

DOI: <https://doi.org/10.18454/RULB.2024.49.37>

## КОНЦЕПЦИЯ МОДЕЛИРОВАНИЯ СЕМАНТИЧЕСКОЙ РАЗМЕТКИ КОРПУСОВ ТЕКСТОВ НА БАЗЕ СОБЫТИЙНОЙ ОНТОЛОГИИ

Научная статья

Михеев А.А.<sup>1,\*</sup>, Ярыгина Е.С.<sup>2</sup>

<sup>1</sup>ORCID : 0000-0003-1119-6654;

<sup>1,2</sup>Марийский государственный университет, Йошкар-Ола, Российская Федерация

\* Корреспондирующий автор (scurra.42[at]yandex.ru)

### Аннотация

Данная статья носит обзорно-аналитический характер, в работе конституируется подход к созданию нового типа семантической разметки текстов, уточняются ключевые параметры конечной перспективной разработки, определяющие контуры формулируемой методологии. В статье приводится анализ существующих подходов к построению семантической разметки, применимой для корпусов текстов.

Ключевыми параметрами определяются: во-первых, событийность онтологического описания текста; во-вторых, смена фокуса с семантики отдельных высказываний/предложений, как единиц структурообразующих смысловое пространство, к единицам иного порядка – ключевым событиям, их порядку и системе; в-третьих, обратный порядок составления онтологического описания: не от грамматики (синтаксиса) к семантике, а от «выражаемого» к форме выражения.

В заключение авторами определяются некоторые потенциальные перспективы создаваемой системы разметки, а также вероятные ограничения и сложности в разработке.

**Ключевые слова:** семантика, корпусная лингвистика, Национальный корпус русского языка, онтологический подход, моделирование семантической разметки.

## THE CONCEPT OF MODELLING SEMANTIC MARKUP OF TEXT CORPORA ON THE BASIS OF EVENT ONTOLOGY

Research article

Mikheev A.A.<sup>1,\*</sup>, Yarigina Y.S.<sup>2</sup>

<sup>1</sup>ORCID : 0000-0003-1119-6654;

<sup>1,2</sup>Mari State University, Yoshkar-Ola, Russian Federation

\* Corresponding author (scurra.42[at]yandex.ru)

### Abstract

This article is a review-analytical one; it constitutes an approach to the creation of a new type of semantic text markup, specifies the key parameters of the final prospective development that define the contours of the formulated methodology. The work analyses existing approaches to the construction of semantic markup applicable to text corpora.

The key parameters are defined as follows: firstly, the eventfulness of the ontological description of the text; secondly, the change of focus from the semantics of individual statements/sentences, as units structuring the semantic space, to units of a different order – key events, their order and system; thirdly, the reverse order of the ontological description: not from grammar (syntax) to semantics, but from "expressed" to the form of expression.

The authors conclude by identifying some of the potential prospects for the markup system being developed, as well as the likely limitations and difficulties in development.

**Keywords:** semantics, corpus linguistics, National Corpus of the Russian Language, ontological approach, modelling of semantic markup.

### Введение

Количество накапливаемой научной информации о мире, в том числе о текстах и в виде текстов, постоянно растет. Процесс накопления информации, с одной стороны, позволяет двигаться к открытиям на новом уровне знаний, с другой же – своим объемом затрудняет дальнейшее продвижение: такой объем информации сложно охватить и использовать ограниченными ресурсами интеллекта и опыта отдельного исследователя. Эта проблема актуальна не только для естественных и формальных наук, но и для гуманитарных, в том числе филологии. Традиционно она решается путем разветвления специализаций научных работников, но на современном этапе развития заметно повышается актуальность и другого подхода: работы над созданием и внедрением систем искусственного интеллекта (ИИ), способного оперировать большими массивами данных и стать перспективным помощником в исследовательской деятельности.

Важным в этой связи представляется также формирование научных баз данных, которые могут использоваться как исследователями, так и инженерами для обучения систем ИИ, и в дальнейшем могут быть использованы также самими системами ИИ в процессе работы. Для лингвистов удачным опытом в этой сфере является активно развивающийся Национальный корпус русского языка [1]. С момента создания корпуса был раскрыт его колоссальный потенциал в

области применения в образовании [2], [3], компьютерной лингводидактике [4], а также для исследовательских задач в области компьютерной лингвистики [5].

К 2023 году корпус содержит объём текстов свыше 2 млрд слов, в составе которого синтаксический корпус, насчитывающий 1,5 млн слов. В дополнение к синтаксической, в корпус добавлены морфологическая, словообразовательная и семантическая разметка. Последняя, однако, реализована лишь частично, в числе доступных инструментов таксономия, мереология, разряды и др. общие признаки отдельных предметных, непредметных имен, имен собственных, прилагательных, числительных, местоимений, глаголов и наречий.

Учеными проводились исследования в области создания систем более детализированной и специализированной семантической разметки, но менее универсальной в сравнении с грамматической, применяемой к корпусам текстов ограниченной сферы, создаваемой вручную на базе онтологического подхода в содействии с экспертами в предметной области, что отдельно декларировалось авторами разработок: «При создании специализированных корпусов текстов обычно производится лингвистическое аннотирование (морфологическое, синтаксическое), не зависящее от предметной области и осуществляемое автоматически и/или вручную. Семантическая разметка, напротив, предметно ориентирована, поскольку определяется онтологией предметной области и производится экспертами» [6, С. 674].

В зарубежных исследованиях предпринимались попытки более широкого подхода к проблеме интерпретации естественного языка с помощью онтологического моделирования в рамках концепции Семантической паутины [7]. Авторы начинают с введения в понятие «онтологии» и на примере лингвистической модели «The Lexicon Model of Ontologies» («LeMOn») демонстрируют, как лингвистическая информация формулируется в онтологии. Далее эта онтологическая база служит основой для автоматической генерации грамматик, которые используются для композиционного построения смысловых представлений в соответствии с лежащей в основе онтологией лингвистических формализмов. Таким образом, исследователи пытаются пройти путь от «лингвистических формализмов» к семантике. В исследовании рассматривается ряд трудностей и ограничений такого подхода, связанного, к примеру, с омонимией, а также возможными способами их компенсации.

С похожей логикой развивается в данный момент проект «SemOntoCor» под руководством д. филол. н., проф. И. М. Богуславского: «SemOntoCor можно рассматривать как следующий шаг в развитии синтаксического корпуса SynTagRus, имеющего несколько уровней разметки. При разметке SemOntoCor на вход поступает морфо-синтаксическая разметка в формате SynTagRus, а на выходе строится базовая семантическая структура (BSemS). Эта структура представляет непосредственное значение предложения в терминах онтологических концептов, соединенных семантическими отношениями» [8, С. 13].

Ранее нами также предпринимались попытки описания морфологических свойств русского языка в рамках онтологического подхода с использованием модели «LeMOn» [9], [10].

Представляется весьма перспективной разработка модели семантической разметки на базе онтологического подхода, которая в перспективе может использоваться для дополнения существующих корпусов текстов новым инструментарием, расширяющим возможности использования корпуса для решения исследовательских задач.

### **Методы и принципы исследования**

Эксперименты с использованием «The Lexicon Model of Ontologies» и «OntoLex» подтолкнули нас к разработке подхода к онтологическому моделированию семантической разметки, в методологической основе которого будет лежать обратная последовательность: от моделирования сюжетно-событийной семантической структуры целых текстов (события – акты – предметные экземпляры) к подбору адекватной разработанной модели инструментария лингвистических формализмов и тезаурусов (словари и связанные онтологии, отражающие лексико-грамматические свойства конкретного текста), что определяет научную новизну исследования. Направление традиционного подхода обусловлено задачами автоматической обработки текста, ключевой из которых является извлечение «смыслов» (так или иначе интерпретируемых в различных исследованиях в соответствии с конкретными научными задачами) из исходно семантически «неразмеченного» текста, поэтому логичным представляется сначала декодировать его грамматическую, прежде всего синтаксическую структуру (парсинг), отталкиваясь от которой и в связи с которой можно перейти к определению семантики отдельных предложений.

Направление разрабатываемого нами подхода определяется иной, несколько более ограниченной научной задачей: найти оптимальный и универсальный способ описания событийной и связанной с ней предметной составляющих текста в форме онтологической структуры, который позволил бы создать новую систему сюжетно-событийной семантической разметки текстов корпуса, подходящей для решения прикладных исследовательских задач: поиск схожих по тем или иным параметрам сюжетов, «параллельных» описаний одного события, интертекстуальные пересечения текстов и т. п. Кроме того, в качестве основной единицы выступает не отдельное предложение или высказывание в декодируемом тексте, а структуры отдельных событий / процессов / явлений, описываемых в тексте, что также определяет инверсию от «смысла» к языковой «форме» его выражения, а не наоборот.

Достижение этой цели, прежде всего, предполагает конкретизацию контуров конечной разработки, конституирование требований к ней, что определяет тему данной статьи.

### **Основные результаты**

В основе моделирования сюжетно-событийной структуры текстов лежит онтологический подход, однако дополненный спецификой, обусловленной потребностью построения сюжетно-событийных онтологий. Традиционно онтологии используются для построения графов, в качестве объектов в которых используются предметные сущности. Таким образом, онтологии описывают преимущественно текущее состояние той или иной предметной области (к примеру, языка с позиции синхронии) в виде структуры множества классифицированных экземпляров (иерархия единиц, таксономия, классы/подклассы и т. п.). Это соответствует приведённой ранее в пример семантической разметке Национального корпуса русского языка, которая позволяет выделять преимущественно родовидовые

отношения отдельных лексических единиц, их иерархии. Тем не менее доступные на данный момент в НКРЯ инструменты существенно упрощают работу по сюжетно-событийному моделированию текста в плане выделения и классификации предметных экземпляров, что будет отмечено нами далее.

В основе текстов, в свою очередь, зачастую лежат события, процессы, явления. Здесь уместен пример из области синтаксиса: при составлении синтаксических деревьев в грамматике зависимостей в корне располагается сказуемое [11]. Анализ опыта онтологического моделирования семантики текстов показывает, что исследователями уделяется недостаточное внимание к разработке методологии построения структур, в основе которых была бы «событийность», а не «предметность». Существующие исследования чаще всего также касаются корпусов текстов узкой направленности и/или для решения узкого спектра задач, но, тем не менее, имеют потенциал расширения [12].

Таким образом, в основе разрабатываемого онтологического подхода к моделированию семантики текстов должна лежать событийность, пусть и создаваемая теми же инструментами, но иначе: семантическое пространство разделяется на отдельные события, происходящие связано или изолированно, последовательно или параллельно по отношению друг к другу. События, в свою очередь, находятся в «корне» множества отдельных связанных с ним актов, образующих общую структуру ключевого события отдельными системами собственных взаимосвязей. Акты же связаны с «вещной» частью онтологически описываемой структуры, в частности, с лицами, предметами, местами и т. п. Все экземпляры в дальнейшем могут связываться с онтологией, содержащей лингвистическую информацию о том, какими языковыми средствами они вводятся в текст, что определено нами для второго этапа разработки.

В качестве первого, экспериментального текста нами был выбран художественный текст, повесть братьев А. Н. и Б. Н. Стругацких «Понедельник начинается в субботу». Выбор, кроме эстетической и сюжетной привлекательности, обусловлен тем, что произведение имеет продолжающий сюжет текст: «Сказка о Тройке», что открывает дальнейшую перспективу в отработке использования созданной модели разметки на близком по содержанию произведении. Кроме того, писатели создали также комплекс текстов, объединенных в «Мире Полудня», включающего Вселенную, в структуру которой входят как реальные «миры» (планеты, звёзды), так и полностью вымышленные, что позволит апробировать инструментарий, соответствующий задаче «разграничения» реального / дополненного / вымышленного мира. На данный момент были выделены 328 событий и 2364 акта, разрабатывается терминологический аппарат, создана классификация «предметных» экземпляров (актантов, предметов и локаций), проводится активная работа по определению комплекса «универсальных семантических связей», которые позволят сконструировать сюжетно-семантическую онтологию на базе исходного текста. Результаты этапов данной работы будут публиковаться по мере их завершения, описать полноценный пример моделирования структуры целого текста в рамках одной статьи не представляется возможным в силу значительных объемов данных, однако, возможно привести некоторые иллюстративные примеры результатов работы над отдельными отрывками.

В первой главе повести нами была выделена и классифицирована цепь из двух последовательно происходящих и связанных друг с другом причинно-следственными отношениями ключевых событий: «Перемещение в Соловец» и «Заселение». Каждое из событий разделено на составляющие его акты в двух уровнях. Так, в первом событии определено четыре акта первого уровня («Встреча с охотниками на дороге», «Поездка», «Прибытие в НИИЧАВО»), которые, в свою очередь, разделены на 168 отдельных акта (отдельные высказывания / вопросы / ответы персонажей, их перемещения в пространстве и действия по отношению друг к другу и задействованным предметам). Каждый акт связан с лицами (актантами), иницирующими конкретное действие, лицами и предметами, на которое направлено данное действие, либо так или иначе связано с ними, а также пространства, которые являются местом действия. Каждый актант (три экземпляра в первом событии) и предмет (32 экземпляра в первом событии) связаны с предметными экземплярами второго порядка: детали портрета, одежда персонажей, составные элементы предметов и т.п., если таковые упоминаются в тексте (48 отдельных экземпляров в описании первого события). Приведенные данные демонстрируют высокую трудоёмкость процесса сюжетно-семантического онтологического моделирования, однако анализ существующей семантической разметки НКРЯ позволяет заключить, что весомая доля процессов может быть упрощена. Так, к примеру, корпус располагает инструментами, позволяющими выделить в анализируемых текстах отдельные экземпляры: людей, животных, неодушевлённые предметы и их отдельные структурные части. Следующий трудоёмкий этап, классификация экземпляров и присвоение соответствующего параметрам уникального кода, может производиться с помощью машинного обучения (по мере накопления данных, составляющих обучающую базу). Процесс обработки текста позволяет прийти к предварительным выводам о том, что многие из этапов работы с текстом в перспективе могут быть описаны, алгоритмизированы и автоматизированы, что позволит ускорить процесс накопления данных в перспективной репозитории сюжетно-семантических онтологий текстов, который в дальнейшем будет размещен в открытом доступе.

### Обсуждение

Предлагаемая разработка сюжетно-семантической модели текста на базе онтологии позволит создать поисковые инструменты нового уровня: оказывается доступным поиск описания схожих событий в разных текстах (например: «в каких текстах люди также перемещаются в указанный город?»); запрос может варьироваться и уточняться: «в каких текстах персонаж перемещается на автомобиле, встречает попутчиков и подвозит их?» и т.п.) Количество и качество возможных поисковых результатов зависят от множества параметров: во-первых, определяются количеством онтологически описанных текстов, во-вторых, глубиной и качеством детализации онтологической модели, в том числе, спецификой классификации включённых в неё событийных и предметных экземпляров.

Другим важнейшим для данного исследования аспектом текста, который должен найти своё отражение в формируемой методологии, является его соотношение с реальностью: в текстах могут описываться как реальные события (пересказывающих действительно происходившие известные события), так и дополненные (намеренно или нет) вымыслом, а также полностью вымышленные события (с упомянутыми реальными событиями/лицами/предметами/местами или нет) и миры, предполагаемые события (прогнозируемые или

гипотетически моделируемые) и т. д. Конструирование модели, в которой бы отражались связи реальности и вымысла представляется чрезвычайно сложной, но интересной и перспективной задачей. Этот принцип будет выражаться на всех этапах разработки методологии моделирования семантической событийной онтологии, в том числе в особенностях классификации предметных экземпляров.

Кроме того, разрабатываемый подход должен быть универсальным, система семантической разметки, создаваемая с его помощью, должна быть потенциально применима к любому тексту, содержащемуся в корпусе. По существу, это определяется универсальностью связей, набора предикатов, связывающего экземпляры онтологии. Здесь может быть использована логика и некоторые наработки «Universal Dependencies» («Универсальные связи»), используемых для «банков деревьев» (репозитория синтаксических структур) [13].

Наконец, полученная онтология не должна быть ограничена рамками описываемого ей конкретного текста, как отдельный граф. Онтология, лежащая в основе системы семантической разметки, может описывать все содержащиеся в ней тексты, расширяясь по мере дополнения новыми целыми текстами, либо отдельными событиями, описываемыми в том или ином тексте. Это позволит смоделировать семантическую структуру «текста в контексте» других имеющихся в корпусе текстов, отразить, с одной стороны, интертекстуальные связи, явные и неявные «переклички» текстов друг с другом, с другой – продемонстрировать сходства и различия описаний одних и тех же реальных событий, упоминаемых в различных текстах, имеющих и не имеющих интертекстуальную соотнесённость друг с другом («цитирующих» друг друга или созданных полностью изолированно друг от друга).

### Заключение

Описанные общие принципы, отражающие требования к принципиально новому типу семантической разметки, позволяют рассмотреть контуры конечного продукта, который может показаться чрезмерно сложным в своей разработке. Примером может послужить синтаксически размеченный корпус в составе Национального корпуса русского языка, который изначально создавался «вручную» заинтересованными исследователями. Однако, на данный момент создателями все более активно используются инструменты автоматической обработки текста – «НейроКРЯ» (сервис разметки, созданный на базе нейросетевых технологий). Безусловно, создание семантически размеченного корпуса должно следовать аналогичной траектории: методология разрабатывается на базе ограниченного комплекса текстов различных функциональных стилей и предметных областей. Полученные результаты позволяют алгоритмизировать процессы обработки текстов и подобрать необходимый инструментарий, который позволит сократить ресурсозатратность работы.

Разрабатываемая система семантической разметки в перспективе позволит собирать и анализировать информацию из корпуса текстов не только с точки зрения его «формального» портрета (грамматики, словарной и частотной составляющих корпуса), но также с точки зрения выражаемой в текстах «реальности», позволит продемонстрировать семантические связи текстов друг с другом, находить схожие события, близкие по сюжету тексты и т. д. Безусловно, обозначенная задача является чрезвычайно комплексной и потребует много времени на реализацию даже при условии применения различного инструментария (машинное обучение, использование имеющейся семантической разметки открытых корпусов, искусственный интеллект, автоматическая обработка текста), а результат на первых этапах может оказаться несколько «наивным», недостаточно детализированным. Однако разработка будет открыта для совершенствующих дополнений, а ценность перспективного результата мотивирует дальнейший исследовательский интерес к работе.

### Конфликт интересов

Не указан.

### Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

### Conflict of Interest

None declared.

### Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

### Список литературы / References

1. Национальный корпус русского языка. — URL: [www.ruscorgo.ru](http://www.ruscorgo.ru) (дата обращения: 08.12.2023).
2. Добрушина Н. Р. Как использовать Национальный корпус русского языка в образовании? / Н.Р. Добрушина // Национальный корпус русского языка: 2003 – 2005. — М.: Индрик, 2005. — С. 308-329.
3. Добрушина Н. Р. Корпусные методики обучения русскому языку / Н.Р. Добрушина // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы. — СПб: Нестор-История, 2009. — С. 335-352.
4. Дмитриев А. В. Потенциал корпусной лингвистики в подготовке специалистов в области компьютерной лингводидактики / А.В. Дмитриев, М.С. Коган // Научно-технические ведомости СПбГУ. Гуманитарные и общественные науки. — 2019. — Т. 10, № 4. — С. 69-85.
5. Полицин С. А. Применение комплекса инструментов управления корпусами текстов при решении задач компьютерной лингвистики / С.А. Полицин, Е.В. Полицина // Вестник ВГУ. Серия: Системный анализ и информационные технологии. — 2019. — № 2. — С. 134-142.
6. Загоруйко М. Ю. Система семантической разметки корпуса текстов в ограниченной предметной области / М.Ю. Загоруйко, И.С. Кононенко, Е.А. Сидорова // Компьютерная лингвистика и интеллектуальные технологии: По материалам Международной конференции «Диалог». — Вып. 11 (18). — М.: Изд-во РГГУ, 2012. — С. 674-683.

7. Cimiano P. *Ontology-Based Interpretation of Natural Language* / P. Cimiano, C. Unger, J. McCrae // Morgan & Claypool, *Synthesis Lectures on Human Language Technologies*. — 2014. — 178 p. — URL: <https://aclanthology.org/J15-2006.pdf> (accessed: 21.11.2023)
8. Богуславский И. М. Разработка семантического корпуса русского языка: SemOntoCor / И.М. Богуславский, В.Г. Диконов, Е.С. Иншакова и др. // *Компьютерная лингвистика и интеллектуальные технологии. По материалам международной конференции «Диалог»*. — 2023. — Вып. 22. — С. 12-25.
9. Balysheva K. *Identifying Morphological Properties of Russian Words with the Ontology-Based Analyser* / K. Balysheva, E. Kartashova, K. Kondratiev et al. // *Abstracts and Proceedings of SOCIOINT 2017 — 4th International Conference on Education, Social Sciences and Humanities*. — URL: <http://www.ocerint.org/socioint17%20e-publication/abstracts/a264.html> (accessed: 21.11.2023)
10. Balysheva K. *OntoLex as a Model for Creating the Ontology-Based Dictionary of Russian Grammatical Forms* / K. Balysheva, E. Kartashova, K. Kondratiev et al. // *Proceedings of the LDK 2017 Workshops: 1st Workshop on the OntoLex Model, Shared Task on Translation Inference Across Dictionaries & Challenges for Wordnets co-located with 1st Conference on Language, Data and Knowledge (LDK 2017)*. — 2017. — P. 51-59. — URL: [http://ceur-ws.org/Vol-1899/OntoLex\\_2017\\_paper\\_7.pdf](http://ceur-ws.org/Vol-1899/OntoLex_2017_paper_7.pdf) (accessed: 21.11.2023)
11. Тестелец Я. Г. Слово и предложение. Структура зависимостей / Я.Г. Тестелец // *Введение в общий синтаксис*. — М.: РГГУ, 2001. — С. 61-106.
12. Лядова Л. Н. Формирование событийных рядов с использованием многоаспектных онтологий / Л.Н. Лядова, В.С. Заякин, М.А. Смирнов // *X Международная научно-техническая конференция «Технологии разработки информационных систем» (ТРИС-2020)*. — Таганрог: Южный федеральный университет, 2020. — С. 297-303.
13. Marneffe M.-C. *Universal Dependencies* / Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre et al. // *Computational Linguistics*. — 2021. — Volume 47. — Issue 2. — P. 255-308.

### Список литературы на английском языке / References in English

1. Nacional'nyj korpus russkogo yazyka [National Corpus of the Russian Language]. — URL: [www.ruscorpora.ru](http://www.ruscorpora.ru) (accessed: 12.08.2023). [in Russian]
2. Dobrushina N. R. *Kak ispol'zovat' Nacional'nyj korpus russkogo yazyka v obrazovanii?* [How to Use the National Corpus of the Russian Language in Education?] / N.R. Dobrushina // *Nacional'nyj korpus russkogo yazyka [National Corpus of the Russian Language]: 2003 – 2005*. — М.: Indrik, 2005. — P. 308-329. [in Russian]
3. Dobrushina N.R. *Korpusnye metodiki obucheniya russkomu yazyku* [Corpus Methods for Teaching the Russian Language] / N.R. Dobrushina // *Nacional'nyj korpus russkogo yazyka: 2006-2008. Novye rezul'taty i perspektivy* [National Corpus of the Russian Language: 2006-2008. New Results and Prospects]. — St. Petersburg: Nestor-Istoriya, 2009. — P. 335-352. [in Russian]
4. Dmitriev A.V. *Potencial korpusnoj lingvistiki v podgotovke specialistov v oblasti komp'yuternoj lingvodidaktiki* [The Potential of Corpus Linguistics in Training Specialists in the Field of Computer Linguodidactics] / A.V. Dmitriev, M.S. Kogan // *Nauchno-tehnicheskie vedomosti SPbGU. Gumanitarnye i obschestvennye nauki* [Scientific and Technical Bulletins of St. Petersburg State University. Humanities and Social Sciences]. — 2019. — Vol. 10, N. 4. — P. 69-85. [in Russian]
5. Politsin S. A. *Primenenie kompleksa instrumentov upravleniya korpusami tekstov pri reshenii zadach komp'yuternoj lingvistiki* [Application of a Set of Tools for Managing Text Corpora in Solving Problems of Computer Linguistics] / S.A. Politsin, E.V. Politsina // *Vestnik VGU. Seriya: Sistemnyj analiz i informacionnye tekhnologii* [Bulletin of VSU. Series: System Analysis and Information Technologies]. — 2019. — N. 2. — P. 134-142. [in Russian]
6. Zagorulko M. Yu. *Sistema semanticheskoy razmetki korpusa tekstov v ogranichennoj predmetnoj oblasti* [System of Semantic Markup of a Text Corpus in a Limited Subject Area] / M. Yu. Zagorulko, I.S. Kononenko, E.A. Sidorova // *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Po materialam Mezhdunarodnoj konferencii «Dialog»* [Computer Linguistics and Intellectual Technologies: Based on the materials of the International Conference “Dialogue”]. — Vol. 11 (18). — М.: Publishing House of the Russian State University for the Humanities, 2012. — P. 674-683. [in Russian]
7. Cimiano P. *Ontology-Based Interpretation of Natural Language* / P. Cimiano, C. Unger, J. McCrae // Morgan & Claypool, *Synthesis Lectures on Human Language Technologies*. — 2014. — 178 p. — URL: <https://aclanthology.org/J15-2006.pdf> (accessed: 21.11.2023)
8. Boguslavsky I. M. *Razrabotka semanticheskogo korpusa russkogo yazyka: SemOntoCor* [Development of a Semantic Corpus Russian Language: SemOntoCor] / I.M. Boguslavsky, V.G. Dikonov, E.S. Inshakova [et al.] // *Komp'yuternaya lingvistika i intellektual'nye tekhnologii. Po materialam mezhdunarodnoj konferencii «Dialog»* [Computer Linguistics and Intellectual Technologies. Based on materials from the International Conference “Dialogue”]. — 2023. — Vol. 22. — P. 12-25. [in Russian]
9. Balysheva K. *Identifying Morphological Properties of Russian Words with the Ontology-Based Analyser* / K. Balysheva, E. Kartashova, K. Kondratiev et al. // *Abstracts and Proceedings of SOCIOINT 2017 — 4th International Conference on Education, Social Sciences and Humanities*. — URL: <http://www.ocerint.org/socioint17%20e-publication/abstracts/a264.html> (accessed: 21.11.2023)
10. Balysheva K. *OntoLex as a Model for Creating the Ontology-Based Dictionary of Russian Grammatical Forms* / K. Balysheva, E. Kartashova, K. Kondratiev et al. // *Proceedings of the LDK 2017 Workshops: 1st Workshop on the OntoLex Model, Shared Task on Translation Inference Across Dictionaries & Challenges for Wordnets co-located with 1st Conference on Language, Data and Knowledge (LDK 2017)*. — 2017. — P. 51-59. — URL: [http://ceur-ws.org/Vol-1899/OntoLex\\_2017\\_paper\\_7.pdf](http://ceur-ws.org/Vol-1899/OntoLex_2017_paper_7.pdf) (accessed: 21.11.2023)

11. Testelets Ya. G. Slovo i predlozhenie. Struktura zavisimostej [Word and Sentence. Dependency Structure] / Ya. G. Testelets // Vvedenie v obshchij sintaksis [Introduction to General Syntax]. — M.: Russian State University for the Humanities, 2001. — P. 61-106. [in Russian]
12. Lyadova L.N. Formirovanie sobytijnyh ryadov s ispol'zovaniem mnogoaspektnyh ontologij [Formation of Event Series Using Multi-aspect Ontologies] / L.N. Lyadova, V.S. Zayakin, M.A. Smirnov // X Mezhdunarodnaya nauchno-tekhnicheskaya konferenciya «Tekhnologii razrabotki informacionnyh sistem» (TRIS-2020) [X International Scientific and Technical Conference “Technologies for the Development of Information Systems”]. — Taganrog: Southern Federal University, 2020. — P. 297-303. [in Russian]
13. Marneffe M.-C. Universal Dependencies / Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre et al. // Computational Linguistics. — 2021. — Volume 47. — Issue 2. — P. 255-308.