

DOI: 10.18721/JHSS.10307

УДК: 81.38:004.912

## **АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ МЕНТАЛЬНЫХ ДЕЙСТВИЙ, РЕАЛИЗУЕМЫХ В НАУЧНЫХ ЭМПИРИЧЕСКИХ ТЕКСТАХ**

**В.А. Салимовский<sup>1</sup>, Д.А. Девяткин<sup>2</sup>,  
Л.А. Каджая<sup>1</sup>, В.А. Мишланов<sup>1</sup>**

<sup>1</sup> Пермский государственный национальный исследовательский университет, г. Пермь, Российская Федерация

<sup>2</sup> Институт проблем искусственного интеллекта Федерального исследовательского центра «Информатика и управление» РАН, Москва, Российская Федерация

Статья посвящена вопросам лингвистического и математического обеспечения автоматической идентификации познавательного-коммуникативных действий, реализующихся в двух речевых жанрах научного эмпирического текста – описании нового для науки явления и классификации опытных данных. Авторы исходят из положения о том, что в жанровых формах объективируются скрытые от непосредственного наблюдения когнитивные процессы, поэтому жанры речи оказываются тем звеном в цепи познавательного-коммуникативной деятельности, которое позволяет систематизировать и формализовать эти процессы. Категория речевой системности, базовая в функциональной стилистике и некоторых направлениях дискурсивного анализа, эксплицируется по отношению к жанровому уровню абстракции. Акт речи рассматривается не изолированно от деятельности, а как ее элемент. Предложен метод распознавания объективируемых в речевом произведении ментальных действий, предполагающий процедуры автоматического лингвистического анализа текстов, сопоставления неоднородных семантических сетей высказываний с контекстно свободными шаблонами, а также уточнения содержания совершаемых действий с использованием классификации последовательностей. Представлены результаты автоматической идентификации познавательного-коммуникативных действий, воплощаемых в текстах указанных жанров.

**Ключевые слова:** речевой жанр, познавательного-коммуникативное действие, речевая системность, автоматический анализ текста, когнитивное моделирование

**Ссылка при цитировании:** Салимовский В.А., Девяткин Д.А., Каджая Л.А., Мишланов В.А. Автоматическое распознавание ментальных действий, реализуемых в научных эмпирических текстах // Научно-технические ведомости СПбГПУ. Гуманитарные и общественные науки. 2019. Т. 10, № 3. С. 74-88. DOI: 10.18721/JHSS.10307

## **AUTOMATIC IDENTIFICATION OF MENTAL ACTIONS IN SCIENTIFIC EMPIRICAL TEXTS**

**V.A. Salimovsky<sup>1</sup>, D.A. Devyatkin<sup>2</sup>,  
L.A. Kadzhaya<sup>1</sup>, V.A. Mishlanov<sup>1</sup>**

<sup>1</sup> Perm State University, Perm, Russian Federation

<sup>2</sup> Institute of Artificial Intelligence Problems Federal Research Center "Informatics and Management" RAS, Moscow, Russian Federation

The article is devoted to the issues of linguistic and mathematical support for automatic identification of cognitive-communicative actions that are realized in two speech genres of a scientific empirical text - new phenomenon description and experimental data classification. The authors consider that cognitive processes concealed from direct observation are objectified in genre forms, and the latter are such link in the chain of cognitive-communicative activity that allow to systematize and formalize these processes. The category of speech system, a basic one in functional stylistics and some areas



of discourse analysis, is explicated as a genre-related level of abstraction. A speech act is considered not activity-isolated but as its intrinsic element. We propose a method for identifying mental actions objectified in a text that supposes the procedures of automatic linguistic text analysis, matching heterogeneous semantic networks of sentences with contextually free patterns, and also specifying the content of the performed actions using sequence classification. The results of automatic identification of cognitive-communicative actions realized in the texts of these genres are presented.

**Keywords :** speech genre, cognitive-communicative action, speech system, automatic text analysis, cognitive modeling

**Citation:** V.A. Salimovsky, D.A. Devyatkin, L.A. Kadzhaya, V.A. Mishlanov, Automatic identification of mental actions in scientific empirical texts, St. Petersburg State Polytechnical University Journal. Humanities and Social Sciences, 10 (3) (2019) 74-88. DOI: 10.18721/JHSS.10307

### Введение

В настоящей работе развивается жанроведческий подход к автоматическому анализу текста. Под речевыми жанрами понимаются относительно устойчивые формы (модели) духовной социокультурной деятельности, осуществляющейся в бытовых ситуациях, научной, политической, художественной и других сферах, на ступени ее объективации посредством системы речевых действий, отраженных в тексте как единице общения [1]. Эта дефиниция основывается на положениях жанроведческой теории М.М. Бахтина о речевых жанрах как формах высказывания, словесного выражения мысли и о необходимости их междисциплинарного изучения [2, 3]. Мы опираемся также на базовое для функциональной стилистики понятие речевой системности [4, 5], определяемое как «взаимосвязь равноуровневых языковых и текстовых единиц в конкретной речевой разновидности, основанной на выполнении единой коммуникативной цели и общей функции, обусловленная экстралингвистической базой этой разновидности, прежде всего назначением в обществе соответствующей формы общественного сознания» [5: 458].

Предложенное понимание речевого жанра хорошо согласуется с другими его современными концепциями, разрабатываемыми на деятельностной основе [6, 7, 8, 9 и др.]. Своеобразие нашей трактовки жанра состоит прежде всего в том, что в ней акцентируется операциональный аспект решения некоторой познавательно-коммуникативной задачи с помощью воспроизводимой модели речевых действий, сформировавшейся в той или иной коммуникативной сфере.

Как известно, анализ познавательной деятельности не поддается прямому внешнему наблюдению и требует обращения к материальным средствам ее объективации, прежде всего речевым. Поэтому подход к жанру речи как к форме организации ментальных процессов открывает новые возможности их систематизи-

рованного описания и формализации в целях компьютерного моделирования мышления [10, 11, 12].

Понятие речевого действия (акта) – одно из базовых в лингвистической прагматике. В нашем исследовании, в отличие от теории речевых актов, действия рассматриваются не изолированно от деятельности, а как ее элементы. Для этого в анализ вводятся психологические положения о ее общем строении. Как писал А.Н. Леонтьев, «основными “составляющими” отдельных человеческих деятельностей являются осуществляющие их действия... Действие – это процесс, подчиняющийся сознательной цели... Сколько-нибудь развернутая деятельность предполагает достижение ряда конкретных целей, из числа которых некоторые связаны между собой жесткой последовательностью» [14: 81-82]. Указанные характеристики человеческой деятельности отчетливо проявляются в исследовании познавательных действий, воплощенных в научно-речевом произведении [1].

В лингвостилистике категория речевой системности первоначально разрабатывалась при решении актуальной в 1960-70-е гг. задачи выделения и описания функциональных макростилей. Отсюда понимание коммуникативной цели в общем, широком и объективном смысле – как назначения той или иной формы общественного сознания и соответствующего вида деятельности. Исследуя варьирование речевой системности в научных текстах, мы снижаем уровень абстракции и рассматриваем цели научно-познавательного процесса в том виде, в каком они отражены в его генетической структуре [15, 16 и др.]. Так, основные целеустановки эмпирического исследования состоят в 1) описании фактов, 2) их классификации, 3) обнаружении закономерностей на базе опытных данных. В теоретическом исследовании реализуются целеустановки на 1) определение его исходных принципов, 2) развертывание системы понятий в процессе построения теории и 3)

верификацию последней. Каждый из указанных этапов научной эмпирической и теоретической деятельности предполагает реализацию комплекса более частных целей.

Основываясь на этих положениях, мы изучаем системность научной речи как взаимосвязь на текстовой плоскости языковых единиц разных уровней, определяемую системой целеустановок генетической структуры научно-познавательного процесса. Типовые познавательные действия, реализующие эти установки, воплощаются в тексте с помощью устойчивого набора языковых средств [1, 17], который становится объектом программирования.

Цель нашего исследования – лингвистическое и программное обеспечение автоматического распознавания познавательных коммуникативных действий, образующих форму двух жанров научного эмпирического текста – *описания нового для науки явления и классификации данных наблюдения*. Для этого выявляются и формализуются показатели речевой системности указанных жанров.

Под этим углом зрения были проанализированы тексты 80 статей разных авторов по наукам о земле, биологии, психологии и лингвистике, в которых реализуются рассматриваемые жанровые формы.

В последнее время в работах, посвященных организации научного текста, значительное внимание уделяется рассмотрению способов и средств выдвижения авторского результата [18, 19, 20 и др.]. Эти исследования используются при решении прикладных задач, в частности, для совершенствования методов автоматического реферирования научных публикаций. Так, для реферата выбираются предложения с учетом их места в риторической структуре вводной части статьи как одной из наиболее важных зон аргументации [21, 22, 23 и др.]. Отметим, что в этих работах не рассматривается операциональная организация целого текста

как детерминированная познавательным процессом, а устанавливается состав только тех действий, совершая которые автор-ученый стремится наилучшим образом представить полученные им результаты.

Основная сложность создания программного обеспечения автоматического распознавания познавательных коммуникативных действий в тексте состоит в необходимости охвата максимально широкого набора разноуровневых языковых признаков (морфологических, синтаксических, семантических). Использование моделей типа BERT [24], активно применяющихся в настоящее время и позволяющих в неявном виде учитывать такие признаки, представляется нецелесообразным ввиду небольшого размера имеющихся размеченных корпусов, недостаточного для надежного обучения подобных классификаторов. Модели, оперирующие с признаковым пространством большой размерности, необходимым для явного задания упомянутых признаков, в нашем случае также не являются результативными [11].

Распознавание когнитивных действий потребовало от нас последовательного решения ряда задач: извлечения лингвистических признаков простых предложений, являющихся «носителями» этих действий; применения шаблонов для формирования набора высокоуровневых признаков малой размерности; обучения классификатора последовательностей на объектах, описываемых этими признаками; применения шаблонов для устранения неоднозначности в клаузах с несколькими метками классов. Эти исследовательские процедуры объединены в три этапа распознавания операций. Схема предлагаемого метода представлена на рис. 1.

На первом этапе был выполнен автоматический лингвистический анализ текста. Для этого применялся морфологический анализатор MyStem [25], синтаксический анализатор



Рис. 1. Схема трехэтапного метода распознавания когнитивных действий в научных текстах.

Fig. 1. Diagram of a three-stage method for recognizing cognitive actions in scientific texts.

Примечание: НСС – неоднородные семантические сети.



UDPipe [26] (анализатор SyntaxNet [27], применявшийся авторами в ранее проведенных экспериментах [12], не использовался ввиду нестабильной работы его программной реализации). После этого выполнялся реляционно-ситуационный анализ высказываний и на основе всей извлеченной информации строились неоднородные семантические сети (НСС), для чего использовался семантический анализатор [28]. НСС включают вершины (синтаксемы) и связи между ними, отражающие синтаксические и семантические отношения между вершинами [29].

На втором этапе НСС высказываний сопоставляются с контекстно-свободными шаблонами, содержащими формальное описание разноуровневых лингвистических признаков реализаций когнитивных операций. Каждый шаблон описывает множество НСС, соответствующих группе высказываний, содержащих определенную когнитивную операцию. Всего для выявления ментальных действий двух речевых жанров создано 76 шаблонов. Результатом сравнения простого предложения со списком шаблонов является бинарный вектор, каждый элемент которого принимает значение «1», если клауза соответствует шаблону, и «0», если не соответствует.

На третьем этапе для уточнения характера когнитивных операций применялся метод классификации последовательностей. В качестве объектов классификации использовались простые предложения. Так как при продуцировании одного простого предложения может осуществляться несколько когнитивных действий, решалась задача многозначной классификации (multi-label classification). Была проведена экспериментальная оценка применимости некоторых методов классификации последовательностей. Ввиду небольшого размера размеченного корпуса использовались относительно простые методы классификации (за исключением рекуррентных нейронных сетей):

1. Композиции деревьев решений со скользящим окном — такие, как случайный лес решающих деревьев (Random Forest [30]) и градиентный бустинг (использовалась реализация из библиотеки XGBoost [31]). Размер скользящего окна, подобранный эмпирически, составил 3. Большой размер окна приводил к переобучению классификатора, меньший — к снижению точности выявления категорий операций на «золотом» корпусе.

2. Классификатор, основанный на условных случайных полях (Conditional Random Fields

[32].

Эти два метода не предназначены для многозначной классификации, поэтому исходная задача была сведена к обучению нескольких бинарных классификаторов, по одному для выявления каждой категории.

3. Рекуррентные нейронные сети LSTM (Long-Short Term Memory) [33] и GRU (Gated Recurrent Unit) [34] со слоем внимания (attention layer). Применялись модели с небольшим количеством параметров (до 16 ячеек в рекуррентном слое) с дропаутом (dropout) [35], одно- и двунаправленным рекуррентным слоем.

Обучение второго и третьего классификаторов проводилось на последовательностях длиной до 10 предложений, поэтому более длинные текстовые фрагменты разбивались на пересекающиеся блоки.

Гиперпараметры для указанных выше методов подбирались с помощью перебора по сетке, на основе оценок качества, полученных в ходе статистической процедуры пятикратного перекрестного скользящего контроля (cross-validation [36]). Для реализации классификаторов использовались библиотеки Scikit-learn [37], TensorFlow [38] и CRFSuite [39].

Отметим, что предусмотренная третьим этапом предложенного метода задача устранения неоднозначности в клаузах с несколькими метками классов будет решаться в последующих исследованиях.

Перейдем к анализу текстов.

### Речевая системность исследуемых жанров

В рамках небольшой статьи невозможно описать, пусть даже предельно кратко, все формальные речевые маркеры анализируемых жанров научного текста, поэтому мы охарактеризуем здесь лишь типологически наиболее существенные, наиболее важные с точки зрения задач автоматической идентификации жанров.

I. Тексты, основная коммуникативная цель которых заключается в *описании нового для науки явления*, относятся к генетически первичному жанру научной речи, с которого, в принципе, начинается познавательно-коммуникативная деятельность. В то же время они обязательно содержат и такие формы, которые объективируют дальнейшие этапы научного осмысления мира.

Речевые особенности жанра в целом определяются признаками тех ментально-речевых операций (или категорий), которые и образуют в совокупности данное сложное познавательно-коммуникативное действие, а именно:

1) систематическое описание свойств нового объекта, 2) описание комплекса наиболее важных дифференциальных признаков нового объекта (диагноз), 3) определение места нового объекта в системе известных явлений, 4) Описание местоположения и распространения нового явления.

Содержание высказываний, реализующих первую из названных операций, включает описание качественных (1.1.) и количественных (1.2.) характеристик нового объекта, а также его состава и строения (1.3.).

1.1. Описание качественных характеристик нового объекта объективируется чаще всего в структурных схемах с именным сказуемым и нулевой связкой, в которых позиция подлежащего замещена предметным или, реже, признаковым существительным (отглагольными или отадекватными номинализациями), а именная часть выражена качественным (кратким или полным) либо относительным прилагательным или причастием в краткой (для страдательных причастий) или полной форме в им. п.; например: *Излом неровный; Блеск стеклянный; Тело слегка уплощенное; Зерна минерала прозрачны; Спереди тело закруглено и сужено.*

Нередко сказуемое выражается синтаксемой «творительный падеж с предлогом с» или «родительный падеж с предлогом без» (предмет как целое характеризуется по части, детали): *Листья с округлым основанием, с острой верхушкой; Голова с несколько размытыми серо-коричневыми пятнами. Грудные тергиты без макрохет.*

В число формальных маркеров входят номинации параметров описания ( $N^{\text{параметр}}$ ): *состав, строение, форма, часть, основание, поверхность, сторона, край, вид, цвет, длина, запах* и т.п. (часто в соединении с глаголом *иметь*), производные прилагательные с формантами *-видный, -образный, -подобный*, сложные прилагательные с дефисным написанием, характеризующие цветовые оттенки или особенности формы: *розово-желтый, бледно-фиолетовый, продолговато-округлый, удлинено-таблитчатый*; нередко в этой функции высказывания с предикатными словами (*не*) *обнаружен, (не) виден, (не) выражен*. Ср.: *Листья яйцевидной формы; Листья яйцевидные; Порошок минерала имеет вид бесцветных, немного удлинённых пластинок; Они имеют форму конических трубок; Соматические щетинки не обнаружены; Спайкость или отдельность не выражены.*

Иногда внешний вид (форма) объекта описывается в структурных схемах с  $N^1$  в «правой» (присвязочной) части; ср.: *Плод — орешек; Кристаллы стеклита — гексагональные*

*пластинки.*

1.2. Описание количественных характеристик объекта осуществляется чаще всего посредством предложений с  $N^{\text{параметр}}$  в позиции подлежащего, с нулевой связкой или с предикатами «параметрической характеристики объекта» ( $R^{\text{параметр}}$ ) типа *равен, составлять, (из)меняться, варьировать, колебаться (в пределах..., от до), расти, увеличиваться, уменьшаться, приближаться, стремиться к..., достигать* и др., обычно в сочетании с  $N^{\text{параметр}}$  (*длина, вес, сила и т.п.*) и числовыми выражениями; ср.: *Твердость 4-5; Высота шипиков — 1,5-1,7 мкм; Плотность 2.792 г/см.*

1.3. Описание состава и строения (взаимосвязи частей, местоположения одной части относительно другой) нового объекта реализуется:

а) в моделях с партитивными или посессивными предикатами (*состоять, включать, образовать, входить; иметь, содержать*): *Эпителий шишки состоит из ацидофильных железистых и ресничных клеток; Клетки образуют нитевидные колонии;*

б) в структурных схемах с бытийным предикатом (*быть, иметься, присутствовать, отсутствовать, возникать; наблюдаться, встречаться, обнаруживаться, фиксироваться; виден, замечен*), в начале которых, как правило, имеется детерминант-локатив: *В периферической части створки имеются простые или сложные камеры-альвеолы; На гранях пинакоида наблюдается штриховка; Спиральные утолщения на стенках сосудов отсутствуют; В кристаллах заметны грани призм. Нередко видна грубая штриховка;*

в) в моделях с локативными предикатами типа *находиться, располагаться, пролегать* и т.п., выраженными в том числе метафорами от глаголов движения, перемещения (*идти* в значении «пролегать, быть расположенным где-н., каким-н. образом», *нести* в значении «поддерживать собой что-л.», *отходить* в значении «отклоняться от главного направления», *проходить* в значении «простираться, иметь какое-л. направление», *разворачиваться* в значении «растягиваться в длину»: *Мускульные волокна идут от продольной мускулатуры стенки тела к хоботу; От кинетосом отходят три ленты; Лента Л2 проходит близко от ленты Л3; В основании стебель несет светлые листья; Блоки из зигзагообразных цепочек разворачиваются вокруг осей; Створки ареолы и поры расположены беспорядочно; Глаза расположены в два ряда.*

2. Описание нового объекта в указанных



трех аспектах, как правило, реализуется в крупном фрагменте текста, образованном однотипными предложениями. Подлежащее этих предложений называет новый для науки объект (растение, животное, минерал) и его части (в случае растений и животных – органы), а сказуемое – признаки этого объекта. В результате создается целостное (интегральное) представление о вновь открытом явлении. Однако уже в ходе решения задачи описания интегральных признаков нового объекта может быть поставлена задача выявления его *основных дифференциальных признаков*, когда автор сразу стремится приблизить описание объекта к его диагнозу.

В собственно языковом отношении тексты, описывающие новый объект, по-видимому, мало чем отличаются от текстов описания как такового: различия лежат в иллокутивной (интенциональной) плоскости. Обычно этой категории предшествует заголовок «Диагноз», который встречается также в текстах, реализующих третью из описываемых в этом разделе ментально-речевых категорий – *определение «места» нового объекта в системе известных явлений* (поэтому его, как и многие другие речевые маркеры, нельзя считать показателем исключительно второй или третьей категории). Больше того, система средств, используемых в этих случаях, в целом совпадает с речевыми маркерами текстов, реализующих аналогичный этап классификационной деятельности ученого (см. ниже).

3. *Определение «места» нового объекта в системе известных явлений.* Как уже было отмечено, текстам, реализующим эту категорию, часто предшествуют заголовки, безусловно, являющиеся речевым показателем данной категории. Она регулярно актуализируется с помощью клишированных конструкций с реляционными предикатами (прежде всего компаративными, партитивными и посессивными), широко используемыми также в текстах, представляющих результаты классификационной деятельности (см. ниже). Ср.: *Allopsontus kabaki sp.n. относится к подроду Allopsontus; Ювенильные стадии нового вида отличаются от таковых стадий variatus следующими признаками... От последнего он существенно отличается по химическому составу; Бустамит и мандигит различаются между собой катионом: Наличие фтора сближает арангасит с митриваитом; Новый вид близок к Ph. Grandis; Новый вид сходен с представителями родов Argenna; Mixacarus имеет сходство только с hatanni; Проявляет*

*сходство с Salix ovalifolia* (в последних двух примерах использованы описательные формы предикатов).

В полипредикатной конструкции реляционный предикат принимает обычно форму номинализации (*различие, отличие, особенность, характерный (отличительный) признак, сходство, близость, тождество* и др.), включаемой в синтаксическую структуру с помощью компликативных глаголов *состоять, заключаться, корениться* (в чем, в том, что...): *Главное кристаллохимическое различие между эпистолитом и звягинитом состоит в упорядоченном замещении...*

В числе других средств реализации этой категории регулярно используются закрытые (бинарные) сочинительные конструкции с противительным-сопоставительным союзом *а* в сочетании с частицей *не* (*Соцветие большей частью зонтиковидное, а не раскидистое; Колоски 1-2,7 см. дл., а не 8-15 мм. дл.*), с союзом *тогда как* (*В ферробустамите Fe<sup>2+</sup> доминирует в малом М3-октаэдре, тогда как в остальных М-позициях преобладает кальций*) и др. Отметим компаративный оборот с предлогом *в отличие от* в квалификативной конструкции с предикатами *свойствен, характерен, присущ* и др., формы сравнительной степени прилагательных или наречий (подробней речевые маркеры категории описаны ниже, в разделе, посвященном реализации аналогичной категории классификационного текста).

4. Завершается описание нового объекта его пространственной характеристикой. В соответствующих фрагментах научного текста регулярно используются предложения с локативными и бытийными предикатами (*находиться, располагаться, иметь распространение; обитать, произрастать, водиться*), а также небольшой группой ментальных предикатов (*отметить, зарегистрировать, зафиксировать, приурочить*) и близких по семантике глаголов «перцептивного проявления» (*найти, обнаружить, встретить*, как правило, в пассивном залоге, с обязательной локативной синтаксемой. Ср.: *Новый минерал урамарсит обнаружен в зоне месторождения Бота-Бурум; На Кольском полуострове был найден образец с минералом; Аналогичная фаза впервые была зафиксирована в вулканических эксгалациях.*

К характерным способам реализации этой категории стоит отнести номинации *местонахождение, местообитание, распространение* в роли свободной синтаксемы (заголовка), за которой следует географическое название в им. п. или локатив с географическим термином (*лес,*

*море, озеро, опушка, долина* и т.п.).

В познавательном процессе операция 1 (описание свойств нового объекта) является предпосылкой операции 2 (описания комплекса дифференциальных признаков нового объекта), а операция 2, в свою очередь, — предпосылкой операции 3 (определения места нового объекта в системе известных явлений). Однако при изложении результатов проведенного исследования ученый совсем не обязательно следует собственно познавательной логике, поэтому последовательность категорий в тексте может быть разной.

II. **К классификационному речевому жанру** отнесем такой текст или его фрагмент, который имеет основной (или даже единственной) коммуникативной целью представление результатов особого рода ментальной деятельности — *разграничения неопределенного множества объектов или явлений, образующих по некоторому постулируемому в эксплицитной формулировке или подразумеваемому (очевидному) признаку нечто общее, целостное в представлениях*. Эта деятельность опирается на предшествующие ментально-речевые действия как эмпирического этапа научного познания мира, представляемые в соответствующих жанрах эмпирического научного текста (наблюдение и описание нового объекта, сравнение с уже известными и описанными и др.), так и более высокого уровня познавательного процесса — теоретического, превращающего собственно классификационную (аналитическую) деятельность ученого в органическую часть построения типологии и синтеза системы.

В классификационном тексте в разных пропорциях должны быть вербально представлены следующие речемыслительные действия (категории): 1) *экспликация понятия об исходном множестве объектов*; 2) *изложение принципов классификации* (критериев, оснований деления); 3) *обозначение* (перечисление) *выделенных классов* (родов, видов и разновидностей); 4) *диагноз* — *описание дифференциальных признаков* (или одного признака) *выделенных классов*. Репрезентирующий этапы этой деятельности текст образует поэтому устойчивую (хоть и варьирующую в некоторых пределах) речевую структуру.

Каждому из этих четырех действий могут соответствовать как автономные текстовые единицы, так и их несамостоятельные фрагменты. Например, первая категория может быть объективирована в отдельных высказываниях, содержащих **определение** классифицируемого множества, однако нередки и непредикатив-

ные способы экспликации представления об исходном множестве — в том числе «методом языковой категоризации мира», т.е. с помощью узувального или узко-терминологического его обозначения.

1. *Экспликация понятия о классифицируемом множестве объектов*, как правило, имеет результатом краткое вербальное определение этого понятия, которое может быть представлено в ряде стереотипных структурных моделей с предикатными словами и их аргументами из специальных списков ключевых слов (например,  $N_{\text{класс}}$ , или номинации для обозначения результатов классификационной деятельности ученого: *класс, подкласс, тип, подтип, род, жанр, субжанр, разновидность, вид, подвид, группа, подгруппа*;  $R_{\text{класс}}$ , или предикаты «классификации» (в т.ч. партитивные и посессивные): *классифицировать(ся), различать(ся), группировать(ся), делить(ся), разделять(ся), разбивать(ся), членить(ся), расчленять(ся); состоять из..., входить в ...*).

В числе этих моделей отметим, прежде всего, структурные схемы с именным сказуемым, выраженным сущ. в им. или тв. п. и глагольными связками (*это [есть], [это] и есть, являться, представлять собой*) например: *Концепт — дискретная содержательная единица коллективного сознания или идеального мира, хранимая в национальной памяти носителя языка в вербально обозначенном виде...; Фундаментальные компоненты значения в языке — это функциональные компоненты... Именные компоненты структурной схемы заполняются включенными в списки ключевых слов номинациями наиболее общих понятий (*единица, элемент, объект, материал, конструкция, конструкт, компонент, часть, образование; категория; вещество, минерал, личность (человек/люди), существо, организм* и др.), существительными из списка  $N_{\text{класс}}$ , а также номинациями-терминами ( $N_{\text{терм}}$ ) из открытого списка, формальным признаком которых могут быть словообразовательные маркеры — префиксы *алло-, квази-, полу-* (*подлещик, алломорф, аллофон, аллолекс, квазиподлежащее*), суффиксы *-ит-* (*кварцит, люизит*), *ант-* (*дезодорант, антидепрессант*), *-о-ид-* (*пластиды, циклоиды, шизоиды, суффиксоиды...*), *-ин-* (*клофелин, аспирин*), *-ем-* (*синтаксема, таксема*), *-а-ци-я-, а/ени-е-, -изм-, -ость-, -ств-о* и др.*

Распространены в этой функции предложения с предикатами именования (*обозначить, назвать*) и интерпретации (*определить, понимать, трактовать, интерпретировать* и др.); первые реализуются в схемах с тв. п. «левого» аргумента, вторые

конструкциями с союзом-связкой **как** и вин. п. сущ. (например: *Шизоидами или циклоидами мы называем колеблющиеся между здоровьем и болезнью патологические личности... Он определяет концепт как «условно-ментальную единицу»*).

Классифицируемое множество объектов может быть определено через их функциональную характеристику ( $N^1$  *служит, функционирует, выступает, существует – для чего, чем, в качестве чего...*), например: *Способами представления знаний в концептосфере служат такие когнитивные структуры, как концепты, сценарии, скрипты, фреймы, схемы, гештальты и т.д.* В этом случае формальными маркерами могут быть сущ. из группы «эпистемических» имен – такие, как *понятие, представление, категория, концепт ...* (в соединении с согласованным или несогласованным определением: *понятие морфемы введено для...; представление о жанре появилось... (Первоначально понятие макиавеллизма было введено Христи и Гейсом для обозначения такой характеристики, как направленность человека на контроль других)*, а также производные предлоги, выражающие отношение назначения, функции (*для, с целью, в качестве, в роли, на роль, в функции* и др.). Ср.: *В качестве таксономических единиц предлагается использовать иерархическую последовательность – водоносная система, водоносный комплекс, водоносный (водоупорный) горизонт [= ‘водоносная система, водоносный комплекс... – это таксономические единицы...’]*.

Если классифицируется «непредметное» множество (признаков, явлений, процессов, обозначенных, как правило, номинациями с деривационными маркерами типа *-а-ци-я-, а/ени-е, -изм, -ость, -ств-о*), то его определение может включать указание на способ, характер проявления признаков, протекания процесса; например: *Дезактуализация предложения проявляется в виде по меньшей мере трех типов процессов: 1) номинализации, 2) атрибутивации, 3) перевода вопроса в новый контекст.*

В качестве формальных речевых показателей данной ментально-речевой категории могут использоваться вводные обороты типа *как считает И.И. Иванов, по мнению И.И. Петрова, вслед за И.И. Сидоровым и т.п.: Дискурс мы, вслед за А. А. Кибриком, понимаем как единство двух сущностей – процесса языковой коммуникации и получающегося в ее результате объекта, то есть текста.*

Весьма часто для вербальной экспликации исходного множества объектов используется номинативный способ – обозначение этого

множества существительным-термином или терминологическим словосочетанием, употребленным в независимой синтаксической позиции (в качестве заголовка текста или какого-либо его раздела). Нередко номинативный способ используется в рамках высказывания (обычно в теме высказывания), реализующего какие-либо другие категории классификационной деятельности. Ср., например, высказывание *Литолого-фациальный анализ пермских отложений даёт возможность выделения крупных трансгрессивно-регрессивных этапов, отвечающих по времени эпохам*, порожденное для реализации второй категории – выбора критериев классификации. Этот способ экспликации представления о классифицируемом множестве весьма часто проявляется в заголовках.

2. Вторая категория речемыслительных действий, образующих классификационную деятельность (*выбор оснований классификации*), реализуется в стереотипных по структуре и семантике полипредикатных высказываниях с  $R_{\text{класс}}$  (*делить, классифицировать, группировать*, включая аналитические варианты таких предикатов с провербами *проводить, осуществлять, производить, строить* и др.) и компликативными предикатами типа *основываться на..., базироваться на..., опираться на..., учитывать, исходить из..., принимать во внимание, принимать за основу* и т.п. Чаще всего избираются «имплицитно-агентивные» конструкции страдательного залога (*Что касается изъяснительных конструкций, то они могут быть разбиты на группы по семантике опорного слова и способу подчинения*). В роли объектного актанта, как правило, выступают отмеченные выше номинации ( $N_{\text{класс}}$ ,  $N_{\text{терм}}$ ). К формальным маркерам должны быть отнесены также предлог *по* (с дат. п. непредметных имен – *признак, свойство, характеристика* и др.) и довольно большая группа новообразованных предложных сочетаний типа *на основе, на базе, с учетом, в соответствии с, с опорой на...; в ... аспекте / отношении / плане, с ... точки зрения, под ... углом зрения* (с согласованным определением перед существительным); *по аналогии с...* и др. Перечисленные предложные группы сочетаются с признаковыми именами и номинализациями глагольных или именных предикатов, в том числе с так называемыми неполными номинализациями, формальными приметами которых является гипотактический показатель *то (тот факт), что... / то, как...* (Т-местоимение может сочетаться с любым из перечисленных здесь квазипредлогов). Кроме того, в число марке-



ров описываемой категории следует включить «аспектуальные» наречия (точнее, их суффиксы) типа *синтаксически, исторически, геологически, структурно* и т.д.

3. Речевые маркеры третьей категории (*перечисление выделенных классов*) играют важнейшую роль с точки зрения задач автоматической идентификации классификационных текстов. Дело в том, что речевые маркеры прочих категорий классификационного текста встречаются также в текстах многих других жанров научного стиля. Так, экспликация представления о классифицируемом множестве объектов или явлений мало чем отличается от любого определения понятия в научном тексте; речевые маркеры текстов, излагающих критерии классификации, во многом совпадают с показателями аргументации, обоснования выдвигаемых теоретических положений и выводов, поэтому они, как и речевые признаки текстов, описывающих отличительные особенности чего-либо, встречаются также во фрагментах тех научных текстов, общая (основная) коммуникативная установка которых не связана с намерением представить результаты классификационной деятельности ученого. Идентификация неклассификационного текста может быть осуществлена, очевидно, с учетом контекста (в классификационном тексте категории представлены обычно в совокупности) и по отсутствию маркеров третьей категории – *перечисления выделяемых классов*.

Данная категория реализуется в предикативных или полупредикативных конструкциях с  $R_{\text{класс}}$  (*делить, классифицировать, группировать, отметить, выявить*) или с предикатами речемыслительного и коммуникативного плана (*описать, обозначить, перечислить, назвать, зарегистрировать, изобразить*) и рядом однородных именных компонентов – номинаций выделенных (терминальных) классов, предваряемых обобщающей номинацией  $N_{\text{класс}}$ , как правило, с согласованным или несогласованным определением. Однородному ряду обычно предшествуют «маркеры перечисления»: местоименный компонент (*такой, следующий, такой как...*), а также вводное слово *например* или присоединительные (уточняющие) коннекторы *а именно, то есть*, в том числе «голое» двоеточие перед перечислительным рядом. Ср.: *В пределах трех гидрогеологических этажей выделяются пять структурно-гидрогеологических комплексов: антропоген-олигоценый, палеогеновый, верхнемеловой, нижнемеловой и палеозойский.* В группу языковых средств, реализующих данную категорию, следует вклю-

чить также посессивные и партитивные предикаты, в своем первичном значении обозначающие отношения принадлежности и целого и части, например,  $[N^1 \text{ (род как целое)}]$  *состоять (складываться, образовываться) из*  $[N^2 \text{ (вид как часть), } N^2 \text{ ...}]$ ;  $[N^1]$  *включать (в состав)*  $[N^4, N^4 \text{ ...}]$  и др.; ср.: *Все виды рыб, обитающие на данном участке р. Валы, по классификации Никольского относятся к 2 пресноводным фаунистическим комплексам...: понто-каспийский комплекс представлен голавлем, уклейей и быстряжкой, бореальный равнинный комплекс – плотвой, щукой, щиповкой.*

Специфическим маркером категории выступает также связочная конструкция с модальным глаголом *мочь* в форме наст. вр. (выражающая не столько модальность возможности, сколько отношение чередования чего-либо в пространстве / времени). Ср.: Риторические отношения *могут быть также симметричными и асимметричными.*

Весьма надежным маркером является нумерация выделяемых классов (арабские и римские цифры в комбинации с пунктуационными знаками, литеры кириллического или латинского алфавита и др.).

С высокой частотностью высказывание, воплощающее данную категорию, включает в своем начале неприсловный определитель (детерминант) с локативным, темпоральным, ситуативным или каузальным значением (*На территории района выделено три гидрогеологических этажа...*).

Как уже было отмечено, для данной категории характерно наличие рядов однородных членов ( $N^{\text{терм}}$ ), для которых типичен синтаксический и лексический параллелизм, реализуемый в пределах одного высказывания, сверхфразового единства, когда каждый из выделяемых классов оказывается темой («левым» аргументом предиката) автономного высказывания Ср., например: *На территории района выделено три гидрогеологических этажа. Верхний этаж, включающий грунтовые водоносные горизонты ... Средний этаж охватывает подземные воды верхнемеловых отложений ... Нижний этаж включает водоносные горизонты нижнемеловых отложений...* Кроме того, структурный параллелизм может быть реализован как объединение параллельных (однотипных по структуре) СФЕ в пределах целого текста (главы, параграфа) – когда отдельному выделенному классу посвящен особый структурно-композиционный раздел, как правило, снабженный порядковым номером и озаглавленный именем класса. Нередко обозначение



(перечисление) видов реализуется в виде таблиц, в столбцах которых обычно перечисляются выделяемые виды и разновидности, а также описываются их отличительные признаки.

4. Типовые средства четвертой категории (описания дифференциальных признаков выделенных классов), помимо указанных выше ключевых номинаций и предикатов, включают: номинативные единицы с общим значением 'признак' – такие, как *признак, свойство, черта, характеристика, особенность, отличие*; качественные и относительные прилагательные из специального списка ( $Adj_{интенс}$ : *яркий, четкий, резкий; средний, нейтральный, главный, важный* и др., включая однокорневые прилагательные с приставками *не-, а-*), обозначающие степень проявления признака;  $N_{параметр}$  в соединении с реляционными предикатами типа *равняться, составлять*; предикатные слова речемыслительной и коммуникативной семантики ( $R_{мент}$ : *описывать, характеризовать, анализировать* и др.); обобщенные предикатные слова с общим значением 'иметь свойство, признак' ( $R_{призн}$ ): *характерный (характеризоваться), специфичный, свойственный (свойствен, свойство), носить (иметь) характер, иметь (обладать)*; реляционные предикаты ( $R_{реляц}$ ) для обозначения тождества/подобия/различия/контраста: *тождествен, идентичен, равен, близок* и др.; компликативные предикаты *состоять, заключаться, корениться в..., определяться, обуславливаться, связан с...* и т.п.; бытийные предикаты с локативной (*в/на  $N^6$  / среди  $N^2$* ) или ситуативной (*при  $N^6$ , в условиях /в ходе  $N^2$* ) синтаксемой; реляционные предлоги *в отличие от..., по сравнению с..., сравнительно с... относительно..., по отношению к..., применительно к..., в сопоставлении с...*; наречия меры и степени: *значительно, резко, скачком, существенно, постепенно, явно, четко, преимущественно, (не) достаточно, несколько...*; прилагательные или наречия в сравнительной степени: *(наи)более / (наи)менее; наибольший/наименьший, более/менее/сравнительно/относительно +Adj* (в соединении с  $R_{призн}$  или  $R_{реляц}$ ).

Ярким формальным признаком рассматриваемой категории является распространение номинативных компонентов моделей из списков ключевых номинаций ( $N_{класс}$ ,  $N_{терм}$  и др.) согласованными с ними оборотами с прилагательными и причастиями, относительными предложениями с местоимениями *который*, несогласованными определениями, а также наличие конструкций с относительными придаточными (типа *то, что..., тот, кто...*) или компликативных предикатов, подчиняющих

*что-придаточные*; ср.: *Конверсия. Данный тип синтаксического преобразования состоит в том, что главная и зависимая клаузы диаметрально меняются позициями; По тектоническим нарушениям фундамента воды зоны гидрометагенеза взаимодействуют с водами нижней части осадочного чехла, которые характеризуются повышенной газонасыщенностью гелием, водородом и др.*

Частотными в текстовых фрагментах данной категории являются сравнительные и сопоставительные конструкции: предложения с собственно компаративными маркерами (предикатами *сравнивать, сопоставлять, соотносить*; прилагательными и наречиями в сравнительной и превосходной степени и придаточными с компаративным союзом *чем*); бинарные сложные предложения сопоставительной семантики с союзами *а, же, между тем, в то же время, тогда как, в то время как, между тем как, чем..., тем...*; бинарные структуры с вводными словами *напротив, наоборот* во второй предикативной конструкции.

Как и в случае с объективацией третьей категории классификационной деятельности, диагнозы выделяемых родов и видов могут быть вербализованы в параллельных (однотипных) конструкциях текста, когда отдельному классу, выделяемому из множества объектов, посвящен особый «однородный член» текста – целый структурно-композиционный раздел (в число маркеров в этом случае входят  $N_{класс}$  /  $N_{терм}$ ). Для таких конструкций текста характерны лексические повторы предикатных слов и однородность синтаксем, представляемых указанными ключевыми номинациями.

Завершая разбор формальных речевых показателей познавательной-коммуникативной деятельности, подчеркнем, что во многих случаях тот или иной маркер не может быть привязан лишь к какой-то одной категории. Так, одна и та же номинация может быть употреблена при экспликации представления об исходном множестве классифицируемых объектов либо как обозначение класса, выделяемого из множества более высокого ранга. Ряд формальных речевых признаков текстов, реализующих *систематическое описание свойств нового объекта*, совпадает с маркерами, обнаруживаемыми в описаниях *дифференциальных признаков нового объекта*, а также – вполне ожидаемо – в аналогичных по интенции фрагментах классификационного текста. Очевидно, что проблема автоматической идентификации текстов, реализующих определенную категорию, должна решаться с

**Таблица 1.**  
**Результаты оценки шаблонов когнитивных операций**

**Cognitive Operations Pattern Evaluation Results**

Код операции	P	R	F1
1.1	0.97	0.95	0.96
1.2	0.83	0.84	0.83
1.3	0.96	0.97	0.96
1.4	0.64	0.85	0.73
2.1	0.79	0.65	0.71
2.2	0.80	0.91	0.85
2.3	0.88	0.94	0.91
2.4	0.64	0.85	0.73

*Примечание.* P – точность, т.е. отношение числа правильно классифицированных речевых фрагментов, реализующих операцию, к общему количеству фрагментов, помеченных как реализующие ее. R – полнота: отношение числа правильно классифицированных речевых фрагментов, реализующих операцию, к общему количеству фрагментов, реализующих ее. F1 – мера, являющаяся средним гармоническим точности и полноты.

учетом максимально широкого круга маркеров – лексических (заданных списками ключевых слов) и грамматических, а также сверхсегментных, графических и пунктуационных.

**Результаты и их обсуждение**

Оценка качества выявления фрагментов текста, реализующих когнитивные операции,

проводилась на корпусе, размеченном вручную. Корпус состоит из более чем 550 небольших фрагментов научных статей (фрагмент включает от одного до десяти предложений, по 10-100 слововхождений в каждом), содержащих реализации определенных когнитивных операций, и около 150 фрагментов, не содержащих реализаций этих операций. Результаты оценки качества приведены в таблице 1.

Необходимо отметить, что причиной относительно низкой точности (P) выявления когнитивных операций является использование контекстно-свободных шаблонов, тогда как к снижению полноты (R) в основном приводят ошибки синтаксического анализатора. В целом полученные оценки говорят о применимости созданных шаблонов для формирования признаков описаний клауз.

Для сравнения методов классификации последовательностей использовался размеченный корпус. Он содержит около 200 фрагментов научных публикаций на русском языке, относящихся к исследованному нами ранее речевому жанру «Сообщение об эмпирической закономерности причинно-следственного типа» [12]. При составлении корпуса использовались статьи по различным научным направлениям, из которых выделялись фрагменты по 10-50 предложений. Из-за несбалансированности категорий в корпусе оценки качества при сравнении считались с микроусреднением (Таблица 2).

Наилучшее качество выявления фрагментов статей, соответствующих когнитивным операциям, было получено с помощью метода на

**Таблица 2.**  
**Результаты оценки методов классификации последовательностей для распознавания когнитивных операций**  
**Evaluation results of sequence classification methods for recognition of cognitive operations**

Классификатор	P	R	F <sub>1</sub> -микро
LSTM+слой внимания	0.59±0.07	0.68±0.08	0.62±0.02
GRU+слой внимания	0.63±0.06	0.65±0.09	0.63±0.05
Случайный лес	<b>0.75±0.09</b>	<b>0.69±0.07</b>	<b>0.71±0.07</b>
Градиентный бустинг	0.72±0.10	0.63±0.07	0.67±0.09
Условные случайные поля	0.70±0.09	0.68±0.05	0.69±0.09

*Примечание.* F1-микро для каждой категории вычисляется как  $2 * (2 * tp + fn + fp)$ , где tp – количество правильно классифицированных фрагментов, принадлежащих категории, fn – количество фрагментов, принадлежащих этой категории, но ошибочно отнесенных классификатором к другим категориям, fp – количество фрагментов, принадлежащих другим категориям, но ошибочно отнесенных классификатором к оцениваемой категории.



основе случайного леса деревьев решений, что может быть связано с относительно небольшим размером размеченного корпуса. Полученные оценки качества ( $F_1 > 0.7$ ) позволяют говорить о возможности применения предложенного метода для решения прикладных задач.

На следующем этапе исследования предполагается сформировать шаблоны для последнего этапа предлагаемого метода – устранения неоднозначности для клауз, помеченных несколькими категориями операций.

### Заключение

Анализ подтверждает исходное положение работы о том, что жанры речи, рассматриваемые как устойчивые формы организации познавательно-коммуникативных процессов, могут служить важным объектом когнитивного моделирования, поскольку поверхностно-речевой план жанра (взаимосвязанные на текстовой плоскости разноуровневые языковые единицы) воплощают внутреннюю (ментальную) деятельность субъекта, направленную на решение той или иной познавательной задачи.

При программировании представленных в лингвистическом описании маркеров познавательно-коммуникативных действий успешно решается задача охвата широкого набора морфологических, синтаксических и лексико-семантических единиц. Предложен метод

распознавания в научных текстах различных когнитивных действий субъекта, предусматривающий: 1) автоматический лингвистический анализ текстов определенного речевого жанра; 2) сопоставление неоднородных семантических сетей высказываний с контекстно-свободными шаблонами, содержащими формальное описание языковых признаков воплощения ментальных действий; 3) уточнение характера этих действий с использованием классификации последовательностей.

Автоматическое распознавание типовых познавательно-коммуникативных действий, образующих устойчивую форму двух речевых жанров – описания нового для науки явления и классификации опытных данных, имеет показатели качества от 0,71 до 0,96 (в среднем по рассмотренным действиям – 0,84), что свидетельствует о применимости созданных шаблонов для формирования признаков описаний предложений. Перспективы исследования связаны с уточнением идентификации познавательно-коммуникативных действий с применением методов классификации последовательностей.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 17-29-07049 офи\_м.

### СПИСОК ЛИТЕРАТУРЫ

1. Салимовский В.А. Жанры речи в функционально-стилистическом освещении (научный академический текст). Пермь: Изд-во Перм. ун-та, 2002.
2. Бахтин М.М. Под маской. Маска третья. Волошинов В.Н. Марксизм и философия языка. М.: Лабиринт, 1993.
3. Бахтин М.М. Проблема речевых жанров // Он же. Эстетика словесного творчества. М.: Искусство, 1979. С. 237-280.
4. Винокур Г.О. Поэтика. Лингвистика. Социология // Винокур Г.О. Филологические исследования. Лингвистика и поэтика. М.: Наука, 1990. С. 22-31.
5. Кожина М.Н. Речеведение. Теория функциональной стилистики. М.: Флинта: Наука, 2014.
6. Дементьев В.В. Теория речевых жанров. М.: Знак, 2010.
7. Седов К.Ф. Общая и антропоцентрическая лингвистика. М.: Изд. дом ЯСК, 2016.
8. Долинин К.А. Речевые жанры как средство организации социального взаимодействия // Жанры речи. Саратов: 1999. Вып. 2. С. 7-13.
9. Swales J.M. Genre analysis: English in academic and research settings // Cambridge University Press, 1990.
10. Осипов Г.С., Девяткин Д.А., Кузнецова Ю.М., Швец А.В. Возможности интеллектуального анализа научных текстов на основе построения их когнитивной модели // Искусственный интеллект и принятие решений. 2018. №1. С. 41-52.
11. Девяткин Д.А., Кузнецова Ю.М., Чудова Н.В. Методы автоматического выявления ментальных действий в текстах научных публикаций. Часть 1 // Искусственный интеллект и принятие решений. 2018. №2. С. 36-46. DOI: 10.14357/2071859418020.
12. Девяткин Д.А., Каджая Л.А., Салимовский В.А. Жанры речи как объект компьютерного анализа (на материале научных текстов) // Жанры речи. 2019. № 2 (22). С. 86-104. DOI: 10.18500/2311-0740-2019-2-22-86-104.
13. Салимовский В.А., Девяткин Д.А., Станкевич М.А. Речевая системность как опосредующее звено машинного анализа познавательной

- деятельности (на материале научных текстов) // Теория речевой деятельности: вызовы современности. Материалы XIX Международного симпозиума по психолингвистике и теории коммуникации. М.: Изд-во Канцлер, 2019. С. 228-229.
14. **Леонтьев А.Н.** Деятельность. Сознание. Личность. М.: Смысл; Академия, 2004.
15. **Майданов А.С.** Методология научного творчества. М.: ЛКИ, 2008.
16. **Мамчур Е.А., Овчинников Н.Ф., Огурцов А.П.** Отечественная философия науки: предварительные итоги. М.: Российская политическая энциклопедия, 1997.
17. **Каджая Л.А.** Системность научного специализированного дискурса: использование жанровых форм и языковых средств (на материале русских и английских гидрогеологических текстов) / Перм. гос. нац. исслед. ун-т. Пермь, 2016.
18. **Чернявская В.Е.** Научный дискурс: движение результата как коммуникативная и языковая проблема. М.: Ленанд, 2017.
19. **Беляева Л.Н., Чернявская В.Е.** Научный и технический текст и информация 4.0: ключевые задачи при создании структурированного контента // Научно-технические ведомости СПбГПУ. Гуманитарные и общественные науки Т. 10, № 2, 2019. С.53-63. DOI: 10.18721/JHSS.10205
20. **Swales J.M.** Genre analysis: English in academic and research settings // Cambridge University Press, 1990.
21. **Teufel S., Carletta J, Viens M.** An annotation scheme for discourse-level argumentation in research articles // Proceedings of EACL'99: Ninth Conference of the European Chapter of the Association for Computational Linguistics, 8-12 June 1999, University of Bergen, Norway. P. 110-117.
22. **Heffernan K., Teufel S.** Identifying problems and solutions in scientific text // Scientometrics, 2018. V. 116. Issue 2. P. 1367–1382. <https://doi.org/10.1007/s11192-018-2718-6>.
23. **Kirschner C., Eckle-Kohler J., Gurevych I.** Linking the Thoughts: Analysis of Argumentation Structures in Scientific Publications // Proceedings of the 2nd Workshop on Argumentation Mining, Denver, Colorado, 2015. Association for Computational Linguistics. P. 1–11.
24. **Devlin J. et al.** Bert: Pre-training of deep bidirectional transformers for language understanding //arXiv preprint arXiv:1810.04805. 2018.
25. **Segalovich I., Titov V.** «1997-2014». MyStem [Software].
26. **Straka M., Hajic J., Straková J.** UDPipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, pos tagging and parsing //Proceedings of the tenth international conference on language resources and evaluation (LREC 2016). 2016. P. 4290-4297.
27. **Andor D., Alberti C., Weiss D., et al.** Globally normalized transition-based neural networks. // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany, Association for Computational Linguistics. 2016. P. 2442–2452.
28. **Shelmanov A., Devyatkin D.** Semantic Role Labeling with Neural Networks for Texts in Russian// Annual International Conference "Dialog". Moscow: Russian State University for Humanities, 2017. Vol.1. P. 245-257.
29. **Osipov G.S., Smirnov I.V., Tikhomirov I.A.** Relational-situational method for text search and analysis and its applications //Scientific and Technical Information Processing. 2010. Vol. 37(6). P. 432-437.
30. **Breiman L.** Random forests. Machine learning. 1(45), 5-32. (2001).
31. **Friedman J. H.** Greedy function approximation: a gradient boosting machine. Annals of statistics. 1189-1232 (2001).
32. **Lafferty J., McCallum A., Pereira F.C.N.** Conditional random fields: Probabilistic models for segmenting and labeling sequence data. (2001).
33. **Hochreiter S., Schmidhuber J.** Long short-term memory. Neural computation. 8(9), 1735-1780 (1997).
34. **Chung J. et al.** Empirical evaluation of recurrent neural networks on-line modeling. ArXiv preprint arXiv: 1412.3555 (2014).
35. **Srivastava N. et al.** Dropout: a simple way to prevent neural networks from overfitting //The journal of machine learning research. 2014. Vol. 15(1) P. 1929-1958.
36. **Efron B., Gong G.** A leisurely look at the bootstrap, the jackknife, and cross-validation //The American Statistician. 1983. Vol. 37. (1). P. 36-48.
37. **Pedregosa F. et al.** Scikit-learn: Machine learning in Python. Journal of machine learning research. No. Oct (12), 2825-2830. (2011).
38. **Abadi M. et al.** Tensorflow: A system for large-scale machine learning. 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), pp. 265-283. (2016).
39. **Okazaki, N.** Crfsuite: a fast implementation of conditional random fields (crfs) (2007).

*Статья поступила в редакцию 5.08.2019*



## REFERENCES

- [1] **V.A. Salimovskiy**, Zhanry rechi v funktsionalno-stilisticheskom osveshchenii (nauchnyy akademicheskiiy tekst) [Speech genres in functional stylistic perspective (scientific text)]. Perm: Izd-vo Perm. un-ta, 2002.
- [2] **M.M. Bakhtin**, Pod maskoy. Maska tretya. Voloshinov V.N. Marksizm i filosofiya yazyka. M.: Labirint, 1993.
- [3] **M.M. Bakhtin**, Problema rechevykh zhanrov, On zhe. Estetika slovesnogo tvorчества. M.: Iskustvo, (1979) 237-280.
- [4] **G.O. Vinokur**, Poetika. Lingvistika. Sociologiya, Vinokur G.O. Filologicheskoye issledovaniya. Lingvistika i poetika. M.: Nauka, (1990) 22-31.
- [5] **M.N. Kozhina**, Rechevedeniye. Teoriya funktsionalnoy stilistiki. [Speech studies. Theory of functional stylistics] M.: Flinta: Nauka, 2014.
- [6] **V.V. Dementyev**, Teoriya rechevykh zhanrov. [Theory of speech genres] M.: Znak, 2010.
- [7] **K.F. Sedov**, Obshchaya i antropotsentricheskaya lingvistika [General and anthropocentric linguistics] M.: Izd. dom YaSK, 2016.
- [8] **K.A. Dolinin**, Rechevye zhanry kak sredstvo organizatsii sotsialnogo vzaimodeystviya, Zhanry rechi [Speech genres as means of organizing social interaction], Zhanry rechi [Speech Genres] Saratov. 2 (1999) 7-13.
- [9] **J.M. Swales**, Genre analysis: English in academic and research settings, Cambridge University Press, 1990.
- [10] **G.S. Osipov, D.A. Devyatkin, Yu.M. Kuznetsova, A.V. Shvets**, Vozmozhnosti intellektualnogo analiza nauchnykh tekstov na osnove postroyeniya ikh kognitivnoy modeli [Possibilities of intellectual analysis of scientific texts on the basis of building their cognitive models], Iskustvennyy intellekt i prinyatiye resheniy. [Artificial intelligence and decision-making]. 1 (2018) 41-52.
- [11] **D.A. Devyatkin, Yu.M. Kuznetsova, N.V. Chudova**, Metody avtomaticheskogo vyyavleniya mentalnykh deystviy v tekstakh nauchnykh publikatsiy. Chast' I [Methods of automatic detection of mental actions in the texts of scientific publications. Part I], Iskustvennyy intellekt i prinyatiye resheniy. [Artificial intelligence and decision-making]. 2 (2018) 36-46. DOI: 10.14357/2071859418020
- [12] **D.A. Devyatkin, L.A. Kadzhaya, V.A. Salimovskiy**, Zhanry rechi kak obyekt kompyuternogo analiza (na materiale nauchnykh tekstov) [Speech genres as an object of computer analysis], Zhanry rechi [Speech genres]. 2 (22) (2019) 86-104. DOI: 10.18500/2311-0740-2019-2-22-86-104
- [13] **V.A. Salimovskiy, D.A. Devyatkin, M.A. Stankevich**, Rechevaya sistemnost' kak oposreduyushcheye zveno mashinnogo analiza poznavatel'noy deyatel'nosti (na materiale nauchnykh tekstov) [Speech system as a mediating link in machine analysis of cognitive activity (based on scientific texts)], Teoriya rechevoy deyatel'nosti: vyzovy sovremen'nosti. Materialy XIX Mezhdunarodnogo simpoziuma po psikholingvistike i teorii kommunikatsii [Theory of speech activity: challenges of the present. Materials of the XIX International Symposium on Psycholinguistics and Theory of Communication] M.: Izd-vo Kantsler, 2019. Pp. 228-229.
- [14] **A.N. Leontyev**, Deyatel'nost'. Soznaniye. Lichnost'. [Activity. Consciousness. Personality] M.: Smysl; Akademiya, 2004.
- [15] **A.S. Maydanov**, Metodologiya nauchnogo tvorчества [Methodology of scientific creativity]. M.: LKI, 2008.
- [16] **Ye.A. Mamchur, N.F. Ovchinnikov, A.P. Ogurtsov**, Otechestvennaya filosofiya nauki: predvaritel'nyye itogi [Native philosophy of science: preliminary results]. M.: Rossiyskaya politicheskaya entsiklopediya, 1997.
- [17] **L.A. Kadzhaya**, Sistemnost' nauchnogo spetsializirovannogo diskursa: ispolzovaniye zhanrovnykh form i yazykovykh sredstv (na materiale russkikh i angliyskikh gidrogeologicheskikh tekstov) [Systematic nature of the scientific specialized discourse: the use of genre forms and linguistic means (on the basis of Russian and English hydrogeological texts)] / Perm. gos. nats. issled. un-t. Perm, 2016.
- [18] **V.Ye. Chernyavskaya**, Nauchnyy diskurs: vydvizheniye rezultata kak kommunikativnaya i yazykovaya problema [Scientific discourse: putting forward the result as a communicative and linguistic problem] M.: Lenand, 2017.
- [19] **L.N. Belyaeva, V.E. Chernyavskaya**, Scientific and Technical Texts in the framework of Information 4.0: content analysis and text synthesis, St. Petersburg State Polytechnical University Journal. Humanities and Social Sciences, 10 (2) (2019) 53-63. DOI: 10.18721/JHSS.10205
- [20] **J.M. Swales**, Genre analysis: English in academic and research settings, Cambridge University Press, 1990.
- [21] **S. Teufel, J. Carletta, M. Viens**, An annotation scheme for discourse-level argumentation in research articles, Proceedings of EACL'99: Ninth Conference of the European Chapter of the Association for Computational Linguistics, 8-12 June 1999, University of Bergen, Norway. Pp. 110-117.
- [22] **K. Heffernan, S. Teufel**, Identifying problems and solutions in scientific text, Scientometrics, 116 (2) (2018) 1367-1382. <https://doi.org/10.1007/>

s11192-018-2718-6

[23] **C. Kirschner, J. Eckle-Kohler, I. Gurevych**, Linking the Thoughts: Analysis of Argumentation Structures in Scientific Publications, Proceedings of the 2nd Workshop on Argumentation Mining, Denver, Colorado, 2015. Association for Computational Linguistics. Pp. 1–11.

[24] **J. Devlin et al.**, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805. 2018.

[25] **I. Segalovich, V. Titov**, «1997-2014». MySystem [Software].

[26] **M. Straka, J. Hajic, J. Straková**, UDPipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, pos tagging and parsing, Proceedings of the tenth international conference on language resources and evaluation (LREC 2016). 2016. Pp. 4290-4297.

[27] **D. Andor, C. Alberti, D. Weiss et al.**, Globally normalized transition-based neural networks, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany, Association for Computational Linguistics. 2016. Pp. 2442–2452.

[28] **A. Shelmanov, D. Devyatkin**, Semantic Role Labeling with Neural Networks for Texts in Russian, Annual International Conference “Dialog”. Moscow: Russian State University for Humanities, 1 (2017) 245-257.

[29] **G.S. Osipov, I.V. Smirnov, I.A. Tikhomirov**, Relational-situational method for text search and

analysis and its applications, Scientific and Technical Information Processing. 37(6) (2010) 432-437.

[30] **L. Breiman**, Random forests. Machine learning. 1(45) (2001) 5-32.

[31] **J.H. Friedman**, Greedy function approximation: a gradient boosting machine. Annals of statistics. (2001) 1189-1232.

[32] **J. Lafferty, A. McCallum, F.C.N. Pereira**, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, (2001).

[33] **S. Hochreiter, J. Schmidhuber**, Long short-term memory. Neural computation. 8(9) (1997) 1735-1780.

[34] **J. Chung et al.**, Empirical evaluation of recurrent neural networks on-line modeling. ArXiv preprint arXiv: 1412.3555, (2014).

[35] **N. Srivastava et al.**, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research. 15(1) (2014) 1929-1958.

[36] **B. Efron, G. Gong**, A leisurely look at the bootstrap, the jackknife, and cross-validation, The American Statistician. 37 (1) (1983) 36-48.

[37] **F. Pedregosa et al.**, Scikit-learn: Machine learning in Python. Journal of machine learning research. Oct (12) (2011) 2825-2830.

[38] **M. Abadi et al.**, Tensorflow: A system for large-scale machine learning. 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), (2016) (2016).

[39] **N. Okazaki**, Crfsuite: a fast implementation of conditional random fields (crfs), (2007).

*Received 5.08.2019*

## СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

**Салимовский Владимир Александрович**

**Salimovsky Vladimir A.**

E-mail: salimovsky@rambler.ru

**Девяткин Дмитрий Алексеевич**

**Devyatkin Dmitry A.**

E-mail: devyatkin@isa.ru

**Каджая Людмила Алексеевна**

**Kadzhaya Lyudmila A.**

E-mail: kadzhaya@psu.ru

**Мишланов Валерий Александрович**

**Mishlanov Valery A.**

E-mail: vmishlanov@yandex.ru