

**СЕМАНТИКО-СОБЫТИЙНАЯ ОНТОЛОГИЯ EVENTONTO: БАЗОВЫЙ ТЕРМИНОЛОГИЧЕСКИЙ
АППАРАТ, ПРОЦЕССЫ РАЗРАБОТКИ И ПРИНЦИПЫ КЛАССИФИКАЦИИ ПРЕДМЕТНЫХ ЕДИНИЦ**

Научная статья

Михеев А.А.^{1,*}

¹ ORCID : 0000-0003-1119-6654;

¹ Марийский государственный университет, Йошкар-Ола, Российская Федерация

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

Аннотация

Статья представляет обзор и анализ существующих подходов к созданию семантической разметки текстов. В ней предлагается новый подход, основанный на событийном онтологическом описании текста, фокусировании на ключевых событиях текста и их классификации, а также на обратном порядке составления онтологического описания: от выражаемого к форме выражения. В работе описываются основные этапы разработки событийной онтологии, ее ключевые параметры и потенциальные перспективы, а также ограничения и сложности в ее создании.

Автор описывает разработанную концепцию семантико-событийная онтология EventOnto, которая включает базовый терминологический аппарат, процессы разработки и принципы классификации предметных единиц. Для разработки EventOnto автор предлагает пройти 8 этапов, каждый из которых подробно описан в статье (отбор подкорпуса текстов для составления экспериментальной модели событийной онтологии, выделение в избранных текстах предметных экземпляров, выделение событий и актов, моделирование связей актов, формирующих структуру событий, моделирование связей актов с предметными экземплярами и предметных экземпляров, расчёт ресурсов и составление прогнозов о ресурсозатратности обработки текстового материала, работа с лингвистическими данными, оптимизация процессов работы с лингвистическими данными).

Ключевые слова: семантика, корпусная лингвистика, Национальный корпус русского языка, онтологический подход, семантико-событийная онтология EventOnto.

**SEMANTIC AND EVENT ONTOLOGY EVENTONTO: BASIC TERMINOLOGY, DEVELOPMENT PROCESSES
AND PRINCIPLES OF SUBJECT UNIT CLASSIFICATION**

Research article

Mikheev A.A.^{1,*}

¹ ORCID : 0000-0003-1119-6654;

¹ Mari State University, Yoshkar-Ola, Russian Federation

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

Abstract

The article presents a review and analysis of existing approaches to the creation of semantic markup of texts. It proposes a new approach based on event ontology description of text, focusing on key events of text and their classification, as well as on the reverse order of ontology description compilation: from the expressed to the form of expression. The work describes the main stages in the development of an event ontology, its key parameters and potential prospects, as well as the limitations and difficulties in its creation.

The author describes the developed concept of semantic and event ontology EventOnto, which includes basic terminological apparatus, development processes and principles of classification of subject units. To develop EventOnto, the author proposes to go through 8 stages, each of which is described in detail in the article (selection of a sub-corpus of texts to compose an experimental model of the event ontology, selection of subject instances in selected texts, selection of events and acts, modelling of links of acts forming the structure of events, modelling of links of acts with subject instances and subject instances, calculation of resources and making predictions about the resource-cost of processing textual material, work with linguistic data, optimization of linguistic data processes).

Keywords: semantics, corpus linguistics, National Corpus of the Russian Language, ontological approach, semantic and event ontology EventOnto.

Введение

В данной статье описываются итоги одного из этапов обширной программы исследований по разработке методологии конструирования событийных онтологий. Проект разработки событийной онтологии, общая концепция которой была представлена ранее [11], требует уточнений, касающихся, прежде всего, терминологического аппарата и поэтапного плана исследовательских работ, что в совокупности составляет цель данной работы.

Ключевым в нашей концепции является понятие событийной онтологии, под которой (в наиболее сжатой формулировке) мы предлагаем понимать *концептуализацию информации о событиях, формирующих смысловое пространство текста*. Концептуализация предполагает организацию эмпирического материала, предметных смысловых единиц, в схемы, позволяющие отобразить функции данных единиц в контексте события, а также их связи друг с другом. Соотношения единиц формулируются в виде связанной совокупности RDF-утверждений, позволяющей составить общую онтологию включенных в нее текстов.

События определяются как запечатленный в тексте факт изменения предметных единиц в пространстве-времени, представляющий собой совокупность отдельных актов (явлений физического мира, направленных действий, а также

иницированных действиями процессов), характеризующихся общей завершенностью (достижением результата), что определяет саму возможность умозрительно выделить его как отдельное событие в тексте. Наиболее иллюстративным, на наш взгляд, будет пример какого-либо известного яркого исторического события, которое имеет результат, продуцирующий ряд других событий. К таковым может относиться, к примеру, переход Юлия Цезаря через Рубикон со своими легионами 10 января 49 г. до н.э. Произошедшее событие имеет конкретный результат – объявление гражданской войны, включающей множество других отдельных событий: сражений, военных маневров и перемещений, политических изменений и пр. При этом каждое из событий является уникальным, что определяется как его внутренним содержанием (совокупность лиц, предметов, отдельных действий, явлений и процессов, имевших отношение к конкретному событию в конкретном месте), так и его положение в контексте других событий, происходивших до, после и одновременно с ним, иными словами, положение события в контексте времени. Безусловно, не каждое событие, отраженное в том или ином тексте, имеет столь глобальное значение, как в представленном примере: это может быть совместное перемещение героев текста в какой-либо город, встреча, обед, создание чего-либо и т. п.. Определение более четких границ событий, их классификация в контексте реализации проекта, будут возможны на более поздних этапах работы в связи с накоплением достаточного объема данных. На текущем этапе, в рамках которого проводятся работы по выделению и классификации предметных экземпляров, мы оперируем несколько абстрактным определением понятия «событие», которое будет уточняться в процессе работы над проектом. Общая интерпретация позволяет провести некоторые иллюстративные параллели с принципами выделения в тексте сверхфразовых единств (сложных синтаксических целых) на синтаксическом уровне, представляющих собой совокупность связанных между собой предложений, в предикативном ядре которых мы чаще всего находим единицы, выражающие отдельные действия, процессы явления (сказуемые), при этом контуры сложного синтаксического целого определяются возможностью умозрительно зафиксировать его общую микротему (в нашем случае – общий результат события) [7, С. 356].

Разграничение явлений, действий и процессов, своими параметрами и системой связей определяющих формальную структуру события, также требует уточнения. Под действиями подразумеваются отдельные акты, инициируемые действующими лицами (актантами), обладающими относительной свободой воли в принятии решений (здесь наблюдается близость, но не тождественность с категорией одушевленности), иллюстративным действием может быть, к примеру, произнесение приказа переправиться через реку. Процессами мы называем цепочки актов, инициированных действиями (и, косвенно, актантами), но в дальнейшем происходящими независимо от него направлено или циклично под его контролем или вне контроля, к примеру, падение брошенного камня, работа какой-либо запущенной машины и т.п. В свою очередь, к явлениям здесь относятся факты природы, происходящие независимо, изолированно от актантов, являющиеся контекстом, условиями и обстоятельствами действий и процессов.

Необходимо также четко представлять контуры той «информации», которую представляется возможным извлечь из текста для дальнейшей онтологической концептуализации. Прежде всего, выделяется «явная» информация, которая прямо следует из текста (например, из предложения «Иван читает книгу» мы знаем, что актором здесь является «Иван», действие, которое он совершает – «читает», а объектом является «книга»), а также «тёмное поле» семантики, которое, в свою очередь, разделяется на «восстановимое» и «неизвестное». Восстановимой может быть явно (с высокой степенью вероятности) следующая из текста дополнительная информация (например, «Иван» – лицо мужского пола). Контекстуально-предположительной является неоднозначная дополнительная информация, которую можно «предположить» с той или иной степенью вероятности. Так, например, мы можем предположить, что книга сделана из бумаги, но может использоваться электронный носитель информации, либо, напротив, это может быть древняя книга, написанная на специально выделанной коже животного («вель»). Достоверность того или иного умозаключения, дополняющего основной выводимый смысл текста, зависит от контекста окружающих действие событий, предметов и локаций: вероятность использования «древнего» текста будет выше, если действие происходит в библиотеке музея, где хранятся такие книги, а электронного носителя – если действие происходит в современную эпоху. Следует подчеркнуть, что в рамках текущего проекта мы ограничиваем себя «явной» информацией, углублённая концептуализация событийной информации текста с расширением в область дополняющих текст вероятных логических выводов мыслится возможным перспективным направлением развития текущего проекта в том случае, если его базовая составляющая будет успешно реализована. Так, к примеру, по схожей логике в отечественных исследованиях велась работа с использованием многофункционального лингвистического процессора «ЭТАП-3» в рамках реализации проекта «Референциальные аспекты компьютерной семантики», поддержанного Российским научным фондом [8].

Работы по составлению корпусов и словарей «пропозиций» (называемыми в зарубежной научной литературе «proposition banks» или «propbanks»), в предикативном центре которых находится глагол, связанный с теми или иными предметными единицами (лицами и объектами), обширно представлены на англоязычном материале. Прежде всего, к ним относятся проекты FrameNet [6], VerbNet [4], [5]. В отечественных исследованиях также поднимался вопрос о перспективах создания русскоязычного аналога, представлена адекватная для русского языка модель словаря [9]. Несмотря на некоторую схожесть идей, предлагаемая нами разработка концептуально отличается от приведенных примеров: предполагается разработать связанный массив данных о событийно-предметной структуре текстов, а не репозиторий отдельных аннотированных конструкций. Некоторые идеи реализованных проектов тем не менее могут быть использованы в работе над текущим проектом. Особо следует подчеркнуть успешно развивающийся в данный момент проект Лаборатории № 15 (ИППИ им. А. А. Харкевича) SemOntoCor под руководством д. филол. н., профессора И. М. Богуславского [2, С. 12-25]. Разработчики также ставят своей целью создание семантического корпуса на основе лингвистических и онтологических ресурсов, создаваемого на базе синтаксического корпуса SynTagRus. Несмотря на близкую идеологию проектов, описываемый здесь событийно-предметный подход онтологической концептуализации информации о тексте, безусловно, отличается в своих целях и методологической основе. Представляется невозможным также сопоставлять проекты, поскольку поставленные задачи, искомые результаты являются абсолютно различными. Предполагаемая разработка, получившая рабочее название EventOnto, не

ставит своей целью создание семантического анализатора, способного интерпретировать смысл тех или иных предложений, написанных на естественном языке, либо использовать естественный язык в искусственно сгенерированной речи, т.е. ключевых задач автоматической обработки естественного языка и Искусственного интеллекта. Перед нами стоит задача систематизации событийных данных, извлекаемых из текста, в машиночитаемом формате онтологии, которые предполагается использовать для создания семантико-событийной разметки, которая, в свою очередь, позволит создавать новые поисковые инструменты по корпусам текстов. К примеру, в перспективе это может помочь в создании инструментов, позволяющих быстро подобрать схожие по структуре цепочки событий в различных текстах (например, схожие с переходом Рубикона поворотные исторические события, где пересечение некой условной «красной черты» приводит к цепочке связанных событий, схожих по тем или иным параметрам, интенсивности, с античными событиями гражданской войны), отдельные схожие события (например, путешествия из одного города в другой, столкновения с непреодолимым препятствием, т. н. «бродячие сюжеты»), подчеркивать сходства и различия отдельных выделенных пользователем событий. На данный момент не существует корпусов текстов с подобными семантико-событийными поисковыми инструментами, разработка данного инструментария по существу является экспериментальной работой, что определяет некоторую «гибкость» подхода, не позволяя, в свою очередь, сформулировать достаточно четкие количественно-качественные параметры результата работы до завершения всех ее ключевых этапов. Полученная база данных машиночитаемой информации в дальнейшем также может использоваться для обучения систем на базе искусственного интеллекта, однако потенциал такого использования также может быть определен лишь результатом реализации текущего проекта.

Более подробно процессы работы над проектом конкретизированы в поэтапном плане. Первые три этапа частично завершены к текущему моменту, в том числе:

1-й этап: определение конечного продукта, его параметров, а также исходных требований к нему. Потенциал использования событийных онтологий представляется достаточно широким, однако определение конкретной разработки и детализация её параметров (насколько это представляется возможным для исходного этапа) позволит адекватно рассчитать распределение ресурсов в проекте и позволит наглядно продемонстрировать некоторые прикладные перспективы формируемого подхода. Конечная разработка представляет собой онтологический граф, ключевыми элементами которого станут отраженные в избранном корпусе текстов события, представляющие собой совокупность связанных актов (отдельных действий, процессов и явлений) и предметных экземпляров. Событийная онтология будет дополнена связанной с исходными текстами лингвистической информацией о лексике, грамматике и стилевой специфике. Полученная онтология будет применима к существующим текстовым корпусам, что позволит создать набор поисковых инструментов по текстам: поиск сюжетно-событийных интертекстуальных связей различных текстов внутри корпуса, описаний схожих событий в различных текстах (их сходства и различия в семантическом, событийно-предметном аспектах, а в перспективе – связанных грамматических и стилистических аспектах), степень реальности / вымышленности событий текста. Данные инструменты могут использоваться как для филологических (лингвистических и литературоведческих) исследований, так и для специфичных творческих задач. Полученные метаданные о текстах корпуса потенциально могут использоваться также для задач при обучении генеративных языковых моделей, что, к примеру, может повысить сюжетно-композиционное качество текстов, генерируемых языковыми моделями ИИ.

2-й этап: отбор подкорпуса текстов для составления экспериментальной модели событийной онтологии, адекватного задачам, следующим из требований и параметров конечного продукта. Подход должен быть адаптирован к использованию текстов различных жанров и функциональных стилей, в частности, нами будут использоваться тексты с художественным повествованием, официально-деловые документы, тексты описания исторических событий, научные тексты, тексты СМИ. Первым текстом для построения опытной модели событийной онтологии избран текст повести А.Н. и Б.Н. Стругацких «Понедельник начинается в субботу» [14]. Разработка данной модели позволит отобрать тексты других жанров и функциональных стилей, событийно связанные с исходным текстом [10].

3-й этап: выделение в избранных текстах предметных экземпляров, их классификация и присвоение уникальных номеров. Под предметными экземплярами событийной онтологии понимаются единицы материального мира (лица и предметы), которые могут выполнять различные роли в событии. Согласовываясь с основными функциями предметных единиц, нами было выделено три ключевых класса: актанты, вещи и пространства. Принципы выделения и классификации данных единиц на базе избранного для экспериментальной событийной онтологии текста (повесть А.Н. и Б.Н. Стругацких «Понедельник начинается в субботу») составляют предмет данного исследования. Следует подчеркнуть, что на текущем этапе оставлены без внимания «нематериальные» единицы и абстрактные понятия. Включение единиц такого уровня может быть реализовано в дальнейшем на базе созданной «наивной» модели событийной онтологии.

Следующие этапы работы над проектом относятся к числу незавершенных:

4-й этап: выделение событий и актов, моделирование связей актов, формирующих структуру событий. При построении событийно-актовой структуры не будет учитываться временные величины, вполне позволяющие отразить течение времени, включая календарные даты описываемых событий. Нельзя также представить данную модель вполне соответствующей этерналистскому подходу к природе времени, где все события происходили бы «одномоментно». В базовой модели событийной онтологии учитываются причинно-следственные связи, отношения событий друг с другом в позиционных отношениях «до», «после», «одновременно». Механизмы описания событийно-актовых связей будут предметом отдельного исследования на материале результатов завершения четвертого этапа работы над проектом EventOnto.

5-й этап: моделирование связей актов с предметными экземплярами и предметных экземпляров друг с другом.

6-й этап: оптимизация процессов 3 и 4 этапа, расчёт ресурсов и составление прогнозов о ресурсозатратности обработки текстового материала.

Следует отметить, что работа с 1 по 6 этапы будет производиться циклично. Анализ и обработка первого текста, включающая выделение и классификацию предметных и событийно-актовых экземпляров, а также моделирование

онтологии, производится исключительно «вручную» с использованием существующих ресурсов Национального корпуса русского языка [12]. Ход проведенных работ будет проанализирован, что позволит изучить возможности использования ресурсов автоматической обработки языка, разработать предложения по оптимизации процессов. Результаты шестого этапа, проведенные на основе анализа работ с текстом повести «Понедельник начинается в субботу», будут использованы для работы над следующим событийно связанным текстом авторов «Сказка о Тройке» [13, С. 5-176]. В дальнейшем будут использованы тексты других авторов, жанров и стилей. Выбор художественного текста для первичной обработки мотивирован тем, что для литературного произведения характерны целостность и определенное развитие сюжета, последовательность событий, система действующих лиц с определенными взаимоотношениями, что позволяет определить его наиболее удобной отправной точкой исследований.

7-й этап: работа с лингвистическими данными (лексика, грамматика, стилистические показатели отдельных текстов). В качестве идейно схожей разработки здесь интересен опыт взаимодействия онтологии верхнего уровня SUMO и лингвистической базы данных WordNet, проанализированной в зарубежной научной литературе [1]. Данный этап также требует комплексного изучения.

8-й этап: оптимизация процессов работы с лингвистическими данными.

Важно также подчеркнуть, что предварительный поэтапный план работы над проектом в целом зависит от промежуточных результатов, что обуславливает вероятность его актуальных корректировок в дальнейшем.

К ключевым рискам проекта относится, в первую очередь, необходимость работы с колоссальным объемом данных. Создать онтологию достаточно крупного корпуса текста вручную представляется абсолютно невозможным. Поэтому объем текстов, необходимых для создания экспериментальной модели, является достаточно небольшим. Результат этой работы позволит оценить необходимые ресурсы и оптимизировать работу, автоматизировать процессы обработки текста (этапы 6 и 8). Иными словами, ручная обработка небольшого объема текстов для конструирования экспериментальной событийной онтологии позволит проанализировать и определить как и в каком объеме для получения качественно схожего результата могут быть применены существующие разработки и технологии (к примеру, перспективы использование машинного обучения для классификации предметных экземпляров). В несколько обобщенном смысле, это предполагает анализ того, как уже существующие технологии работы с языковым материалом могут быть применены для создания технологий другого качества на примере конкретной задачи.

Методы и принципы исследования

В методологической основе третьего этапа работы над проектом EventOnto лежит классификация предметных экземпляров. В основу данной классификации заложены адекватные исходным требованиям к разработке параметры выделения предметных единиц, в частности, отражена специфика реальности-вымышленности экземпляров, определена их таксономия (отношения часть-целое). Обработка первого текста («Понедельник начинается в субботу») производится исследователями самостоятельно, исходя из результатов анализа произведения. Для дальнейшего моделирования связей предметных единиц необходимым является присвоение им уникальных номеров, отражающих их позицию в общей классификации.

В дальнейшем моделировании мы исходим из ряда общих принципов, которые будут влиять на работу с отдельными экземплярами. Прежде всего, необходимо разделить в структуре разработки уникальные, семантически значимые единицы текстов, а также универсальные единицы, которые будут содержаться в тезаурусах верхнего уровня.

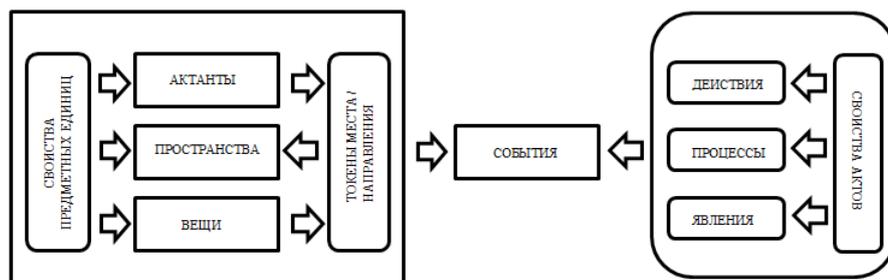


Рисунок 1 - Разработанная классификация событий

Все основные единицы предметного мира (актанты, вещи и пространства) относятся к уникальным. Экземпляры свойств предметных единиц и актантов, действия, процессы и явления, а также токены пространства и направления (прежде всего, сюда относятся предлоги «в», «на», «из», «от» и т.п.) содержатся в отдельных универсальных тезаурусах. С точки зрения интерпретации содержания текста это значит буквально следующее: всякое упоминаемое в тексте действующее лицо (актант), вещь или пространство являются уникальными единицами описываемого реального или вымышленного мира, обладают присвоенным именем («Александр Иванович Привалов», SAA0000AATON; «Похититель дивана 1», SAA0006AATON) или названием («Автомобиль Привалова», OAA0008AATOM). Действие, процесс или явление может исполняться различными актантами в тех или иных последовательностях или одновременно, связано друг с другом или параллельно, к примеру, «читать» может кто угодно и где угодно, само действие не является уникальным. Однако конкретный читатель, отдельный человек, является уникальной единицей, равно как и конкретная книга является уникальной единицей предметного мира. Чтение книги становится уникальным действием текста, когда связывает между собой уникальные единицы (субъект и объект действия, чтеца и книгу). Таким образом, единицы, выражающие то или иное действие / процесс / явление,

относятся к числу универсальных и могут заключаться в общий тезаурус, могут использоваться для неограниченного количества лиц или предметов в онтологических описаниях событий текста. Если количество предметных единиц ограничено пределами реального или вымышленного мира, отраженного в текстах, то количество возможных действий, процессов и явлений ограничено ресурсами языка, на котором текст написан. Иными словами, возможно создавать тексты со сколь угодно большим количеством персонажей, присваивая им новые имена, но количество их вероятных действий будет ограничено. Гипотетически можно составлять текст, перечисляя и описывая в нем бесконечное множество лиц или предметов, но мы не сможем составить текст, в котором будет перечислено бесконечное количество действий, которое может совершить конкретное лицо, даже если наделим его какими-либо невероятными свойствами, такой перечень будет ограничен ресурсами языка и нашего опыта, а также будет приблизительно совпадать с перечнем возможных действий других представителей того же класса. Тем же свойством ограниченности обладают и единицы, указывающие на свойства предметных единиц (цвет, форма, качество и т.д.) и актов (скорость, качество и т.д.). Тем не менее акты, связываясь с предметными единицами, совместно образуют структуру уникальных событий текста, являющихся ключевой единицей EventOnto.

Основные результаты

Выделенные предметные единицы распределены на три основных класса: актанты, вещи и пространства. Под актантами (в анализируемом тексте выявлено 195 единиц) понимаются реальные или потенциальные инициаторы действий и процессов, связанных с другими предметными единицами или самими собой («мыться») или частью себя («умыть лицо»). К классу «вещи» (в тексте выделено 389 единиц) относятся материальные предметы как естественного происхождения («Солнце», OAA0007RN), так и созданные актантами («Спальный мешок», OAA0021AATOM). Для двух указанных классов применимы общие принципы классификации.

S актант	AA0000 уник.№	R подлин.				N ест.
O вещь		A вымышл.	P прототип.			M иск.
			A сотвор.	T типич.	O обыкн.	
					S надел.	
				U уник.		

Рисунок 2 - Классификация предметных единиц «актанты» и «вещи»

Отдельные характеристики актанта или вещи влияют на присваиваемое ему уникальное название. Первая буква экземпляра определяется его классом: S (актант) или O (вещь). Следующие шесть символов представляют собой уникальный номер, включающий две буквы и четыре цифры (количество вероятных комбинаций позволит присвоить уникальный номер каждому отдельному экземпляру формируемой экспериментальной событийной онтологии). Далее следует распределение подлинных (существовавших или существующих единиц, лиц или вещей, реального мира, действующих в пределах известной реальности) и вымышленных (единиц, созданных авторами для описания того или иного события). Вымышленные единицы распределяются на «прототипичные» (P), т. е. имеющие некий реальный прототип в известной исторической действительности, но действующий в пределах созданной автором «новой» реальности (к примеру, Иван «Грозный» в событиях фильма «Иван Васильевич меняет профессию»), либо «сотворённые» (A), т. е. полностью созданные автором для описания события. «Сотворённые» автором единицы распределяются на «типичные» (T), т. е. имеющиеся в действительности как класс актанты (человек, кот и т.п.) или вещи (автомобиль, стол, часы и т.п.). К «уникальным» (U) относятся единицы, не имеющие в реальном мире аналогичных существ или предметов, относящихся к тому же классу (таким образом, к примеру, мы могли бы классифицировать Сета, упоминаемого в древнеегипетской мифологии в качестве одного из богов). «Типичные» экземпляры распределяются на «обыкновенные» (O), действующие и функционирующие аналогично представителям своего класса в действительности, а также «наделённые» (S), обладающие некими свойствами, представляющиеся неординарными для представителей «родного» класса в действительности («Кот Василий», SAA0004AATSN, который является котом, но при этом способен разговаривать на человеческом языке). Закрывающий символ названия экземпляра определяет тип его происхождения: естественное (N) либо искусственное (M), что позволяет выделить созданные другими актантами существа (к примеру, искусственный интеллект) или вещи (к примеру, предметы материального мира, созданные людьми). Развитие проекта EventOnto позволит в дальнейшем представить также сопоставительно-статистическую информацию о характеристиках предметных экземпляров в текстах различных жанров и функциональных стилей.

Актанты и вещи, используемые в тексте в комбинации с предлогами или наречиями, образующими в совокупности семантическую единицу места действия (в комнате) или его пространственного направления (в комнату, из комнаты), образуют класс экземпляров «локации» (в тексте выделено 114 единиц). Данные единицы в своей новой роли получают новое уникальное именование, основанное на другой классификации.

L пространство	AA0000 уник.№	W путь	R подл.	N ест.
		P место	A вым.	M иск.

Рисунок 3 - Классификация предметных единиц пространства

Ключевым параметром, связанным с характеристиками токена места или направления, является разграничение пространства на два типа: путь (W) и место (P). Один и тот же объект может дублироваться в различных событийных контекстах: «на Лукоморье» (LAA0022PAM) и «по ул. Лукоморье» (LAA0022WAM). Характеристики по достоверности («подлинное» (R) / «вымышленное» (A)) и по происхождению («естественное» (N) / «искусственное» (M)) наследуются от характеристик корневого по отношению к пространству актанта или объекта («Ул. Лукоморье», OAA0048AATOM).

Таким образом, главный герой текста, «Александр Иванович Привалов» (SAA0000AATON), определяется как актант, обладающий способностью инициировать действия и процессы, является вымышленным персонажем текста, созданным автором текста и не имеющим известного прототипа в действительности, но представляющим типичный класс («человек»), не обладающим при этом нетипичными для человека свойствами. «Кот Василий» (SAA0004AATSN) также относится к числу актантов, является вымышленным персонажем, не имеющим отношения к какому-либо известному прототипу в действительности, представляющим типичный класс («кот»), но обладающим нетипичным свойством (умение говорить). «Солнце» (OAA0007RN) является реальным объектом естественного происхождения, а «Спальный мешок» (OAA0021AATOM) – вымышленным объектом, созданным автором, типичным, обыкновенным предметом материальной культуры, созданным людьми.

Обсуждение

События, формирующие онтологию EventOnto, в весьма обобщенном смысле содержат в себе информацию «кто» / «с кем» > «используя что» / «из чего» / «с чем» и т.п. > «где» > и, в конечном счете, «что сделал». Массив предметных экземпляров позволяет сформировать базу для ответов на первый ряд этих вопросов. Выделенные характеристики предметных экземпляров позволят в дальнейшем охарактеризовать достоверность запечатленной в событиях текстов действительности и распределить её на три типа реальности: подлинную, вымышленную и дополненную. К последнему типу реальности будут относиться вымышленные события, в которые были добавлены реальные лица, предметы или пространства, а также подлинные события, к участию в которых добавлены вымышленные лица и предметы, либо действия перенесены в иные локации (вымышленные, либо реальные, но не достоверные). Глубина вымысла «дополненной реальности» в том или ином тексте может быть детально конкретизирована, что коррелирует с качеством классификации предметных экземпляров онтологии и моделирования событий.

Заключение

Известные на данный момент семантические корпуса можно условно распределить на две ключевые группы: во-первых, корпуса, в которых представлено семантическое аннотирование отдельных слов (к их числу на данный момент относится и Национальный корпус русского языка), но в которых не представлено никакой информации об их связи внутри текстов; во-вторых, корпуса, выражающие смысловые отношения между словами, представляющие собой семантические структуры, построенные в соответствии с определенной лингвистической теорией или теоретико-нейтральные [3]. Проект EventOnto по существу не относится ни к одному из представленных типов и выходит за рамки устоявшейся парадигмы формальной семантики, фокусируя свое внимание не на языке (как структуре) и его способности в тех или иных комбинациях единиц выражать те или иные смыслы, но на концептуализации запечатленной в различных текстах событийной информации (отдельные события, их структура, связь и отношения друг с другом). Сформулированная в формате машиночитаемых RDF-утверждений информация о структурно-событийном содержании текстов позволит в дальнейшем создать комплекс поисковых инструментов в аннотированном корпусе. Данные инструменты открывают возможность формулировать запросы по корпусу или избранным подкорпусам, касающиеся событийно-сюжетной организации текстов, структуре и связей запечатленных в текстах событий, их интертекстуальных пересечений, специфику вымысла отдельных событий и их соотношение с действительностью, а также связанную статистическую информацию. Качество поисковых результатов во многом коррелирует со спецификой классификации предметных экземпляров текста, представленной в данной работе. Сформулированная классификация предметных единиц текста в дальнейшем может быть дополнена в процессе работы над пятым этапом реализации проекта EventOnto, что будет отражено в публикациях по результатам завершения описанного этапа.

Финансирование

Статья опубликована за счет средств внутривузовского научного гранта ФГБОУ ВО «Марийский государственный университет» в рамках проекта «Речевые стратегии в художественном дискурсе цифровой эпохи метамодернизма».

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.

Funding

The article was published with funds from an intra-university scientific grant from the Mari State University as part of the project "Speech Strategies in the Artistic Discourse of the Digital Age metamodernism".

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. Alvez J. Commonsense Reasoning Using WordNet and SUMO: a Detailed Analysis / J. Alvez, I. Gonzalez-Dios, G. Rigau. — 2019. — URL: https://www.researchgate.net/publication/335651329_Commonsense_ReasoningUsing_WordNet_and_SUMO_a_Detailed_Analysis (accessed: 15.04.2024)
2. Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог» (Москва, 16-19 июня 2021 г). — Москва : РГГУ, 2021. — Вып. 20 (27).
3. Boguslavsky I. Semantic Representations in Computational and Theoretical Linguistics: the Potential for Mutual Enrichment / I. Boguslavsky, V. Dikonov, E. Inshakova [et al.] — 2021. — P. 127-141. — DOI: 10.28995/2075-7182-2021-20-127-141
4. Flidner G. Deriving FrameNet Representations: Towards Meaning-Oriented Question Answering / G. Flidner // Natural Language Processing and Information Systems. NLDB 2004. Lecture Notes in Computer Science / F. Meziane, E. Métais. — 2004. — Vol 3136. — DOI: 10.1007/978-3-540-27779-8_6
5. Kipper K. A large-scale classifications of English verbs / K. Kipper, A. Korhonen, N. Ryant [et al.] // Language Resources and Evaluation Journal. — 2008. — Vol. 42. — № 1. — P. 21-40
6. Ruppenhofer J. FrameNet II: Extended Theory and Practice / J. Ruppenhofer, M. Ellsworth, R.L. Petruck Miriam [et al.] — URL: https://framenet.icsi.berkeley.edu/the_book (accessed 15.04.2024)
7. Валгина Н.С. Современный русский язык: Учебник / Н.С. Валгина, Д.Э. Розенталь, М.И. Фомина. — М.: Логос, 2002. — 528 с.
8. Карточка проекта фундаментальных и поисковых научных исследований поддержанного российским научным фондом № 16-18-10422. — URL: <https://rscf.ru/project/16-18-10422/> (дата обращения: 15.04.2024).
9. Ляшевская О.Н. Русский фреймнет: к задаче создания корпусного словаря конструкций / О.Н. Ляшевская, Ю.Л. Кузнецова // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27-31 мая 2009 г). — М.: РГГУ, 2009. — Вып. 8 (15). — С. 306-312.
10. Макарова А.А. Формальное представление семантики художественного текста на базе онтологического подхода (на примере повести братьев Стругацких «Понедельник начинается в субботу») / А.А. Макарова, А.А. Розина // Молодой исследователь: от идеи к проекту: материалы VII студенческой научно-практической конференции. — Йошкар-Ола: Марийский гос. ун-т, 2023. — С. 705-707.
11. Михеев А.А. Концепция моделирования семантической разметки корпусов текстов на базе событийной онтологии / А.А. Михеев, Е.С. Ярыгина // Russian Linguistic Bulletin. — 2024. — №1(49). — DOI: 10.18454/RULB.2024.49.37
12. Национальный корпус русского языка. — URL: www.ruscorpora.ru (дата обращения: 10.04.2024).
13. Стругацкий А.Н. 1967-1968. Сказка о Тройке. Сказка о Тройке-2. Обитаемый остров / А.Н. Стругацкий, Б.Н. Стругацкий. — М.: АСТ, 2019. — 608 с.
14. Стругацкий А.Н. Понедельник начинается в субботу (Сказка для научных сотрудников младшего возраста) / А.Н. Стругацкий, Б.Н. Стругацкий. — М.: АСТ, 2023. — 352 с.

Список литературы на английском языке / References in English

1. Alvez J. Commonsense Reasoning Using WordNet and SUMO: a Detailed Analysis / J. Alvez, I. Gonzalez-Dios, G. Rigau. — 2019. — URL: https://www.researchgate.net/publication/335651329_Commonsense_ReasoningUsing_WordNet_and_SUMO_a_Detailed_Analysis (accessed: 15.04.2024)
2. Komp'juternaja lingvistika i intelektual'nye tehnologii: Po materialam ezhegodnoj mezhdunarodnoj konferencii «Dialog» (Moskva, 16-19 ijunja 2021 g) [Computer linguistics and intellectual technologies: Proceedings of the annual international conference "Dialogue" (Moscow, 16-19 June 2021)]. — Moscow : RSGU, 2021. — Iss. 20 (27). [in Russian]
3. Boguslavsky I. Semantic Representations in Computational and Theoretical Linguistics: the Potential for Mutual Enrichment / I. Boguslavsky, V. Dikonov, E. Inshakova [et al.] — 2021. — P. 127-141. — DOI: 10.28995/2075-7182-2021-20-127-141
4. Flidner G. Deriving FrameNet Representations: Towards Meaning-Oriented Question Answering / G. Flidner // Natural Language Processing and Information Systems. NLDB 2004. Lecture Notes in Computer Science / F. Meziane, E. Métais. — 2004. — Vol 3136. — DOI: 10.1007/978-3-540-27779-8_6

5. Kipper K. A large-scale classifications of English verbs / K. Kipper, A. Korhonen, N. Ryant [et al.] // Language Resources and Evaluation Journal. — 2008. — Vol. 42. — № 1. — P. 21-40
6. Ruppenhofer J. FrameNet II: Extended Theory and Practice / J. Ruppenhofer, M. Ellsworth, R.L. Petruck Miriam [et al.] — URL: https://framenet.icsi.berkeley.edu/the_book (accessed 15.04.2024)
7. Valgina N.S. Sovremennyj russkij jazyk: Uchebnik [Modern Russian Language: Textbook] / N.S. Valgina, D.Je. Rozental', M.I. Fomina. — M.: Logos, 2002. — 528 p. [in Russian]
8. Kartochka proekta fundamental'nyh i poiskovyh nauchnyh issledovanij podderzhannogo rossijskim nauchnym fondom № 16-18-10422 [Card of the project of fundamental and search scientific research supported by the Russian Science Foundation № 16-18-10422]. — URL: <https://rscf.ru/project/16-18-10422/> (accessed: 15.04.2024). [in Russian]
9. Ljashevskaja O.N. Russkij frejmnjet: k zadache sozdanija korpusnogo slovarja konstrukcij [Russian framenet: to the task of creating a corpus dictionary of constructions] / O.N. Ljashevskaja, Ju.L. Kuznecova // Komp'juternaja lingvistika i intellektual'nye tehnologii: Po materialam ezhegodnoj Mezhdunarodnoj konferencii «Dialog 2009» (Bekasovo, 27-31 maja 2009 g) [Computer linguistics and intellectual technologies: Proceedings of the annual International Conference "Dialogue 2009" (Bekasovo, May 27-31, 2009)]. — M.: RSGU, 2009. — Iss. 8 (15). — P. 306-312. [in Russian]
10. Makarova A.A. Formal'noe predstavlenie semantiki hudozhestvennogo teksta na baze ontologičeskogo podhoda (na primere povesti brat'ev Strugackih «Ponedel'nik nachinaetsja v subbotu») [Formal representation of the semantics of the fiction text on the basis of the ontological approach (on the example of the story "Monday Begins on Saturday" by Strugatsky brothers)] / A.A. Makarova, A.A. Rozinova // Molodoj issledovatel': ot idei k proektu: materialy VII studencheskoj nauchno-praktičeskoj konferencii [Young researcher: from idea to project: materials of the VII Student Scientific and Practical Conference]. — Yoshkar-Ola: Mari State University, 2023. — P. 705-707. [in Russian]
11. Miheev A.A. Konceptija modelirovanija semantičeskoj razmetki korpusov tekstov na baze sobytijnoj ontologii [The concept of modelling semantic partitioning of text corpora on the basis of event ontology] / A.A. Miheev, E.S. Jarygina // Russian Linguistic Bulletin. — 2024. — №1(49). — DOI: 10.18454/RULB.2024.49.37 [in Russian]
12. Nacional'nyj korpus russkogo jazyka [National corpus of the Russian language]. — URL: www.ruscorpora.ru (accessed: 10.04.2024). [in Russian]
13. Strugackij A.N. 1967-1968. Skazka o Trojke. Skazka o Trojke-2. Obitaemyj ostrov [Tale of the Troika. Tale of Troika 2. Prisoners of Power] / A.N. Strugackij, B.N. Strugackij. — M.: AST, 2019. — 608 p. [in Russian]
14. Strugackij A.N. Ponedel'nik nachinaetsja v subbotu (Skazka dlja nauchnyh sotrudnikov mladshego vozrasta) [Monday Begins on Saturday (A Tale for Early Career Researchers)] / A.N. Strugackij, B.N. Strugackij. — M: AST, 2023. — 352 p. [in Russian]