comparing Neural Lexical Models of a Classic National Corpus and a Web Corpus: The Case for Russian // Lecture Notes in Computer Science. 2015. Pp. 47–58. DOI: 10.1007/978-3-319-18111-0\_4
12. **Zorkina M.** Describing Objects in Tang Dynasty Poetic Language: A Study Based on Word Embeddings // Journal of Chinese Literature and Culture. 2018. Т. 5, №. 2. С. 250–275.
13. **Bonch-Osmolovskaya A. et al.** Tolstoy semanticized: Constructing a digital edition for knowledge discovery // Journal of Web Semantics. 2019. Т. 59. С. 100483.
14. **Baroni M., Dinu G., Kruszewski G.** Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors // Proceedings of the 52<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2014. Pp. 238–247.
15. **Бобрик М.А.** Проза Осипа Манделъштама «Холодное лето» (1923). Языковой комментарий. СПб.: Издательство Европейского университета в Санкт-Петербурге, 2018. 224 с.
16. **Ланская О.В.** Концепт 'любовь' в романе Л.Н. Толстого «Война и мир» // Вопросы когнитивной лингвистики. 2008. №. 4. С. 74–77.
17. **Базилевская А.К.** «Роды любви» в раннем творчестве Л.Н. Толстого // Вестник культурологии. 2008. №. 3. С. 58–74.
18. **Гельфонд М.Л.** Философия любви Л.Н. Толстого // Известия Тульского государственного университета. Гуманитарные науки. 2011. №. 1. С. 3–13.



19. **Ефимова Е.В.** Семантика лексемы «Любовь» в повести Л.Н. Толстого «Крейцерова соната» // Известия Волгоградского государственного педагогического университета. 2014. № 2 (87). С. 124–127.

20. **Вялов А.И.** Понятие любви в учениях К.Д. Кавелина и Л.Н. Толстого // Гуманитарные ведомости ТГПУ им. Л.Н. Толстого. 2012. № 3. С. 46–49.

21. **Мелешко Е.Д.** Метафизика и этика любви Л.Н. Толстого: единство эроса и агапэ // Гуманитарные ведомости ТГПУ им. ЛН Толстого. 2017. № 1 (21). С. 5–21.

## REFERENCES

[1] **Tarasova I.A.** Evolyutsiya idiostilya Georgiya Ivanova v zerkale yego slovarya [Evolution of Georgi Ivanov's idiostyle in the mirror of his dictionary], *Avtorskaya leksikografiya i istoriya slov*. M.: Azbukovnik, 2013. Pp. 313–318

[2] **Zobnin A.I., Nosyrev G.V.** Morfologicheskiy analizator MyStem 3.0 [Morphological analyzer MyStem 3.0], *Trudy Instituta russkogo yazyka im. V.V. Vinogradova*, 6 (2015) 300–310.

[3] **Bonch-Osmolovskaya A., Orekhov B., Tolstaya F.** Digital Edition of Complete Tolstoy's Heritage: OCR Crowd Sourcing Initiative, Literary Scholarship and User Scenarios, *Digital Humanities 2023: Book of Abstracts: Zentrum für Informationsmodellierung – Austrian Centre for Digital Humanities, University of Graz*, 2023. Pp. 288–289.

[4] **Shestakova L.L.** Russkaya avtorskaya leksikografiya: Teoriya, istoriya, sovremennost [Russian author's lexicography: Theory, history, modernity]. M.: Languages of Slavic cultures, 2011.

[5] **Herbelot A.** The semantics of poetry: A distributional reading, *Digital Scholarship in the Humanities*, 30 (4) (2015) 516–531.

[6] **Firth J.R.** *Papers in Linguistics 1934–1951*. London: Oxford, 1957.

[7] **Grayson S. et al.** Novel2Vec: Characterising 19<sup>th</sup> Century Fiction via Word Embeddings, *AICS*, 2016. Pp. 68–79.

[8] **Orekhov B.V.** Stikhi i proza cherez prizmu distributivnoy semantiki [Poetry and prose through the prism of distributive semantics], *Ostrova lyubvi BorFed: Sbornik k 90-letiyu Borisa Fedorovicha Yegorova [BORFED's Love Islands: A collection dedicated to the 90<sup>th</sup> anniversary of Boris Fedorovich Egorov]*. SPb.: Rostok, 2016. Pp. 652–655.

[9] **Grechachin V.A.** Yazyk russkoy i bashkirskoy poezii XX veka v svete distributivnogo podkhoda [Language of Russian and Bashkir poetry of the XX century in the light of the distributive approach]. The abstract ... Candidate of Philological Sciences. Ufa, 2018.

[10] **Orekhov B.V.** Bashkirskiy stikh XX veka. Korpusnoye issledovaniye [Bashkir verse of the XX century. Case study]. SPb.: Aleteyya, 2019.

[11] **Kutuzov A., Kuzmenko E.** Comparing Neural Lexical Models of a Classic National Corpus and a Web Corpus: The Case for Russian, *Lecture Notes in Computer Science*, (2015) 47–58. DOI: 10.1007/978-3-319-18111-0\_4

[12] **Zorkina M.** Describing Objects in Tang Dynasty Poetic Language: A Study Based on Word Embeddings, *Journal of Chinese Literature and Culture*, 5 (2) (2018) 250–275.

[13] **Bonch-Osmolovskaya A. et al.** Tolstoy semanticized: Constructing a digital edition for knowledge discovery, *Journal of Web Semantics*, 59 (2019) 100483.

[14] **Baroni M., Dinu G., Kruszewski G.** Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors, *Proceedings of the 52<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2014. Pp. 238–247.

[15] **Bobrik M.A.** Proza Osipa Mandelshtama “Kholodnoye leto” (1923). Yazykovoy kommentariy [Osip Mandelstam's prose “Cold Summer” (1923). Language commentary]. SPb.: Publishing House of the European University in St. Petersburg, 2018.

[16] **Lanskaya O.V.** Kontsept'lyubov v romane LN Tolstogo «Voyna i mir» [The concept of love in the novel “War and Peace” by LN Tolstoy], *Voprosy kognitivnoy lingvistiki [Questions of cognitive linguistics]*, 4 (2008) 74–77.

[17] **Bazilevskaya A.K.** «Rody lyubvi» v rannem tvorchestve LN Tolstogo [The “types of love” in the early work of LN Tolstoy], *Vestnik kulturologii*, 3 (2008) 58–74.



[18] **Gelfond M.L.** Filosofiya lyubvi L.N. Tolstogo [L.N. Tolstoy's philosophy of love], Izvestiya Tulskego gosudarstvennogo universiteta. Gumanitarnyye nauki [Proceedings of Tula State University. Humanities], 1 (2011) 3–13.

[19] **Yefimova Ye.V.** Semantics of the lexeme “love” in the story by L.N. Tolstoy “the kreutzer Sonata”, Izvestia of the Volgograd State Pedagogical University, 2 (87) (2014) 124–127.

[20] **Vyalov A.I.** Ponyatiye lyubvi v ucheniyakh K.D. Kavelina i L.N. Tolstogo [The concept of love in the doctrines of KD Kavelin and L.N. Tolstoy], Gumanitarnyye vedomosti TGPU im. L.N. Tolstogo, 3 (2012) 46–49.

[21] **Meleshko Ye.D.** Metaphysics and ethics of love of Leo Tolstoy: unity of Eros and Agape, Gumanitarnyye vedomosti TGPU im. L.N. Tolstogo, 1 (21) (2017) 5–21.

#### **СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR**

**Орехов Борис Валерьевич**

**Boris V. Orekhov**

E-mail: [nevmenandr@gmail.com](mailto:nevmenandr@gmail.com)

<https://orcid.org/0000-0002-9099-0436>

*Поступила: 16.08.2023; Одобрена: 10.11.2023; Принята: 15.11.2023.*

*Submitted: 16.08.2023; Approved: 10.11.2023; Accepted: 15.11.2023.*