

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»

На правах рукописи

Зиновьева Анастасия Юрьевна

МОДЕЛЬ МНОГОЯЗЫЧНОГО ИНТЕЛЛЕКТУАЛЬНОГО
КОНТЕНТ-АНАЛИЗА (НА МАТЕРИАЛЕ АНГЛО-, ФРАНКО-
И РУССКОЯЗЫЧНЫХ НОВОСТНЫХ СООБЩЕНИЙ
О ТЕРРОРИСТИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ)

Специальность 10.02.21 — Прикладная и математическая лингвистика

Диссертация на соискание ученой степени
кандидата филологических наук

Научный руководитель:
доктор филологических наук, доцент
Шереметьева Светлана Олеговна

Челябинск — 2022

ОГЛАВЛЕНИЕ

Введение.....	5
Глава 1. Интеллектуальный контент-анализ как объект моделирования.....	14
1.1 Модель и моделирование в лингвистических исследованиях	14
1.2 Основы контент-анализа как метода исследования	16
1.2.1 История развития и определение контент-анализа	16
1.2.2 Типы контент-анализа	21
1.3 Интеллектуальный контент-анализ.....	23
1.3.1 Стандартная процедура качественного контент-анализа	23
1.3.2 Определение и процедура интеллектуального контент-анализа	26
1.3.3 Ресурсы интеллектуального контент-анализа	29
1.4 Онтология как способ формализации знаний для интеллектуального контент-анализа	31
1.4.1 Определение и структура онтологии	31
1.4.2 Классификация онтологических ресурсов	35
1.4.3 Особенности разработки предметных онтологий	44
1.4.4 Особенности разработки онтологий терроризма.....	49
Выводы по главе 1	57
Глава 2. Корпусный анализ подъязыка новостных сообщений предметной области «Терроризм» на материале русского, английского и французского языков	61
2.1 Предварительные замечания.....	61
2.2 Методика анализа подъязыка	62
2.3 Результаты анализа русского корпуса	66
2.4 Результаты анализа английского корпуса	80

2.5 Результаты анализа французского корпуса.....	89
Выводы по главе 2.....	97
Глава 3. Построение модели многоязычного интеллектуального контент-анализа новостных сообщений предметной области «Терроризм».....	101
3.1 Методологические основы построения модели.....	101
3.2 База знаний модели.....	102
3.2.1 Онтология предметной области «Терроризм».....	103
3.2.2 База экземпляров.....	115
3.2.3 Онтолексиконы и ономастиконы.....	116
3.2.4 Правила онтологического анализа.....	119
3.2.5 Проблема концептуальной неоднозначности при онтоанализе.....	120
3.2.6 Прочие компоненты базы знаний.....	129
3.3 Алгоритм модели.....	130
3.4 Платформа концептуального аннотирования.....	132
3.5 Примеры использования модели.....	142
3.5.1 Пример первый: анализ тенденций в англоязычном корпусе.....	143
3.5.2 Пример второй: анализ репрезентации террористов на материале русскоязычного корпуса.....	147
3.5.3 Пример третий: репрезентация отдельного теракта в русско-, англо- и франкоязычных СМИ.....	152
Выводы по главе 3.....	157
Заключение.....	160
Список литературы.....	164

Приложение А Морфологическая репрезентация глаголов в подязыке новостных сообщений предметной области «Терроризм»	183
Приложение Б Концептуальные классы лексических единиц предметной области «Терроризм».....	186
Приложение В Иерархическая структура концептов онтологии терроризма	200
Приложение Г Фреймы концептов онтологии терроризма	207
Приложение Д Фреймы экземпляров онтологии терроризма	250
Приложение Е Аксиомы онтологии предметной области «Терроризм»	254

ВВЕДЕНИЕ

Настоящая диссертационная работа посвящена методологическим и практическим аспектам моделирования интеллектуального контент-анализа неструктурированной текстовой информации на примере новостных сообщений предметной области «Терроризм» на английском, французском и русском языках.

Актуальность исследования обусловлена тем, что наблюдаемый в настоящее время постоянный рост объема неструктурированной информации на различных языках требует систем автоматизации ее анализа, среди которых первостепенное значение имеют системы интеллектуального контент-анализа, которые могли бы обеспечить высокое качество извлечения из текстов соответствующего информационному запросу пользователя контента. Создание таких систем невозможно без решения проблем моделирования процессов интеллектуального контент-анализа, которые, несмотря на то что в нашей стране и за рубежом исследованиям в этой области уделяется серьезное внимание, еще не решены.

Настоящее исследование актуально и в том отношении, что в центре его внимания находится многоязычность, поскольку многоязычная модель интеллектуального контент-анализа делает возможным ее повторное использование для обработки текстов на различных языках с минимизацией времени, затрат и усилий разработчиков. Актуальность исследования определяется также тем, что оно охватывает не только английский и французский языки, на которые ориентированы многие исследования в данной области, но и русский, интеллектуальному контент-анализу которого до настоящего времени еще не уделяется достаточно внимания, что отражено в значительно меньшем количестве работ по теме.

Разрабатываемая модель многоязычного интеллектуального контент-анализа ориентирована на предметную область «Терроризм», что также обеспечивает актуальность настоящей работы, поскольку борьба с терроризмом является одной из главных задач современности, и контент-анализ текстов данной предметной области имеет большое значение для аналитической деятельности в сфере политического прогнозирования и контртерроризма.

Степень разработанности проблемы. Существующие исследования по разработке моделей и систем интеллектуального контент-анализа можно разделить на два крупных класса: исследования, ориентированные на обработку текстов любой тематики, в результате чего страдает глубина извлечения информации из узконаправленных текстов; и исследования, посвященные решению точечных задач интеллектуального контент-анализа в конкретной предметной области и на конкретном языке, отсюда невозможность применения таких систем контент-анализа к текстам других предметных областей и на других языках. Наиболее реалистичным на данный момент и поэтому наиболее разрабатываемым направлением исследований является построение моделей интеллектуального контент-анализа для конкретных областей на материале конкретных языков. Исследований по разработке многоязычных моделей интеллектуального контент-анализа значительно меньше, и подавляющее большинство из них также ориентировано на конкретный подязык.

Большинство доступных зарубежных работ по интеллектуальному контент-анализу нацелены на обработку английского или иного национального (не русского) языка, например системы Intelligent Content Analysis, OntoText, работы А. С. М. Fong (2006), I. Flaounas (2010), L. Weber (2013) и др. Среди исследований, близких по тематике к нашему и выполняемых на русскоязычном материале, можно отметить работы, посвященные системам (как одноязычным, так и многоязычным) ВААЛ, «Гитика», АBBYY Compreno, АИРЕ, NearIdx, OntosMiner, Texterra. Проект NearIdx нацелен на обработку русского языка в широкой предметной области «Национальная безопасность» (включая раздел «Терроризм») в основном в целях информационного поиска, но также в нем рассматриваются определенные аспекты формализации естественного языка, релевантные для интеллектуального контент-анализа, в частности рубрикация и фасетный анализ.

В приведенных работах представлены различные подходы к ключевому для интеллектуального контент-анализа этапу категоризации текстовых единиц и формализации экспертных знаний. Наиболее распространенным является подход на основе онтологий, позволяющий проводить глубокий семантический анализ.

Несмотря на большое количество исследований по рассматриваемой теме, проблема моделирования интеллектуального контент-анализа не может считаться решенной в полной мере ввиду комплексности, неоднозначности естественного языка и сложности его формализации.

Цель исследования состоит в разработке модели многоязычного интеллектуального контент-анализа на основе онтологических знаний на материале новостных сообщений предметной области «Терроризм» на английском, французском и русском языках с возможностью последующего ее применения к другим языкам и предметным областям.

Поставленной целью продиктованы следующие **задачи исследования**:

- 1) уточнить понятие интеллектуального контент-анализа и наиболее перспективные пути его реализации;
- 2) разработать методику создания модели многоязычного интеллектуального контент-анализа ограниченной предметной области;
- 3) выявить ограничения и закономерности подъязыка новостных сообщений предметной области «Терроризм» на уровне структуры релевантности и суперструктуры, морфосинтаксическом и лексико-семантическом уровнях на материале английского, французского и русского языков;
- 4) построить онтологическую базу знаний модели многоязычного интеллектуального контент-анализа на основе данных анализа подъязыка;
- 5) разработать алгоритм интеллектуального контент-анализа;
- 6) апробировать модель на материале новых текстов английского, французского и русского языков, не относящихся к тренировочному корпусу.

Объектом исследования является подъязык новостных сообщений предметной области «Терроризм» на английском, французском и русском языках.

Предметом исследования является моделирование концептуальной структуры рассматриваемого подъязыка и формализация извлечения проблемно-ориентированного контента из текстов предметной области.

В качестве **материала исследования** для построения модели многоязычного интеллектуального контент-анализа были использованы три псевдопараллельных

(сопоставимых) корпуса новостных интернет-сообщений о террористической деятельности за 2014–2020 гг. равного объема общим объемом более 600 000 словоупотреблений, для доработки модели — корпуса текстов за 2019–2020 гг. объемом 20–40 тыс. словоупотреблений. Все тексты были получены методом целевой выборки из интернет-источников по ключевым словам. В качестве источников были выбраны сайты англо-, франко- и русскоязычных информационных агентств и газет BBC News, Bloomberg, Reuters, Sky News, 20 Minutes, AFP, L'Express, Le Figaro, RT France, Sputnik France, «Вести», «Интерфакс», «Лента», «НТВ», «РИА Новости», «ТАСС» и др. На отдельных этапах исследования использовались случайные выборки из вышеуказанных корпусов, концептуально размеченные сотрудниками научно-образовательного центра «Лингво-инновационные технологии» Южно-Уральского государственного университета, в числе которых был автор настоящей работы. Дополнительным материалом для проверки работы модели послужили новые корпуса текстов на трех языках разного объема за 2019–2020 гг.

Материал исследования представлен корпусами текстов на трех языках как по лингвистическим, так и по экстралингвистическим причинам. Во-первых, в настоящее время английский, французский и русский являются одними из самых распространенных языков в мире, официальными языками ООН. Во-вторых, эти языки принадлежат к разным языковым группам и, как следствие, имеют разные характеристики. В частности, различия наблюдаются на морфосинтаксическом уровне: английский язык является в большей степени аналитическим, русский язык — синтетическим, французский язык занимает промежуточное положение, сохраняя черты синтетизма в своей письменной форме. Наконец, использование в исследовании трех языков обусловлено необходимостью создания независимой от конкретного естественного языка онтологии и, следовательно, модели интеллектуального контент-анализа, что позволит в будущем с минимальными усилиями разработчиков распространить ее на другие языки.

Научная новизна работы заключается в том, что впервые разработаны методологические и практические аспекты моделирования интеллектуального контент-анализа неструктурированной информации на примере новостных сообщений

предметной области «Терроризм» на английском, французском и русском языках, что обусловлено применением совокупности современных лингвистических и компьютерных методов к анализу языкового материала; выявлением и уточнением ограничений и закономерностей подязыка новостных сообщений предметной области «Терроризм» на уровне структуры релевантности и суперструктуры, морфосинтаксическом и лексико-семантическом уровнях; исследованием проблемы концептуальной неоднозначности; разработкой предметно-ориентированной модели контент-анализа с акцентом на интеллектуальность и многоязычность; построением многоязычной онтологии предметной области «Терроризм»; созданием онтолексиконов для английского, французского и русского языков; разработкой правил онтологического анализа, логического вывода и формирования динамических концептуально-лексических фреймов для представления результатов контент-анализа; а также разработкой алгоритма интеллектуального контент-анализа, который позволяет извлекать не только явно выраженную, но и имплицитную информацию.

Теоретическая значимость исследования состоит в развитии методологических аспектов многоязычного интеллектуального контент-анализа, в том числе в уточнении определения интеллектуального контент-анализа, ввиду чего результаты исследования способны внести вклад как в данное направление научных исследований, так и в смежные с ним направления: лексикографию, лексикологию и терминоведение, компьютерную и корпусную лингвистику, информационный поиск, автоматическое реферирование и аннотирование, машинный перевод и т. п.

Практическая ценность результатов работы заключается в возможности создания системы многоязычного интеллектуального контент-анализа на базе разработанной модели. Потенциальная область применения модели не ограничивается новостными интернет-сообщениями о террористической деятельности; лежащие в ее основе принципы могут быть экстраполированы на другие языки и предметные области. Отдельные положения могут быть включены в курсы по прикладной лингвистике, функциональной стилистике, общественно-политическому переводу.

В работе использованы следующие **методы исследования**: метод лингвистического моделирования, методы целевой и случайной выборки для создания

исходных и тестовых корпусов, статистический и сопоставительный анализ, компонентный анализ, дистрибутивный анализ, метод оппозиций, контекстный метод.

Теоретико-методологическую базу исследования составили работы Ю. Д. Апресяна (1966), И. И. Ревзина (1962, 1966, 1977), С. О. Шереметьевой (2000, 2017, 2018, 2020) в области *лингвистического моделирования*; положения В. П. Захарова (2016) в области *корпусной лингвистики*, работы А. А. Бызова (2019), Ю. П. Воронова (2005); В. И. Шалака (2002), В. Berelson (1952), С. Cioffi-Revilla (2014), L. A. Kort-Butler (2016), К. Krippendorff (1967, 1980, 2018) и др. в области *контент-анализа*, в том числе работы С. О. Шереметьевой (2020), А. С. М. Fong (2006), Ph. Mayring (2014), L. Weber (2013) и др. в области *интеллектуального контент-анализа*; работы по *неоднозначности* Л. Н. Иорданской (1967), В. Н. Полякова (2004), Е. В. Рахилиной и др. (2006); *построению одноязычных и многоязычных баз знаний для автоматической обработки текстов* Е. Б. Козеренко (1995), С. О. Шереметьевой (1998, 1999, 2012); *онтологической семантике* С. Ниренбурга и В. Раскина (2004), К. Mahesh (1996), В. Onyshkevych (1997); *построению онтологий* И. М. Богуславского (2012), Н. В. Лукашевич (2011), О. А. Митрофановой и Н. С. Константиновой (2015), Т. Gruber (1993), N. Guarino (1997), О. Lassila и D. McGuinness (2001) и др.; *извлечению информации на основе онтологических знаний* С. О. Шереметьевой (2020), А. Konys (2015), Е. Iosif (2012); а также прикладные исследования по разработке *онтологий терроризма* С. Л. Мишлановой (2012), А. Mannes (2005, 2007), М. D. Turner (2011), L. Wendelberg (2021) и др.; работы в области исследования *подъязыков* Н. Д. Андреева (1971), С. О. Шереметьевой (1994, 2017), Z. Harris (1968), J. Lehrberger (1982), в том числе *подъязыка новостных текстов* Т. Г. Добросклонской (2008), М. К. Özgüven и J. Tsujii (1992), Т. А. van Dijk (1988). В целях автоматизации исследования применялись универсальный экстрактор лексики, разработанный С. О. Шереметьевой (2012), и корпус-менеджер AntConc авторства L. Anthony (2020).

На защиту выносятся следующие **положения**:

1. Интеллектуальный контент-анализ представляет собой извлечение из корпусов текстов соответствующего информационному запросу пользователя

контента, его интерпретацию и представление в удобной для пользователя форме; включает качественный этап категоризации текстовых единиц и количественный этап подсчета выделенных категорий.

2. Создание модели многоязычного интеллектуального контент-анализа текстов ограниченной предметной области для обеспечения высокого качества результатов осуществляется на основе онтологической базы знаний.

3. Построение онтологической базы знаний осуществляется на основе анализа подъязыка соответствующей предметной области на материале многоязычных текстов и предусматривает последовательное выполнение трех стадий разработки, которые включают в себя циклические этапы и подэтапы: анализ лингвистического материала рассматриваемой предметной области, создание исходной модели многоязычного интеллектуального контент-анализа на основе результатов анализа лингвистического материала и совершенствование созданной модели с использованием дополнительного текстового материала предметной области.

4. Подъязык новостных сообщений предметной области «Терроризм» ограничен, его ограничения имеют двойственную природу, что обусловлено жанром новостных сообщений и предметной областью «Терроризм».

5. База знаний модели многоязычного интеллектуального контент-анализа состоит из концептуальных (онтология предметной области), эпизодических (база экземпляров) и лингвистических (онтолексиконы, ономастиконы) знаний, правил онтологического анализа и логического вывода, динамических концептуально-лексических фреймов для представления результатов контент-анализа.

6. Алгоритм модели многоязычного интеллектуального контент-анализа включает этап постановки задачи контент-анализа и набор процедур для ее последовательного решения на основе сочетания автоматических и ручных методов, а именно отбор текстового материала, онтологический анализ, определение релевантных концептуальных тегов, заполнение концептуально-лексических фреймов, представление результатов контент-анализа.

Достоверность и научная обоснованность теоретических и практических результатов обеспечивается опорой на авторитетные положения работ

отечественных и зарубежных ученых, применением современных методов исследования, адекватных его цели и задачам; использованием в качестве фактического материала англо-, франко- и русскоязычных корпусов текстов значительного объема; согласованностью теоретических выводов с практическими результатами. Научные положения и выводы, сформулированные в работе, подкреплены фактическими данными, представленными в приведенных таблицах и рисунках.

Апробация работы. Основные положения исследования неоднократно обсуждались на заседаниях кафедры лингвистики и перевода ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)», а также были представлены на конференциях различного уровня: конференции аспирантов и докторантов ЮУрГУ (Челябинск, 2019–2021), международной конференции «Цифровые трансформации и глобальное общество» (Санкт-Петербург, 2018), 14-ой Международной научно-технической конференции «Интерактивные системы: проблемы человеко-компьютерного взаимодействия» (Ульяновск, 2019), конференции «Интернет и современное общество» (Санкт-Петербург, 2020), 64-й Всероссийской научной конференции МФТИ (Москва, 2021), Международной научно-практической конференции «Язык. Культура. Перевод: межкультурная коммуникация в цифровую эпоху» (Одинцово, 2022). Основные положения исследования отражены в 10 печатных работах, четыре из которых опубликованы в журналах, рекомендованных ВАК, три публикации — в изданиях, индексируемых международной базой Scopus.

Диссертационная работа состоит из введения, трех глав, заключения, списка использованной литературы и приложений. Объем основного текста диссертации составляет 163 страницы. Диссертация содержит 21 рисунок, 35 таблиц. Список литературы состоит из 175 источников, из них 109 — на иностранных языках.

Во **введении** указывается цель работы, обосновывается ее актуальность, новизна, теоретическая и практическая значимость, приводятся задачи, предмет и объект исследования, описывается и обосновывается материал исследования, используемые методы, указываются положения, выносимые на защиту, и приводятся сведения об апробации результатов работы.

В первой главе «Интеллектуальный контент-анализ как объект моделирования» рассматриваются общие проблемы определения и моделирования интеллектуального контент-анализа и применения онтологий в качестве основы базы знаний модели интеллектуального контент-анализа; анализируются практические работы по созданию онтологий предметной области «Терроризм».

Во второй главе «Корпусный анализ подязыка новостных сообщений предметной области „Терроризм“ на материале русского, английского и французского языков» проводится анализ подязыка новостных сообщений предметной области «Терроризм»; по результатам проведенного анализа определяются разноуровневые характеристики подязыка, необходимые для разработки базы знаний модели интеллектуального контент-анализа, а также основные концептуальные классы предметной области для будущего преобразования их в концепты онтологии.

В третьей главе «Построение модели многоязычного интеллектуального контент-анализа новостных сообщений предметной области „Терроризм“» описывается разработанная модель и ее компоненты: база знаний и алгоритм, а также инструмент для автоматизации работы модели; демонстрируется работа модели на примере интеллектуального контент-анализа новостных сообщений о террористической деятельности на английском, французском и русском языках.

В заключении излагаются выводы по проведенному исследованию.

Приложения содержат иллюстративный материал для ряда практических положений диссертации.

ГЛАВА 1.

ИНТЕЛЛЕКТУАЛЬНЫЙ КОНТЕНТ-АНАЛИЗ КАК ОБЪЕКТ МОДЕЛИРОВАНИЯ

1.1 Модель и моделирование в лингвистических исследованиях

В широком смысле под *моделированием* подразумевается создание модели некоего объекта познания для изучения его свойств и характеристик. В Большом энциклопедическом словаре моделирование описывается как «одна из основных категорий познания, [на идее которой] по существу базируется любой метод научного исследования — как теоретический (при котором используются различного рода знаковые, абстрактные модели), так и экспериментальный (использующий предметные модели)» [Большой энцикл. словарь, 2004].

Продуктом моделирования является *модель* — этот термин появился в научном обиходе благодаря работам математиков Э. Бельтрами и Ф. Клейна и был заимствован в языкознание американским лингвистом З. Харрисом, который впервые использовал его для описания методологии Э. Сепира в середине XX века.

Несмотря на частое применение моделей в лингвистических исследованиях, этому термину довольно трудно подобрать определение, что не раз отмечалось учеными [Лосев, 2004; Ревзин, 1962]. В частности, И. И. Ревзин описывает модель как «некоторое гипотетическое научное построение, некоторый конструкт» [Ревзин, 1962, с. 9] и предлагает характеризовать модель через ее свойства, которые заключаются в соответствии исходных данных модели некоторой существенной части совокупности исходных объектов, а ее итоговых данных — некоторой существенной части совокупности заключительных объектов [Ревзин, 1967, с. 25]. В более поздней работе И. И. Ревзин также отмечает такое существенное свойство модели, как ее упрощенный, огрубленный характер [Ревзин, 1977, с. 62].

Я. Г. Неуймин в работе «Модели в науке и технике» рассматривает понятие обобщенной научной модели, которая состоит из четырех компонентов: субъект, решаемая им задача, объект-оригинал и язык описания (способ материального

воспроизведения модели) [Неуймин, 1984, с. 45]. Как отмечает З. И. Комарова, такое понимание модели указывает на ее приблизительный характер, обусловленный тем, что в целях описания одного объекта познания для решения разных задач могут быть созданы сразу несколько моделей [Комарова, 2014]. З. И. Комарова также приводит ряд свойств, которые характеризуют обобщенную научную модель: субъективность, двойственность (модель как репрезентует исследуемый объект-оригинал, так и сама подвергается исследованию), трансформируемость, компактность, специфическая информативность (абстрактная форма модели позволяет выделить свойства, не представленные в оригинале) [Там же, с. 314].

Лингвистические модели в свою очередь обладают особой спецификой. По определению, приведенному в Лингвистическом энциклопедическом словаре, лингвистическая модель — это «искусственно созданное лингвистом реальное или мысленное устройство, воспроизводящее или имитирующее своим поведением (обычно в упрощенном виде) поведение какого-либо другого устройства (т. е. оригинала) в лингвистических целях» [Лингв. энцикл. словарь, 1990, с. 304].

Исследователь В. В. Доброва пишет о необходимости разграничения понятий лингвистической и языковой модели: «Языковая модель является естественно-языковым образцом, по которому построен ряд языковых образований, а лингвистическая модель — это теоретическое построение, описывающее то или иное языковое явление, в том числе и языковую модель. Лингвистическая модель [...] создана в целях лингвистического исследования» [Доброва, 2015, с. 146].

Профессор Ю. Д. Апресян предлагает разделение лингвистических моделей на три типа в зависимости от характера моделируемого лингвистического объекта: *модели речевой деятельности* воспроизводят конкретные процессы и явления языка, речевую деятельность человека; *модели лингвистического исследования* воспроизводят исследовательские процедуры; *метамодели*, или метатеории, воспроизводят готовые лингвистические описания [Апресян, 1966, с. 99].

В последние несколько десятилетий набирает популярность построение многоязычных моделей с целью моделирования лингвистических объектов и явлений на нескольких языках [Козеренко, 1995; Montiel-Ponsoda, 2011; Sheremetyeva, 2000,

2017, 2018, 2020]. Ориентация на многоязычные модели позволяет экономить исследовательские усилия при создании новых ресурсов [Sheremetyeva, 2017].

Среди многообразия моделей, используемых в лингвистике, особый интерес в рамках настоящей работы представляют компьютерные модели, основанные на алгоритмическом представлении последовательности действий. Компьютерные модели считаются особенно эффективными в тех случаях, когда исследование включает обработку большого количества информации [Доброва, 2015].

Реализация метода моделирования в прикладной лингвистике предполагает последовательное решение следующих задач, которым мы намереваемся следовать в ходе настоящего исследования:

- 1) формулировка проблемы, постановка цели и задач моделирования;
- 2) анализ проблемной области;
- 3) формирование метаязыка описания предметной области (ПО);
- 4) применение метаязыка и получение итогового представления ПО;
- 5) проверка модели: оценка ее объяснительной и предсказательной силы, компьютерная реализация или эксперимент [Баранов, 2001, с. 10].

Настоящее исследование направлено на разработку многоязычной прикладной компьютерной модели, воспроизводящей исследовательскую процедуру интеллектуального контент-анализа. Структура диссертации соответствует приведенным выше этапам моделирования: в первой главе формулируется проблема, ставятся цели и задачи моделирования; во второй главе проводится анализ проблемной области; остальные задачи решаются в третьей главе.

1.2 Основы контент-анализа как метода исследования

1.2.1 История развития и определение контент-анализа

Предпосылки к формированию контент-анализа как метода оценки текстов обнаруживаются еще в XVIII веке, как отмечает К. Довринг в работе «Quantitative Semantics in 18th Century Sweden» [Dovring, 1954]: к этому времени относится

первый должным образом задокументированный случай количественного анализа текста печатных материалов в Швеции, где анализу были подвергнуты религиозные книги сектантов в попытке определить наличие в них скрытых «опасных мыслей», подрывающих шведскую государственность и позиции официальной религии. В дальнейшем систематический количественный анализ текста, по данным современного исследователя в области контент-анализа К. Криппендорфа, использовался для анализа содержания СМИ и литературных источников в США, странах Европы, в том числе в России [Марков, 1913], однако европейские исследования, «несмотря на более сильную теоретическую базу, [...] почти не повлияли на развитие метода в Северной Америке» [Krippendorff, 1967, с. 9]. При этом именно в США происходило основное развитие методологии контент-анализа.

В истории развития контент-анализа выделяют пять этапов:

1. Количественный анализ газет (конец 1890 — 1920-е гг.), который представляет собой не что иное, как попытки журналистов с помощью простейших подсчетов оценить содержание газетных текстов, причем результаты такого анализа были не лишены предвзятости и оценочных суждений [Там же, с. 11].

2. Изучение общественных ценностей в материалах СМИ (1930-е гг.). Этот этап был обусловлен появлением более мощных по сравнению с газетами средств массовой информации (например, радио), экономическим кризисом и готовностью общественных наук расширять свою теоретико-методологическую базу. На данном этапе в качестве материала контент-анализа начали использовать не только газетные тексты, но и радиопередачи, фильмы, речи политиков и др.

3. Анализ пропаганды (1940-е гг.), связанный с участием США во Второй Мировой войне и характеризовавшийся применением контент-анализа преимущественно для изучения пропагандистских материалов. На данном этапе контент-анализ перестал быть исключительно описательным и начал использоваться как индуктивный метод. В это десятилетие были созданы классические контент-аналитические работы Г. Лассуэлла, который являлся активным сторонником количественного контент-анализа [Lasswell, 1949].

4. Становление контент-анализа как междисциплинарного метода оценки текстов (1950-е гг.). В этот период засекреченные прежде данные стали доступными, и стало возможно оценить надежность результатов, полученных ранее методом контент-анализа. В то же время появилось множество контент-аналитических исследований, решающих более широкий круг задач, чем раньше.

5. Компьютерный контент-анализ (конец 1950-х гг. — настоящее время). Начало этого периода ознаменовалось активным внедрением компьютерных технологий в практику контент-анализа в конце 1950-х — начале 1960-х гг. Именно в это время контент-анализ начал применяться советскими исследователями, причем почти одновременно специалистами в области общественных наук [Баранов, 1967; Ядов, 1968] и прикладной лингвистики [Звегинцев, 1968].

В 1960–80-е гг. были разработаны первые программы для автоматизированного контент-анализа сначала на основе списков релевантных слов [Stone и др., 1966], затем на основе тезаурусов [Miller и др., 1993].

В XXI веке в контент-анализе нашли применение достижения из области искусственного интеллекта: так, появились примеры контент-анализа с применением искусственных нейронных сетей [Fong, 2006]. Кроме того, начавшаяся в это время популяризация Семантической паутины [Berners-Lee, 2001] привлекла интерес исследователей из разных областей науки к онтологиям как универсальным системам хранения и представления знаний. В аспекте развития контент-анализа это привело к появлению работ по интеллектуальному контент-анализу на основе онтологий [Добров, 2015; Weber, 2013], в том числе многоязычных [Li и др., 2007].

В новейшем издании одной из ключевых для методологии контент-анализа работ, монографии К. Криппендорфа «Content Analysis: An Introduction to Its Methodology» подчеркивается, что в последние пятьдесят лет наблюдается экспоненциальный рост контент-аналитических исследований [Krippendorff, 2018]. Ввиду столь широкого распространения в разных сферах науки контент-анализ не имеет единого определения, которое принималось бы всеми исследователями. Далее рассмотрим существующие определения термина «контент-анализ», чтобы выбрать то из них, которое является наиболее подходящим для настоящего исследования.

Впервые термин «контент-анализ» был употреблен в 1941 году в названии работы американских социологов Д. Уэйплса и Б. Берелсона [цит. по Berelson, 1948], а его первое определение было предложено Б. Берелсоном в 1952 году.

Согласно определению Б. Берелсона, контент-анализ — это «метод исследования для объективного, систематического количественного описания явного содержания коммуникации» [Berelson, 1952, с. 18]. В данном определении отражены такие основополагающие, с точки зрения автора, характеристики контент-анализа, как объективность, систематичность, квантитативность и ориентация на изучение явного (*manifest*) содержания коммуникации. Объективность предполагает отсутствие пристрастности; иными словами, контент-анализ должен выполняться на основе заданных правил, а не субъективного видения исследователя. Систематичность обеспечивает регулярное и последовательное применение заданных правил ко всему массиву анализируемых текстов. Ориентация на явное содержание указывает на то, что под содержанием (контентом) Б. Берелсон понимал нечто *содержащееся* в тексте, неотъемлемое от него.

Многие авторы в ранние годы развития контент-анализа настаивали на его квантитативности, в частности [Berelson, 1952; Lasswell, 1949], однако в более поздних формулировках квантитативность как неотъемлемый атрибут контент-анализа, по-видимому, подвергается сомнению и исключается из определения. В 1961 году термин *content analysis* был зафиксирован в словаре Вебстера с определением: «подробное изучение и анализ различных типов коммуникации (газет, радиопередач, пропагандистских фильмов) посредством классификации, приведения к табличной форме и оценки ключевых символов и тем для того, чтобы определить их значение и возможное воздействие» [цит. по Krippendorff, 1967, с. 2].

В 1969 году политолог О. Холсти определил контент-анализ как «любой метод для формулирования выводов посредством объективной и систематической идентификации определенных характеристик сообщений» [Holsti, 1969, с. 14]. В данном определении также отсутствует указание на квантитативность метода и его направленность на изучение явного содержания, что свидетельствует о расширении объема понятия. Кроме того, в интерпретации О. Холсти содержание

рассматривается как свойство источника сообщения, т. е., контент-анализ позволяет делать выводы об источнике сообщения на основе написанного им текста.

В 1980 году К. Криппендорф предложил следующее определение контент-анализа: «метод исследования, позволяющий делать воспроизводимые и достоверные выводы из текстов (или других содержательных материалов) в контексте их использования» [Krippendorff, 1980, с. 18]. Основная идея К. Криппендорфа состоит в том, что содержание возникает в процессе анализа текста относительно некоторого контекста. Согласно идеям К. Криппендорфа, тексты сами по себе не имеют единственного смысла, который может быть в них обнаружен, и в целом не существуют без читателя и (или) аналитика.

В русскоязычных источниках английскому термину *content analysis* соответствует несколько эквивалентов: общеупотребительные «контент-анализ» [Пашиных, 2012], «анализ содержания» [Звегинцев, 1968] и менее распространенные «содержательный анализ» [Пастухов, 2014], «формализованный анализ» [Власова, 2005], «„контентный“ анализ», «текстовый анализ» [Баранов, 1967]. Следует отметить, что термины «контент-анализ» и «анализ содержания» («содержательный анализ») либо рассматриваются как синонимы [Пашиных, 2012; Пастухов, 2014; Таршис, 2002; Чернобровкина, 2011], либо противопоставляются [Шалак, 2002]. В частности, В. И. Шалак подчеркивает, что контент-анализ — не «просто содержательный анализ текстов», а «строгий метод оценки текстов» [Там же, с. 117].

Исходя из вышеизложенного, можно заключить, что в научной литературе существует несколько подходов к определению термина «контент-анализ», которые различаются взглядами на количественность контент-анализа и интерпретацией центрального для данного метода понятия «содержание» («контент»). В свою очередь мы придерживаемся подхода, основанного на использовании количественной меры, и под содержанием (контентом) понимаем то, что содержится в тексте в явном виде или может быть выведено на основе эксплицитного содержания текста с применением фоновых знаний исследователя. Кроме того, вслед за К. Криппендорфом мы считаем, что контент-анализ должен быть воспроизводимым, т. е. должен содержать понятные и однозначные правила кодирования единиц анализа.

1.2.2 Типы контент-анализа

Закономерно наличие множества подходов к интерпретации контент-анализа привело к наличию у данного метода множества вариаций. В работах, посвященных изучению контент-анализа, выделяют несколько параметров, на основе которых может быть произведена его типологизация: количественная мера, объект и предмет, элементы кодирования, степень учета контекста, автоматизации.

И. А. Пашинян подразделяет контент-анализ на количественный и неколичественный по использованию количественной меры. Количественный контент-анализ основан на квантитативности и направлен на получение количественной структуры содержания текста в рамках принятой системы категорий анализа. Неколичественный контент-анализ в свою очередь базируется на нечастотной модели содержания текста, для оценки которого не используются количественные меры; вместо этого фиксируется только наличие элемента содержания [Пашинян, 2012].

Л. А. Корт-Батлер выделяет четыре типа контент-анализа в зависимости от объекта анализа: контент-анализ вербальных (естественноязычных) текстов, изображений, аудио и аудиовизуальных текстов [Kort-Butler, 2016].

В зависимости от предмета анализа выделяют содержательный (предметный) и структурный контент-анализ [Rich и др., 2018]. Если в рамках содержательного контент-анализа исследователя интересует, *что* сказано, то в рамках структурного — *как* сказано. При структурном контент-анализе могут рассматриваться, в частности, следующие показатели: количество времени, уделенное определенной теме в источнике; объем газетной колонки, отведенный кандидату во время предвыборной кампании; количество и динамика публикации научных статей, посвященных конкретной проблеме и др. При содержательном контент-анализе выполняется подсчет и сравнение частот выбранных элементов анализа: слов и словосочетаний, категорий, тем, сообщений. Структурный и содержательный контент-анализ не исключают друг друга и могут проводиться одновременно [Там же].

И. Л. Дженис по основанию выбора элементов кодирования выделяет прагматический, семантический и формальный (sign-vehicle) типы контент-анализа.

При прагматическом контент-анализе выбор элементы кодирования производится на основе их возможных причин и воздействия (например, подсчитывается количество употреблений выражения, которое с высокой вероятностью вызывает у реципиент определенные эмоции); при семантическом — на основе значений (например, подсчитывается количество употреблений в тексте какого-либо понятия вне зависимости от того, какими лексемами оно выражено); при формальном — на основе психофизических свойств элементов (подсчитывается количество употреблений конкретных элементов: слов, словосочетаний и т. п.) [Janis, 1965].

Д. Л. Алтейде в зависимости от степени учета контекста и личности исследователя выделяет количественный и этнографический контент-анализ, в рамках которых различается подход к сбору, анализу и интерпретации данных [Altheide, 1987]. Этнографический контент-анализ, в отличие от количественного, является рефлексивным и цикличным, используемые в нем концепты не всегда задаются заранее и часто появляются в ходе анализа. Помимо статистического анализа данных, в этнографическом контент-анализе используется текстовый анализ, а результаты анализа представляются не только в виде таблиц, характерных для количественного контент-анализа, но и в виде связного текста [Там же, с. 67].

Наконец, по степени автоматизации выделяют ручной (выполняемый аналитиком), машинный (выполняемый компьютером) и смешанный контент-анализ [Krippendorff, 2018]. Исторически первым является контент-анализ, выполняемый вручную. Машинный контент-анализ появился в конце 1950-х — начале 1960-х гг., когда автоматическая обработка естественного языка начала казаться исследователям реализуемой: в частности, в те годы были проведены первые успешные опыты машинного перевода. В это время была разработана первая версия контент-аналитической системы *General Inquirer* для анализа текстов в различных областях: от политологии и психотерапии до маркетинга и художественной литературы. Однако многие сопоставительные исследования ручного и машинного контент-анализа привели к неоднозначным результатам: они указывали, с одной стороны, на значительное несовершенство машинного контент-анализа по сравнению с ручным,

с другой стороны, на возможность использования компьютера в качестве вспомогательного инструмента при контент-анализе [Там же].

В настоящее время компьютерные технологии позволяют вывести смешанный контент-анализ на новый уровень: например, А. Н. Олейник указывает на возможность масштабирования фрагментарного ручного контент-анализа на большие массивы текстов посредством машинного обучения [Олейник, 2021].

Несмотря на столь широкое разнообразие типов контент-анализа, на практике его дифференциация зачастую производится на основании дихотомии «количественный — качественный». Количественный контент-анализ подразумевает поверхностный подсчет слов и словосочетаний в анализируемом тексте. Качественный контент-анализ основан на категоризации текстовых единиц. Эта дихотомия часто подвергается обоснованной критике, поскольку качественный контент-анализ, помимо собственно качественного этапа (кодирования единиц текста категориями), содержит также количественный этап (анализ частоты встречаемости категорий) [Gauch, 2002; Krippendorff, 2018; Mayring, 2014].

В рамках настоящей работы основное разграничение проводится между классическим количественным контент-анализом, нацеленным на поверхностную обработку текста, и качественным (семантическим в терминологии И. Л. Джениса) контент-анализом, направленным на категоризацию единиц текста и подсчет встречаемости выявленных категорий. В дальнейшем речь в настоящем исследовании будет идти именно о качественном контент-анализе.

1.3 Интеллектуальный контент-анализ

1.3.1 Стандартная процедура качественного контент-анализа

Стандартная процедура качественного контент-анализа состоит из нескольких этапов [Бызов, 2019; Воронов, 2005; Kort-Butler, 2016]. Сначала исследователь ставит задачу контент-анализа и на ее основе выбирает единицы кодирования,

после чего выполняются процедуры кодирования и категоризации, проведение которых возможно в разном порядке в зависимости от подхода исследователя.

Существует три подхода к порядку процедур кодирования и категоризации: дескриптивный, прескриптивный и суммативный. В случае дескриптивного («снизу вверх», bottom-up) кодирования система анализа не задана заранее, вместо этого исследователь определяет коды и категории анализа, релевантные поставленной задаче, в ходе анализа текста [Воронов, 2005]. Напротив, в случае дескриптивного («сверху вниз», top-down) кодирования исследователь заранее определяет набор кодов и категорий, который может незначительно изменяться в процессе анализа текста [Бызов, 2019; Kort-Butler, 2016]. Наконец, суммативный подход предполагает прескриптивное определение нескольких кодов перед началом анализа и дескриптивное добавление кодов в процессе анализа [Hsieh, 2005].

Кодирование может проводиться вручную с привлечением нескольких кодировщиков или автоматически с помощью компьютера; во втором случае необходимо составить словарь для компьютерного кодирования. В случае привлечения кодировщиков проводятся предварительные тесты для оценки надежности кодирования и определения процента сходимости результирующих текстов, размеченных разными кодировщиками. Л. Корт-Батлер отмечает, что при необходимости проводится пересмотр правил и категорий кодирования, после чего выполняется повторное тестирование на образце текста. Пересмотр может выполняться в ходе контент-анализа неоднократно, что говорит об итеративном характере метода [Kort-Butler, 2016, с. 5]. После основного этапа кодирования выполняется оценка итоговой надежности. Наконец, закодированные данные приводятся к табличному виду, выполняется их подсчет, сравнение, анализ и интерпретация.

Описанная выше процедура классического качественного контент-анализа имеет ряд недостатков, вызванных в первую очередь привлечением кодировщиков.

Во-первых, кодирование текста вручную представляет собой длительный и трудоемкий процесс, особенно в случае необходимости обработки значительных объемов текста. Обеспечение согласованности результатов такого кодирования между несколькими кодировщиками требует долгого обучения и тренировки,

при этом отмечается, что кодировщики не всегда могут воспроизвести даже собственное кодирование [Cioffi-Revilla, 2014].

Во-вторых, ручное кодирование утомительно, и даже хорошо обученные кодировщики со временем совершают все большее количество ошибок. Кроме того, под влиянием размеченных текстов интерпретация категорий кодировщиками может измениться, что в конечном счете приведет к ошибочной классификации и низкой надежности кодирования [Kort-Butler, 2016].

В-третьих, количество неструктурированных текстов постоянно растет, из-за чего их ручная разметка становится проблематичной. Иногда это приводит к использованию выборки вместо генеральной совокупности и, как результат, к снижению гранулированности результатов анализа [Бызов, 2019].

Наконец, кодировщик может пропустить в тексте сложные паттерны: сетевые структуры, временные зависимости, скрытые свойства [Cioffi-Revilla, 2014].

Все перечисленное ведет к ухудшению одной из главных характеристик контент-анализа — воспроизводимости, что закономерно снижает ценность результатов такого анализа. Однако эти проблемы могут быть решены с помощью автоматического контент-анализа, в ходе которого исследователь составляет список паттернов для последующего их поиска в массиве текстов. В качестве таких паттернов могут быть использованы любые единицы анализа: слова, словосочетания, концепты, а также сетевые структуры, скрытые свойства и т. п. [Там же].

Результаты автоматического кодирования воспроизводимы как одним и тем же компьютером, так и разными компьютерами, использующими одинаковый набор паттернов. Еще одним преимуществом автоматического кодирования является возможность применения его к массиву текстов большого объема.

Недостаток такого подхода, однако, состоит в том, что набор используемых паттернов ограничен теми, которые были заданы исследователем, и информация, представленная в тексте имплицитно без явных текстовых репрезентаций, будет пропущена при автоматическом контент-анализе. Кроме того, простейшие программы для автоматического кодирования, основанные на словарном методе, не обладают фоновыми знаниями, имеющимися у человека, и могут неправильно

интерпретировать единицы анализа. В работе по контент-анализу текстов для выявления их эмоциональной окраски (сентимент-анализу) Р. Стайна приведен следующий пример некорректной интерпретации слов при автоматическом кодировании текста: отзыв о кофеварке «*Have to push its fancy button over and over to get a pot of coffee. Lots of fun first thing in the morning. Not a great way to start the day*» может быть автоматически помечен как положительный, поскольку в нем есть слова с положительной окраской: *fancy, fun, great*; при этом для человека очевидно, что эти слова использованы саркастически [Stine, 2018, с. 17.2].

Итак, классический качественный контент-анализ, выполняемый вручную, при ориентации на большие объемы текста может привести к проблемам с воспроизводимостью результатов, которые могут быть вызваны недостаточной квалификацией, усталостью кодировщиков, сдвигом интерпретации категорий кодировщиками под воздействием анализируемых текстов и др., в связи с чем имеет смысл автоматизация процедуры кодирования. Однако в случае автоматического кодирования на основе поверхностного словарного метода могут быть потеряны или искажены некоторые смыслы анализируемого текста. Следовательно, требуется такая процедура глубокого качественного контент-анализа, которая может быть автоматизирована. Возможным решением является интеллектуальный контент-анализ.

1.3.2 Определение и процедура интеллектуального контент-анализа

Интеллектуальный контент-анализ (англоязычный эквивалент — *intelligent content analysis, ICA*) представляет собой междисциплинарный метод, сочетающий в себе количественную и качественную обработку текста с использованием интеллектуальных технологий. Интеллектуальный контент-анализ тесно связан с такими областями прикладной лингвистики, как автоматическая обработка текста на естественном языке, извлечение информации, знаний и семантическое аннотирование.

В настоящее время существует немалое количество исследований, выполняемых в русле интеллектуального контент-анализа. Примеры таких исследований описаны в работах [Добров, 2015; Лукашевич и др., 2018; Astrakhantsev, 2014;

Chaves, 2010; Efimenko, 2004; Flaounas и др., 2010; Fong, 2006; Li и др., 2007; Manicheva и др., 2012; Martínez, 2014; Streibel, 2010; Weber, 2013]. Несмотря на широкое распространение практико-ориентированных исследований, теоретические положения данного метода описаны в литературе далеко не в полном объеме.

В англоязычных источниках обнаружено два определения интеллектуального контент-анализа:

- 1) «попытка достичь семантического понимания контекста, в котором встречаются определенные ключевые слова» [Fong, 2006, с. 172];
- 2) «поиск и извлечение информации из баз данных с автоматическим реферированием и классификацией тем» [Multiling. Comp. and Techn., 1999, с. 12].

Полагаем, что оба приведенных определения сужают объем понятия интеллектуального контент-анализа. В первом определении основной единицей анализа указываются ключевые слова, в то время как исследователя могут интересовать не только слова, но и другие единицы, например концепты или невербальные единицы в контент-анализе, нацеленном на обработку изображений и аудиовизуальных текстов. Второе определение предполагает извлечение информации только из баз данных, то есть структурированных источников, несмотря на то что, как правило, для контент-анализа выбирается именно связный неструктурированный текст.

В русскоязычной научной литературе термин «интеллектуальный контент-анализ» встречается крайне редко; в частности, он обнаружен в статьях [Лазарев, 2013; Лидерс, 2015], однако ни в одной из указанных статей не приводится его определение. При этом С. А. Лазарев и соавторы отмечают, что функцией интеллектуального контент-анализа является категоризация содержимого текста и последующее принятие решений на основе имеющихся знаний.

Анализ доступных исследований, выполняемых в русле контент-анализа с применением интеллектуальных технологий, позволяет разделить их на два крупных класса: исследования, ориентированные на обработку текстов любой тематики (в результате чего страдает глубина извлечения информации из узконаправленных текстов), и исследования, посвященные решению точечных задач интеллектуального контент-анализа в конкретной предметной области на конкретном языке.

Построение моделей интеллектуального контент-анализа для конкретных предметных областей на материале конкретного языка является на данный момент наиболее реалистичным и потому наиболее разрабатываемым направлением исследований. Исследований в области разработки многоязычных моделей интеллектуального контент-анализа значительно меньше, и подавляющее большинство из них также ориентировано на конкретный подязык. Большинство открытых зарубежных работ по интеллектуальному контент-анализу нацелены на обработку английского или иного национального (не русского) языка, например, системы [Intelligent Content Analysis, OntoText], работы [Chaves, 2010; Flaounas и др., 2010; Fong, 2006; Li и др., 2007; Martínez, 2014; Streibel, 2010; Weber, 2013] и др.

Среди исследований, выполняемых на русскоязычном материале, можно отметить работы, посвященные системам (как одноязычным, так и многоязычным) ВААЛ [Шалак, 2002], «Гитика» [Кордонский, 2015], АBBYU Compreno [Manicheva и др., 2012], АИРЕ [Добров и др., 2015], NearIdx [Лукашевич и др., 2018], Ontos-Miner [Efimenko, 2004], Texterra [Astrakhantsev, 2014]. Проект NearIdx нацелен на обработку русского языка в широкой предметной области «Национальная безопасность» (включая раздел «Терроризм») в основном в целях информационного поиска, но также рассматривает определенные аспекты формализации естественного языка, релевантные для интеллектуального контент-анализа, в частности рубрикацию и фасетный анализ [Лукашевич и др., 2018].

Изучение указанных выше работ позволяет выделить следующий набор процедур интеллектуального контент-анализа, который не всегда реализуется в полном объеме и приведенном ниже порядке:

1. Сбор текстов предметной области.
 - a. Классификация или кластеризация текстов.
 - b. Выделение особенностей предметной области.
 - c. Извлечение релевантных для анализа частей текстов.
2. Предварительная обработка текстов.
3. Формализация экспертных знаний.
 - a. Определение категорий анализа.

- b. Определение единиц анализа.
 - c. Разработка баз данных (знаний).
 - d. Разметка текстов.
4. Обработка размеченных текстов.
- a. Классификация текстов.
 - b. Извлечение знаний.
 - c. Представление знаний.
5. Интерпретация извлеченных единиц анализа.
6. Формирование отчета интеллектуального контент-анализа.

Экспертные знания могут быть формализованы разными способами от самых простых (облако ключевых слов) до глубоких с упором на семантические, в частности онтологические, базы знаний. Обработка неструктурированной информации с помощью специализированных онтологий является многообещающим вариантом реализации интеллектуального контент-анализа и поэтому все больше привлекает внимание исследователей. Онтологический анализ позволяет выполнять кодирование (разметку) релевантных для контент-анализа элементов текста семантическими метками и таким образом осуществлять формализацию семантических признаков лексических единиц, обеспечивая их измеримость, что в сочетании с традиционными поверхностно-статистическими характеристиками способно значительно повысить качество извлечения информации [Konys, 2015; Iosif, 2012; Sheremetyeva, 2020; Streibel, 2010] и, соответственно, интеллектуального контент-анализа.

1.3.3 Ресурсы интеллектуального контент-анализа

Интеллектуальный контент-анализ как разновидность обработки естественного языка требует задействования различных ресурсов, которые могут быть разделены на статические и динамические [Witt и др., 2009]. Под статическими понимаются неизменяемые во времени ресурсы, содержащие данные в той или иной форме; к ним могут быть отнесены корпуса текстов, словари (лексиконы), онтологии, списки терминов и т. п. Динамические ресурсы, напротив, представляют собой

инструменты, обеспечивающие создание новых данных и последующую их обработку, а именно инструменты разметки текста, морфологические анализаторы, парсеры, инструменты для извлечения лексики и т. п. В ряде случаев провести границу между статическими и динамическими ресурсами может быть затруднительно: например, при оценке статуса онлайн-лексиконов, которые создаются и обновляются автоматически при каждом доступе к ним и, соответственно, не могут считаться ресурсами, не изменяемыми во времени [Там же, с. 2]. Это замечание справедливо и для онтологий, многие из которых постоянно обновляются и пополняются новыми концептами. Следует отметить, что разделение ресурсов на статические и динамические проводится на основе их функции: статические ресурсы предназначены для хранения данных, динамические — для их создания.

Статические и динамические ресурсы интеллектуального контент-анализа могут быть как тексто-ориентированными, так и элементом-ориентированными в зависимости от природы и размера включенных в них лингвистических объектов [Там же]. В рамках этой классификации к тексто-ориентированным статическим ресурсам относятся корпуса текстов, а к элементом-ориентированным статическим ресурсам, состоящим из отдельных, хотя и связанных друг с другом элементов, относятся различные лексиконы и онтологии. В случае динамических ресурсов интеллектуального контент-анализа к тексто-ориентированным ресурсам относятся парсеры и инструменты, основанные на алгоритмах машинного обучения, в то время как элементом-ориентированными динамическими ресурсами могут быть названы инструменты, предназначенные для разметки текста.

Среди ресурсов, используемых в интеллектуальном контент-анализе, можно также выделить необработанные, первичные и обогащенные [Там же]. Необработанным ресурсом является, например, неразмеченный корпус текстов. Первичные ресурсы в свою очередь представляют собой ресурсы, которые были подвергнуты минимальной обработке: так, если при распознавании некачественной копии документа расшифровщик предпочел одно слово другому, он уже осуществил интерпретацию, приняв во внимание контекст. Наконец, обогащенные ресурсы — это

первичные ресурсы, к которым вручную или автоматически с помощью инструментов разметки текста были добавлены некоторые интерпретации.

Под разметкой понимается присвоение интерпретирующей лингвистической или экстралингвистической информации необработанным или первичным ресурсам. Среди типов разметки выделяют, в частности, морфологическую, синтаксическую, семантическую, метаразметку и т. д. [Захаров, 2016]. Изначально интеллектуальный контент-анализ проводится на необработанных или первичных ресурсах, которые в процессе анализа становятся обогащенными.

Отдельно могут быть выделены многоязычные ресурсы, которые содержат несколько разметок, описывающих лингвистические объекты с точки зрения разных языков [Witt и др., 2009]. Многоязычными могут быть корпуса текстов (параллельные и псевдопараллельные), лексиконы и онтологии. Динамические ресурсы (в частности, инструменты разметки текстов) также могут быть ориентированы на многоязычность [Agić, 2017].

Процедура интеллектуального контент-анализа подразумевает использование как статических, так и динамических лингвистических ресурсов. Сам процесс анализа обеспечивается использованием динамических ресурсов, с помощью которых осуществляется кодирование (разметка) текста, извлечение и обработка информации, при этом база знаний, на которой основывается модель интеллектуального контент-анализа, представляет собой совокупность статических ресурсов.

1.4 Онтология как способ формализации знаний для интеллектуального контент-анализа

1.4.1 Определение и структура онтологии

В настоящей работе в качестве способа формализации экспертных знаний для модели интеллектуального контент-анализа выбрана онтология, поскольку онтологии позволяют осуществлять глубокий семантический анализ неструктурированной информации. В современной науке понятие «онтология» является

многозначным, поэтому представляется целесообразным изучить возможные подходы к определению этого понятия, а также подходы к типологизации онтологий.

Понятие онтологии как учения о бытии имеет многовековую историю в философии. Фундаментальные принципы бытия, его структура и закономерности интересовали философов еще в античности, когда самого слова «онтология» не существовало. В научном обиходе термин «онтология» закрепился в первой половине XVIII века в своем философском значении [Nickles и др., 2007].

В начале 1990-х гг. термин «онтология» был заимствован из философии исследователями в области искусственного интеллекта для обозначения артефакта, описывающего значения и взаимосвязи элементов некой системы. Следует отметить, что, в отличие от онтологии-науки, онтологии-артефакты описывают не сами сущности, а человеческие знания об этих сущностях, то есть фактически онтологии-артефакты связываются с эпистемологией [Nirenburg, 2004].

Классическое определение онтологии, используемое в работах по искусственному интеллекту, информационным технологиям, компьютерной лингвистике, было введено в 1993 году американским ученым Т. Грубером, согласно которому онтология — это «явная спецификация концептуализации», где под концептуализацией понимается «абстрактная, упрощенная модель мира, создаваемая нами для определенных целей» [Gruber, 1993, с. 199]. В прикладных лингвистических исследованиях концептуализация обычно интерпретируется как представление знаний через описание множества концептов и отношений между ними, ориентированное на решение конкретной задачи, и формализуется в виде графа, в узлах которого находятся концепты, а дуги представляют отношения между ними.

В прикладной лингвистике понятие онтологии неразрывно связано с естественным языком, поскольку машиночитаемая онтология позволяет обеспечить «понимание» компьютером текста на естественном языке. Однако среди исследователей наблюдаются разногласия относительно необходимой степени зависимости онтологии от естественного языка.

В рамках одной из концепций исследователи отстаивают безусловную независимость онтологий от лексики и грамматики естественного языка, которая

выражается в отсутствии взаимно-однозначных соответствий между концептами онтологии и значениями слов конкретного языка [Nirenburg и др., 1996]. Другая концепция подразумевает, что независимость ресурса от естественного языка не является определяющим фактором для отнесения его к онтологиям [Brewster и др., 2005]. В рамках этого подхода онтологиями считаются также некоторые тезаурусы, описывающие языковые выражения и отношения между ними. Некоторые исследователи называют такие ресурсы легкими онтологиями [Gómez-Pérez, 2010], терминологическими онтологиями [Wielinga, 1994] или ресурсами онтологического типа [Лукашевич, 2011]. Несмотря на отмечаемую зыбкость границ между понятиями онтологии и тезауруса [Guarino, 1997], исследователи считают разграничение этих понятий необходимым из-за их принципиальных различий [Нариньяни, 2001].

Кроме того, существует концепция, согласно которой онтологии и ресурсы онтологического типа могут быть распределены по спектру зависимости от естественного языка на независимые, частично зависимые и полностью зависимые [Onyshkevych, 1997]. Примерами полностью независимых ресурсов являются коммерческая онтология MikroKosmos (μ K), предназначенная для машинного перевода [Nirenburg, 2004], SUMO (Suggested Upper Merged Ontology), созданная в целях обработки информации и размещенная в открытом доступе [Niles, 2003] и BFO (Basic Formal Ontology), разработанная главным образом для информационного поиска [Arp, 2015]. Наиболее зависимым от языка ресурсом из тех, что традиционно относятся к онтологическим, является тезаурус WordNet, описывающий языковые выражения английского языка [Miller и др., 1990]. Промежуточное положение занимают Сус [Foxvog, 2010] и модель «Смысл \leftrightarrow Текст» И. Мельчука, основанная на глубинно-синтаксическом представлении смысла [Мельчук, 1999]. В современной отечественной лингвистике наиболее многообещающее исследование по созданию независимого от языка онтологического ресурса для анализа текста и машинного перевода связано с разработкой системы ЭТАП [Богуславский, 2012].

Мы считаем, что наиболее подходящей для нашего исследования, нацеленного на многоязычность, является концепция независимости онтологий от

конкретного естественного языка, поскольку такой подход позволяет создать онтологию для работы сразу с несколькими языками.

Далее рассмотрим базовую структуру онтологии, которая, как правило, представлена пятью категориями: концепты, атрибуты, отношения, аксиомы, экземпляры. Основу онтологии составляют концепты. В лингвистических исследованиях понятие «концепт» получает совершенно различные интерпретации в зависимости от конкретной области лингвистики. В компьютерной лингвистике концепты онтологии — это «формальные (математические, компьютерные) модели тех единиц, которые [...] принято называть концептами (в ином значении)» [Добров, 2014, с. 60]. Это «иное» значение концепта, о котором пишет в своей диссертации А. В. Добров, приводится в работе И. С. Кубряковой: «...термин, служащий объяснению единиц ментальных или психических ресурсов нашего сознания или той информационной структуры, которая отражает знание и опыт человека; оперативная содержательная единица памяти, ментального лексикона, концептуальной системы и языка мозга (*lingua mentalis*), всей картины мира, отраженная в человеческой психике» [Кубрякова и др., 1996, с. 90]. Иными словами, концепты онтологии — это формализованные обобщения представителей всех сущностей или явлений, организованные в иерархию и связанные друг с другом посредством различных отношений. Концепты могут быть модифицированы различными атрибутами. Условия соотнесения концептов и отношений определяются аксиомами, которые могут быть включены в онтологию для разных целей: проверки корректности информации, вывода новой информации или определения понятийных и числовых ограничений на значения атрибутов. В структуру онтологии также могут входить экземпляры, т. е. отдельные представители концептов, однако в некоторых онтологиях база экземпляров представляет собой отдельный ресурс, связанный с онтологией, но не являющийся ее частью [Nirenburg, 2004].

Поскольку онтологии, как правило, представляют собой развернутые сети с большим количеством концептов, атрибутов и отношений, их визуализация в виде двухмерной модели вызывает определенные трудности. По этой причине для отображения поверхностной структуры зачастую используется иерархия

(дерево концептов), а для глубинной — фреймовое представление знаний, основным инструментом которого является фрейм.

Фрейм содержит условное название концепта и его определение на естественном языке, а также слоты — свойства концепта, которыми описываются его отличительные черты и связи с другими концептами. Слоты заполняются ограничениями ролей, тип накладываемого ограничения определяется фасетом.

Название концепта является условным, поскольку необходимо в первую очередь для легкого восприятия пользователем-человеком и не определяет значение концепта (значение концепта обусловлено его определением и содержимым слотов). Названия концептов в большинстве онтологий указываются на универсальном языке современной науки — английском и состоят из одного-четырех компонентов, разделенных дефисами или написанных слитно в стиле CAMELCASE.

Перечисленные выше элементы являются базовыми, однако в конкретной онтологии некоторые из них могут отсутствовать в зависимости от намерений разработчика и типа онтологического ресурса.

1.4.2 Классификация онтологических ресурсов

Вопрос классификации онтологий и ресурсов онтологического типа в той или иной мере рассматривался многими зарубежными [Gómez-Pérez, 2010; Guarino, 1997; Nois, 2013; Lassila, 2001; Mizoguchi, 1995; Swartout и др., 1997; Uschold, 1996; Wielinga, 1994] и отечественными учеными [Коваль, 2007; Лукашевич, 2011; Митрофанова, 2015; Рубашкин, 2007, 2013]. Онтологические ресурсы могут существенно различаться по ряду параметров, которые целесообразно использовать как основания для классификации. В качестве основных параметров можно выделить степень формальности онтологического ресурса и его наполнение.

Классификация по степени формальности, приведенная в работе [Lassila, 2001], позволяет отличить истинные онтологии от ресурсов онтологического типа. Истинные онтологии характеризуются наличием ряда структур, которые отсутствуют в ресурсах онтологического типа, — формальные таксономии, экземпляры

и разного рода ограничения. При этом такие структуры, как контролируемые словари, глоссарии, тезаурусы и неформальные таксономии могут иметь место в обоих типах ресурсов. Таким образом, определяющим свойством истинных онтологий считается наличие формальных родовидовых отношений между концептами, в которых соблюдается свойство транзитивности.

В работе [Gómez-Pérez, 2010] представлена аналогичная по содержанию классификация, в рамках которой авторы выделяют «легкие» онтологии, включающие в себя только концепты, свойства и таксономии, и «тяжелые», содержащие также ограничения и аксиомы.

Еще один вариант классификации онтологических ресурсов по степени формальности приводится в работе М. Усколда и М. Грюнингера [Uchold, 1996]. Данная классификация, однако, основана не на структурных компонентах таких ресурсов, а на степени формальности языка описания: для описания неформальных онтологий используется естественный язык, для полунеформальных — контролируемый естественный язык, для полуформальных — искусственный язык, и, наконец, для строго формальных онтологий — язык с формализованной семантикой.

Второй параметр классификации, условно названный нами наполнением онтологического ресурса, определяет его содержание, а также количество и качество включаемых в него понятий.

По-видимому, одна из первых классификаций по содержанию была представлена в 1995 году в статье [Mizoguchi, 1995], авторы которой предложили разделение онтологий на три типа:

- общие онтологии, описывающие наиболее абстрактные понятия (например, ВРЕМЯ, ПРОСТРАНСТВО, ОБЪЕКТ) и являющиеся независимыми от конкретной области или задачи;
- онтологии задач, используемые для решения определенной прикладной задачи и содержащие термины, необходимые при разработке программного обеспечения для выполнения этой задачи;
- предметные онтологии, или онтологии предметной области, описывающие знания в рамках ограниченной предметной области.

В работе [Swartout и др., 1997] представлено два типа онтологий: предметные онтологии и онтологические теории, описывающие набор общих концептов для репрезентации некоторого аспекта действительности.

В работе [Van Heist, 1997] онтологии подразделяются на четыре типа: онтологии представления, верхнего уровня, предметной области и прикладные онтологии. В этой классификации впервые выделяются онтологии представления — метаонтологии, описывающие область представления знаний и язык для спецификации онтологий более низких уровней и содержащие такие концепты, как РЕСУРС, КЛАСС, ОГРАНИЧЕНИЕ и др. Оставшиеся три типа повторяют уже упомянутые ранее: онтологии верхнего уровня соответствуют общим онтологиям в классификации Р. Мидзогути и соавт. и онтологическим теориям в классификации Б. Суортаута и соавт. Прикладные онтологии, содержащие концептуальную модель конкретной задачи или приложения, соответствуют онтологиям задач.

Исследователь Й. Хойс, анализируя классификации, представленные в работах [Hartmann и др., 2005] и [Gómez-Pérez, 2010], предлагает развернутую, более детальную классификацию онтологий, в которой представлены девять типов онтологий, начиная с онтологий представления и заканчивая прикладными онтологиями [Hois, 2013]. Интересно, что в данной классификации, в отличие от предыдущей, разграничиваются понятия прикладной онтологии и онтологии задач, а также понятия общей онтологии и онтологии верхнего уровня. Автор подчеркивает, что между девятью представленными типами нет строгих границ, и зачастую для разработки онтологии достаточно базового разделения на три типа: онтологии верхнего уровня, ядерные онтологии и онтологии предметной области. При этом онтологии верхнего уровня могут быть использованы повторно для разработки ядерных онтологий, а те в свою очередь — для разработки онтологий предметной области. Тип онтологии позволяет разработчику выбрать необходимую степень детализации концептов и отношений, определяемых в онтологии, а следовательно, и общую глубину представляемой информации [Там же, с. 23].

Вариант классификации по содержанию представлен в статье О. А. Митрофановой и Н. С. Константиновой [Митрофанова, 2015], где предлагается

разделение онтологий на зоны (верхнюю, среднюю и нижнюю) с учетом качества и количества включаемых в них концептов. По-видимому, онтологии верхней зоны, содержащие 100–500 наиболее универсальных концептов, соответствуют общим онтологиям, средней зоны (500–10 000 концептов, представляющих мир в целом) — онтологиям верхнего уровня, нижней зоны (200–2000 концептов в рамках ограниченной предметной области) — онтологиям предметной области.

Представляется разумным из большого многообразия типов онтологий в зависимости от их содержания выделить два — онтологию верхнего уровня и онтологию предметной области. Онтологии верхнего уровня включают в себя знания, общие для всех предметных областей и допускают многократное использование. Это ресурсы, в которых происходит первоначальное разделение действительности на самые общие категории, как, например, в упомянутых ранее онтологиях MikroKosmos, SUMO и BFO. Такие онтологии могут использоваться в качестве базовых и универсальных ресурсов для присоединения различных предметно-ориентированных онтологий [Рубашкин, 2007, с. 481], которые содержат знания об определенных предметных областях: культуре, медицине, авиации и др. — и могут применяться только в рамках соответствующих предметных областей, возможно, многократно для решения различных задач обработки текстов.

Онтологии предметных областей в свою очередь подразделяются на базовые онтологии, содержащие основные, наиболее широкие концепты предметной области, и более узкие онтологии, ориентированные на ее отдельные аспекты. Содержание онтологии предметной области должно соответствовать существующему разделению в связанной с таким ресурсом онтологии более высокого уровня.

Помимо трех основных параметров классификации существуют и другие. Например, Б. Вилинга и А. Шрайбер предложили разделение онтологий на терминологические онтологии и онтологии, моделирующие знания [Wielinga, 1994]. Эта позиция было подвергнута критике исследователем Н. Гуарино, который указал на нечеткость критерия и в свою очередь предложил разделение на используемые и справочные, или, иными словами, онлайн- и офлайн-онтологии, где первые представляют собой несложные структуры, например лексиконы, пригодные для

хранения онлайн, а последние — развернутые теории, определяющие значения терминов, включенных в лексикон [Guarino, 1997].

Отечественный исследователь С. А. Коваль предложил классификацию онтологий по наличию экземпляров: экземплярные онтологии описывают сферу практического знания, а безэкземплярные представляют теоретические, концептуальные знания. Отсутствие экземпляров непосредственно влияет на вводимые в онтологию отношения, выводя их на метаязыковой уровень [Коваль, 2007].

В. Ш. Рубашкин в свою очередь подразделяет онтологии в зависимости от их коммуникативных возможностей на открытые для взаимодействия с другими онтологиями и закрытые [Рубашкин, 2013].

О. А. Митрофанова и Н. С. Константинова в статье [Митрофанова, 2015] приводят еще одну классификацию, в рамках которой онтологии делятся на глубинные и поверхностные в зависимости от способа разработки и исследования. При создании глубинных онтологий разработчики основываются на формальной логике (предикатов первого порядка, модальной, дескриптивной и т. д.), в то время как при создании поверхностных они отталкиваются от естественного языка и выполняют построение онтологии на основе корпуса текстов.

Наряду с приведенными выше типами онтологических ресурсов в исследованиях, выполняемых в рамках компьютерной лингвистики, также упоминаются лингвистические и многоязычные онтологии как онтологии, непосредственно связанные с основным объектом лингвистики — естественным языком.

В результате анализа работ, посвященных лингвистическим онтологиям, нами выявлены три основных интерпретации этого понятия [Зиновьева, 2018]:

- модели представления лингвистических знаний;
- онтологии для обработки естественного языка;
- лексические онтологии, созданные по wordnet-модели.

Первая интерпретация ортогональна двум другим, поскольку подразумевает онтологические ресурсы, содержащие концепты предметной области «Лингвистика»: например, WORD-PART, WRITTEN-LANGUAGE-EXPRESSION и т. д. [Соколова, 2008; Farrar, 2003]. Подобные ресурсы необходимы для структурирования знаний

о данной предметной области и имеют ограниченную сферу применения. В рамках настоящего исследования интерес представляют вторая и третья интерпретация.

Интерпретация лингвистической онтологии как ресурса, предназначенного для обработки естественного языка, основана на формальной трактовке понятия «онтология», т. е. лингвистическая онтология — это формальная онтология, обладающая независимостью от конкретного естественного языка и создаваемая в целях его обработки. Такая онтология содержит сеть иерархически организованных концептов реальной действительности в ее восприятии человеком, связанных друг с другом различными отношениями, при этом концепты связаны с лексикой на одном или нескольких языках. Примером такой лингвистической онтологии является MikroKosmos (μ K), позиционируемая как независимая от языка система концептов реальной действительности, разработанная с целью извлечения информации из текстов на разных языках и осуществления машинного перевода в рамках системы машинного перевода, основанной на знаниях [Mahesh, 1996].

Онтология μ K оперирует тремя типами концептов: OBJECT, EVENT и PROPERTY, причем в последней ветви задаются все отношения, связывающие концепты онтологии, и атрибуты. Объекты, события и свойства описываются в виде фреймов со слотами, которые определяют, как концепты онтологии соотносятся друг с другом (посредством отношений), или устанавливают числовые или понятийные ограничения (посредством атрибутов). В проекте проводится «четкое разграничение между понятийным и эпизодическим знанием» [Там же, с. 10], в результате чего онтология содержит только понятийные (концептуальные) знания, эпизодические же знания (экземпляры) хранятся в специальной базе экземпляров.

Связанные с концептами и экземплярами лексические единицы содержатся в лексиконах и ономастиконах. Следует отметить, что процессы развития концептуальных и языковых знаний ведутся независимо друг от друга. Размер лексикона в десятки раз превышает количество онтологических концептов, что обусловлено способом разработки онтологии: вместо детализации концептов путем их дробления разработчица модифицирует существующие концепты различными свойствами. Например, значение слова *асфальтировать* косвенно соотносится

с концептом COVER-EVENT, в качестве инструмента которого выступает концепт ASPHALT. При этом другой тип покрытия можно описать, изменив инструмент, и необходимость создания для этого еще одного концепта отсутствует [Там же].

Итак, как онтология для обработки естественного языка μ К обладает такими преимуществами, как объективность представления концептуальных знаний, независимость от системы лексических значений конкретного языка, равноправие концептуальной структуры и лексиконов, простота и относительная компактность, достигаемая наложением ограничений на свойства концепта. Кроме того, для ее функционирования не требуется опоры на онтологию верхнего уровня, поскольку все свойства и отношения, которыми она оперирует, описаны в ней самой.

В других источниках под термином «лингвистическая онтология» понимаются ресурсы, построенные по wordnet-модели (в некоторых работах такие ресурсы называются соответственно «ворднетами» [Лукашевич, 2011; Madsen и др., 2010], также приводится термин «лексическая онтология» [Данченков, 2010]).

В статье О. А. Митрофановой и Н. С. Константиновой отмечается, что лингвистическими онтологиями в этом значении называются ресурсы, содержащие лексикализованные понятия (слова или термины) совместно с их языковыми свойствами; при этом используемые понятия опираются на выражения естественного языка [Митрофанова, 2015]. В ресурсах такого типа установлены специфические отношения между компонентами: центральными являются таксономические отношения, также могут присутствовать меронимия, антонимия, синонимия и др.

В зависимости от наличия тех или иных отношений лингвистическими онтологиями называют и таксономии, и глоссарии, и тезаурусы [Лукашевич, 2011]. Таксономия в таком случае представляется вырожденной онтологией, содержащей только один тип отношений между компонентами; тезаурус, в свою очередь, обладает более развитой системой отношений. В качестве примера таких онтологий приводят WordNet, РуТез, иногда MikroKosmos [Лукашевич, 2011; Митрофанова, 2015]. С последним, впрочем, трудно согласиться: как уже было отмечено ранее, онтология MikroKosmos, будучи связанной с лексическими единицами разных языков, тем не менее описывает концепты, а не языковые выражения.

Отметим, что большинство лингвистических онтологий (во второй и третьей интерпретациях этого понятия) разрабатывается на основе лексики одного языка, при этом имена и описания содержащихся в них концептов также формулируются на одном языке, причем, как правило, в обоих случаях этим языком является английский, современный *lingua franca*. Однако существуют и многоязычные онтологии — этот термин используется в литературе для обозначения двух типов онтологических ресурсов: предназначенных для обработки текстов на нескольких языках (1) и переведенных на различные языки для лучшего понимания онтологических знаний пользователями, являющимися носителями разных языков (2).

Независимые от языка лингвистические онтологии по определению предполагают многоязычность в первом значении при том, что каждое слово лексикона на определенном языке особым образом соотнесено с концептом онтологии. Использование конкретного естественного языка (как правило, английского) для наименования концептов в большинстве подобных онтологий объясняется исключительно удобством разработчиков и пользователей онтологии.

В целях расширения сферы применения независимых от языка онтологий для обработки текстов на нескольких языках разработчики соотносят выражения добавляемого языка с онтологическими концептами, руководствуясь рядом специальных правил. В данном случае задача разработчиков состоит в том, чтобы сделать работу по соотнесению языковых выражений с концептами менее трудоемкой. Эта задача успешно решена в онтологии MikroKosmos, для которой разработан специальный инструмент, поддерживающий возможность описания значений лексических единиц на английском, испанском и китайском языках при помощи онтологических концептов и ряда ограничений [Nirenburg и др., 1996].

В научной среде также существуют попытки обеспечить многоязычность онтологий во втором значении посредством адаптации онтологий для иноязычных пользователей. Как правило, исследователи предлагают выполнить (вручную или с некоторой степенью автоматизации) локализацию названий концептов, не изменяя при этом саму структуру концептуализации. В работе [Montiel-Ponsoda и др., 2008] выдвинуто предложение об использовании в этих целях специальной

лингвистической модели, в [Espinoza, 2008] описан инструмент для полуавтоматического перевода названий онтологических концептов, работа [Chaves, 2010] посвящена описанию методики ручной разметки концептов названиями на различных языках. Важно отметить, что перевод названий концептов возможен только для узкоспециализированных онтологий, где концептуализации на разных языках теоретически должны иметь похожую структуру.

В рамках направления, интерпретирующего онтологии как зависимые от языка ресурсы, вопрос многоязычности онтологий рассматривается иначе. В работе [Alatrish, 2014] предложен инструмент для полуавтоматического создания отдельных онтологий на разных языках. В работе [Embley и др., 2011] описана методика объединения онтологий, изначально основанных на лексическом составе разных языков, посредством соотнесения данных и метаданных. В [Chrizman, 2014] предложена методика разработки онтологии, зависимой от одного языка, на основе онтологии, зависимой от другого языка. Таким образом, для данного направления многоязычность онтологий заключается скорее в наличии параллельных онтологий для разных языков, нежели в многоязычности одной онтологии.

Итак, по результатам рассмотрения существующих подходов к классификации онтологий и с учетом того, что онтология, которая будет разработана в ходе настоящего исследования, предназначена для модели многоязычного интеллектуального контент-анализа текстов ограниченной предметной области, к будущей онтологии могут быть предъявлены следующие требования. Во-первых, данная онтология должна быть лингвистической во втором значении и многоязычной в первом значении; такое требование обусловлено необходимостью обрабатывать неструктурированные тексты на нескольких естественных языках. Во-вторых, это должна быть онтология предметной области «Терроризм», связанная с какой-либо существующей онтологией верхнего уровня. Это позволит не создавать с нуля собственное деление действительности, что является крайне сложной философской задачей, и сократить объем работ исключительно до описания предметной области.

1.4.3 Особенности разработки предметных онтологий

Одна из центральных задач данного исследования предполагает разработку онтологии предметной области «Терроризм», ввиду чего особенно важно изучить особенности построения предметных онтологий в целом и проанализировать существующие онтологии терроризма в частности, после чего выбрать подходящий для настоящего исследования путь решения данной задачи.

Рассмотрим существующие подходы к разработке онтологий. В зависимости от цели создания онтологии выделяют несколько подходов [Mahesh, 1996]:

- подход, основанный на энциклопедических знаниях, направленный на покрытие всех знаний, представленных в какой-либо энциклопедии (пример онтологии, созданной в русле данного подхода, — Сус);
- подход, основанный на анализе предметной области, в рамках которого разработчики выбирают ограниченную предметную область с целью отражения в онтологии как можно большего числа концептов данной ПО;
- подход, основанный на задаче онтологии, направленный на отбор только тех знаний, которые необходимы для выполнения какой-либо задачи, например объединения баз данных;
- подход, основанный на анализе лексики языка с целью наиболее полного представления лексического состава языка;
- формальный подход, в рамках которого разработчики руководствуются исключительно принципами формальной логики;
- ситуативный подход, в рамках которого процесс разработки онтологии изменяется в зависимости от потребностей разработчиков.

В соответствии с задачами нашего исследования наиболее подходящим представляется подход, основанный на анализе предметной области.

Базовый алгоритм разработки онтологии любого типа включает три этапа:

- 1) идентификация множества концептов, которые обозначают объекты или события в описываемом фрагменте реальной действительности;
- 2) идентификация множества отношений между концептами;

3) установление отношений между концептами.

В случае с предметными онтологиями этот алгоритм может быть расширен. Поскольку предметные онтологии, как правило, связываются с онтологиями верхнего уровня, разработчику после создания основного набора концептов необходимо решить, какая онтология верхнего уровня соответствует задачам текущего проекта, и связать верхние концепты предметной онтологии с соответствующими концептами онтологии верхнего уровня. Кроме того, в предметную онтологию могут быть добавлены экземпляры. Эта процедура выполняется после того, как установлены отношения между концептами, либо сразу после идентификации этих отношений, поскольку в некоторых случаях отношения должны устанавливаться между конкретными экземплярами, а не представляемыми ими концептами. Следует также подчеркнуть, что предметные онтологии, разрабатываемые с целью обработки естественного языка, должны быть связаны с ресурсами лексики (лексиконами), ввиду чего одним из этапов разработки такой онтологии будет соотнесение единиц лексикона с концептами и экземплярами. В случае многоязычных онтологий необходимо соотнесение с концептами онтологий лексем на нескольких интересующих разработчика языках. Этот этап создания онтологии ставит перед разработчиком нетривиальную задачу, поскольку прямые и однозначные соответствия между лексемами и концептами онтологии могут отсутствовать, что может быть особенно заметно именно в случае многоязычных онтологий.

Первый этап разработки предметной онтологии предполагает формирование ядра онтологии посредством получения содержательного знания предметной области. Поскольку разработчик онтологии, как правило, не является экспертом в моделируемой предметной области, для него получение этого знания возможно несколькими способами: привлечением к разработке онтологии соответствующего эксперта, извлечением информации из специализированных текстов [Власов, 2010] или с помощью построения микротеорий [Митрофанова, 2015].

Совместная работа эксперта в предметной области и разработчика считается оптимальным вариантом, поскольку эксперт обеспечивает полноту и актуальность знаний предметной области, в то время как разработчик следит за правильностью

формализации этих знаний. Однако привлечение эксперта к работе над онтологией не всегда осуществимо.

В случае невозможности привлечения эксперта разработчик вынужден прибегать к одному из двух последних способов. Эти способы не только различаются по своей сути, но зачастую приводят к разным результатам. В работе [Митрофанова, 2015] приводится следующий пример: концепт ЦВЕТ можно попытаться толковать с помощью обращения к естественному языку, либо с помощью построения микротеорий. В первом случае разработчик непосредственно отбирает все слова, обозначающие цвет (в том числе, возможно, на разных языках), и производит их классификацию по какому-либо основанию. Во втором случае разработчик создает микротеорию, например следующую: с точки зрения физики, цвет задается длиной волны, следовательно, на основе этого параметра можно определить все конкретные значения цвета. Можно заметить, что во втором случае реальная действительность представлена более объективно, однако такая онтология не вполне подходит для обработки естественного языка. С другой стороны, обращение к языковым примерам, особенно к ограниченной порции таковых, может привести к тому, что в онтологию не будет внесено достаточное количество концептов для обозначения цвета. В связи с этим наиболее эффективной является комбинация указанных способов получения знаний о предметной области.

После сбора знаний одним из представленных способов разработчику следует определиться, с каких концептов лучше начать построение онтологического дерева. Выделяют три подхода к построению онтологического дерева: нисходящий (top-down), восходящий (bottom-up) и промежуточный (middle-out). Первый способ предполагает построение онтологии сверху вниз, от наиболее универсальных концептов к более конкретным; второй же, напротив, предполагает выделять универсальные концепты на основе индивидуальных сущностей. При промежуточном подходе сначала определяются ключевые концепты, после чего происходит их обобщение (для получения более абстрактных концептов) и детализация (для получения более конкретных) [Francesconi и др., 2010; Uschold, 1996].

Использование восходящего подхода подразумевает более высокий уровень детализации, при этом увеличивается выполняемый разработчиком объем работы и становится труднее обнаружить общие черты в связанных концептах. В результате повышается риск возникновения противоречий, что ведет к ошибкам и необходимости вносить существенные изменения в онтологию. Нисходящий подход, напротив, позволяет лучше контролировать уровень детализации, однако в этом случае существует риск, что создание некоторых концептов верхнего уровня будет необоснованно, что в итоге отразится на стабильности модели.

При разработке онтологии может быть использован любой из перечисленных способов, однако оптимальным для создания онтологии предметной области можно считать промежуточный, поскольку благодаря выбору ключевых концептов предметной области в качестве основы онтологии представляется возможным более точно определить связи между концептами; кроме того, этот подход с большей вероятностью избавляет разработчика от необходимости перестраивания онтологии вследствие возникновения ошибок и противоречий.

Первый этап разработки онтологии может быть полностью или частично автоматизирован, особенно в том случае, если он основан на обработке текстовых корпусов. В этом случае применяются различные инструменты и методы корпусной лингвистики и статистической обработки текстов. Например, в работе [Moreno, 2011] описана методика, в соответствии с которой сначала производится автоматическое извлечение ключевых слов из текста, после чего определяются основные семантические поля, и на основе этих данных разработчиками уже вручную выделяются ключевые для предметной области концепты.

Второй этап разработки предметной онтологии представляет собой идентификацию существующих в предметной области отношений между концептами. Нами были выделены следующие способы определения отношений:

- 1) интуитивный (на основе собственных знаний разработчика);
- 2) извлечение отношений из текстов на естественном языке;

3) повторное использование отношений из онтологии верхнего уровня, с которой связана разрабатываемая онтология предметной области, в том числе посредством конкретизации имеющихся отношений [Зиновьева, 2018].

Если в качестве материала исследования используются корпуса текстов, то на данном этапе могут возникнуть некоторые сложности, поскольку, в отличие от концептов, не все отношения представлены в текстах на лексическом уровне.

Существует множество работ, посвященных обнаружению и извлечению из текстов отношений предметной области с использованием лингвистических шаблонов, начиная со статьи [Hearst, 1992], где был предложен метод выявления в тексте родовидовых отношений при помощи маркеров *является, представляет собой, есть, это*, а также некоторых знаков препинания. Другие отношения, как отмечает автор, могут быть выявлены в тексте с помощью аналогичных маркеров. Однако такой метод подходит только для обнаружения в тексте уже идентифицированных отношений и не отвечает на вопрос о том, как разработчик может выявить новые отношения. Тем не менее этот метод может быть полезен при дальнейшей разработке онтологии на этапе соотнесения концептов друг с другом.

Продуктивный метод извлечения из корпуса новых отношений посредством изучения связей между словами в насыщенном знаниями контексте предложен в работе [Moreno, 2011]. Метод предполагает построение конкорданса в целях исследования коллокационного диапазона высокочастотного слова, которое является реализацией одного из концептов предметной области, и определении типа связи между двумя словами. Этот метод, в сущности, также является в некоторой степени интуитивным, однако его преимущество перед чисто интуитивным методом состоит в большей наглядности и относительной автоматизации. Кроме того, метод позволяет выявить маркеры и паттерны, характерные для того или иного отношения, что в дальнейшем также может найти применения при соотнесении концептов друг с другом и расширении онтологии.

Третий этап разработки предметной онтологии заключается в установлении идентифицированных отношений между идентифицированными концептами. В первую очередь требуется построить онтологическое дерево, т. е. определить

родовидовые отношения между концептами. Эта задача может быть решена интуитивно на основе знаний разработчика или при помощи лингвистических паттернов, предложенных в работе [Hearst, 1992].

Наряду с обнаружением родовидовых отношений паттерны могут быть использованы для поиска меронимических отношений [Aguado de Cea и др., 2009], причинно-следственных связей и семантических ролей [Khoо, 2002], отношений ассоциации и несимметричных отношений [Власов, 2010]. Метод лингвистических паттернов может применяться в совокупности со статистическим [Bhatt, 2015] для вычисления вероятности того или иного типа отношений между концептами.

Поскольку отношения в онтологиях, как правило, являются бинарными, для их обнаружения в тексте необходимо образовать лингвистические паттерны вида: *[первый объект отношения] маркер отношения [второй объект отношения]*, при этом в качестве маркера отношения могут выступать лексические (слово или часть слова), синтаксические (часть речи или член предложения) и (или) семантические (семантическая категория) ограничения. Отметим, что лингвистические паттерны, включающие синтаксические и семантические ограничения, могут быть использованы только при работе с размеченными корпусами, в то время как применение лексических ограничений не требует специальной разметки.

1.4.4 Особенности разработки онтологий терроризма

В последние двадцать лет востребованной задачей разработчиков предметных онтологий является концептуализация предметной области «Терроризм» с целью анализа, прогнозирования и предотвращения террористической деятельности. Изучение концепта «Терроризм» и построение онтологий терроризма проводят с позиций разведывательной и контртеррористической деятельности [Benahmed, 2017; Chmielewski и др., 2009; Galjano, 2009; Inyaem и др., 2010; Koufakis и др., 2020; Mannes, 2005, 2007; Turner, 2011], лингвистики [Жданова и др., 2012; Мишланова, 2012; Химинец, 2016], поведенческой психологии [Barhamgi и др., 2018;

Wendelberg, 2021]. Опыт создания этих онтологий может и должен быть принят во внимание при разработке онтологии в рамках настоящего исследования.

Большинство существующих онтологий предметной области «Терроризм» находится в закрытом доступе; некоторые из них доступны по запросу (например, онтология терроризма, основанная на онтологии верхнего уровня SUMO) или фрагментарно (онтология Л. Вендельберг), другие недоступны совсем, однако некоторые сведения о них известны из опубликованных работ. В данном пункте будут рассмотрены онтологии терроризма, описанные в литературных источниках.

Отметим, что идеальная лингвистическая онтология, ориентированная на предметную область «Терроризм», содержит все знания о террористической деятельности. Однако терроризм — это сложное, многостороннее явление, охватывающее значительное количество ситуаций и взаимосвязей, что подтверждается отсутствием общепринятой дефиниции для этого понятия [Benahmed, 2017]. Этим обусловлено отсутствие на данный момент единого ресурса, включающего в себя все аспекты предметной области. По этой причине разрабатываемые в настоящее время онтологии терроризма на практике являются либо узконаправленными и ориентированными на отдельные аспекты предметной области (например, признаки радикализации индивида, сеть террористов и связи между ними и пр.), либо базовыми, не отражающими подробно специфику каждого направления террористической деятельности, причем оба типа ресурсов зачастую представляют собой не конечный продукт, а исследовательский проект в процессе развития. В каждом случае разработчики создают собственные системы концептов и отношений, которые отвечают конкретным задачам онтологии, и целевые методики решения проблем, связанных с разработкой подобных ресурсов [Шереметьева, 2017].

С целью иллюстрации этих положений приведем далее сравнительно-сопоставительный анализ онтологий терроризма, разработанных или разрабатываемых десятью различными исследовательскими коллективами (таблица 1.1). Данные исследования выполнены по большей части на основе английского языкового материала, за исключением проекта пермских ученых, в котором также были задействованы материалы на русском и французском языках.

Работы были проанализированы по таким параметрам, как область применения (конечный пользователь) онтологии, наличие или отсутствие связи с онтологией верхнего уровня, способ построения онтологии и основные концепты, наличие экземпляров, способ представления онтологии. Результаты сопоставительного анализа представлены по каждому параметру отдельно.

Таблица 1.1. Онтологии терроризма в хронологическом порядке

№	Название ресурса	Страна	Источник
1	Онтология PiT	США	[Mannes, 2005, 2007]
2	Онтология терроризма	Польша	[Chmielewski и др., 2009]
3	Онтология терроризма	Россия (Санкт-Петербург)	[Galjano, 2009]
4	Нечеткая онтология терроризма	Таиланд	[Inyaem и др., 2010]
5	Онтология AIT	США	[Turner, 2011]
6	Онтология терроризма	Россия (Пермь)	[Жданова и др., 2012; Мишланова, 2012; Химинец, 2016]
7	Онтология терроризма	Алжир	[Benahmed, 2017]
8	Онтология Risk-Track	Франция, Тунис, Испания	[Barhamgi и др., 2018]
9	Проект CONNEXIONs	Греция	[Koufakis и др., 2020]
10	Проект OFEDR	Норвегия	[Wendelberg, 2021]

Область применения (конечный пользователь). Первый параметр является ключевым, поскольку он во многом определяет дальнейшую работу по созданию онтологии. В зависимости от цели создания онтологии и области ее возможного применения разработчик определяет необходимость связи предметной онтологии с онтологией верхнего уровня, и если такая необходимость существует, то какая онтология верхнего уровня является наиболее подходящей в конкретном случае. Кроме того, область применения онтологии определяет набор концептов и свойств, необходимость включения экземпляров и способ представления.

Онтология для портала Profiles in Terror (далее — PiT) разрабатывалась в стратегических целях разведывательных контртеррористических служб США и представления знаний о структуре террористической сети, под которой понимается совокупность индивидов и организаций, вовлеченных в террористическую деятельность. В качестве конечного пользователя онтологии, соответственно, выступают разведывательные контртеррористические службы США [Mannes, 2005].

Онтология польских исследователей направлена на идентификацию террористической угрозы посредством выявления скрытых семантических связей. Конечным пользователем онтологии являются государственные и военные структуры [Chmielewski и др., 2009].

Онтология петербургских исследователей Ф. Гальяно и В. В. Поповича спроектирована для обеспечения системы мониторинга для оперативного предотвращения террористических актов и (или) минимизации их последствий. Конечный пользователь онтологии — разведывательное управление ВМС США [Galjano, 2009].

Нечеткая онтология терроризма нацелена на извлечение из текстов знаний о террористической деятельности и (в будущем) для принятия стратегических решений в отношении террористических актов, совершаемых в тайландской провинции Чумпхон. Конечный пользователь онтологии не указан, однако можно предположить, что им являются разведывательные службы Таиланда [Inyuet и др., 2010].

Онтология АИТ предназначения для прогнозирования террористических актов на основе данных о террористических организациях, их намерениях и вооружении. Конечным пользователем онтологии являются разведывательные контртеррористические службы США. Онтология АИТ разрабатывалась как часть комплексной системы — базовой онтологии терроризма ВООТ (Basic Ontology of Terrorism), с которой могли бы работать специалисты из различных областей [Turner, 2011].

Онтология терроризма за авторством ученых из Пермского государственного университета разрабатывалась в целях изучения понятия «терроризм» в англо-, русско- и франкоязычном дискурсе международных интернет-новостей, а также проведения исследований на пересечении лингвистики, философии и социологии [Жданова и др., 2012; Мишланова, 2012; Химинец, 2016].

Онтология для проекта CONNEXIONs направлена на мониторинг онлайн-источников и данных с устройств, связанных с интернетом вещей, для сбора разведанных данных, анализа доказательств и, следовательно, эффективного и результативного расследования террористической деятельности. Конечный пользователь онтологии — правоохранительные органы [Koufakis и др., 2020].

Онтология терроризма, разрабатываемая алжирскими учеными, призвана автоматизировать характеризацию и классификацию терактов на ранних стадиях для смягчения их последствий [Benahmed, 2017].

Онтология Risk-Track предназначена для анализа сообщений в социальных сетях на предмет радикализации их автора [Barhamgi и др., 2018].

Цель онтологии OFEDR заключается в усовершенствовании методов раннего обнаружения проявлений радикализации и предотвращении радикализации индивидов [Wendelberg, 2021].

Конечный пользователь последних трех проектов не указан; предположительно, им являются контртеррористические службы соответствующих стран.

Связь с онтологией верхнего уровня. Достоверно утверждать о связи предметной онтологии с онтологией верхнего уровня представляется возможным только в случае онтологии АИТ, поддерживающей разделение реальной действительности на категории, принятое в BFO. Согласно BFO, существует три основные категории бытия: материальные объекты (OBJECT), их качества (QUALITY) и процессы (PROCESS) [Arp, 2015]. В онтологии АИТ концепты верхнего уровня BFO детализируются до получения концептов предметной области «Терроризм».

Прочие анализируемые ресурсы не имеют связи с онтологиями верхнего уровня (либо о ее наличии авторами не сообщается) и используют собственную концептуализацию соответствующей области террористической деятельности и понятия «терроризм» как такового.

Способ построения онтологии и основные концепты. Работа пермских лингвистов, онтология Risk-Track, онтология проекта OFEDR основаны на нисходящем подходе; нечеткая онтология терроризма, — на восходящем. Разработчики онтологий PiT и АИТ, а также Х. Бенахмед и Н. Ассули используют промежуточный (нисходяще-восходящий) подход к построению ресурса. Подход к разработке онтологии для проекта CONNEXIONs, а также онтологий польских и петербургских исследователей на основе имеющихся данных определить не удалось.

В качестве основы для разработки онтологии PiT выступают базы данных с информацией о террористических актах и исполнителях. В ходе развития

онтологии извлеченные концепты дополняются новыми, более детализированными. Например, помимо основных концептов EVENT (собственно террористический акт), PERSON (участник террористического акта) и ORGANIZATION (террористическая группировка), в онтологию также добавлены концепты PERSONAL STATUS, CONTACTS, TECHNIQUE (способы секретной коммуникации) и пр. Кроме того, основные концепты объединяются в более абстрактные, например, класс ACTIVITY был создан для объединения взаимосвязанных событий, групп и индивидов. Онтология PiT содержит 70 концептов и 173 свойства [Mannes, 2005, 2007].

Онтология польских исследователей представляет собой синтез знаний, полученных от экспертов предметной области, открытых баз данных и баз знаний (Global Terrorism Database, Profiles in Terror), и содержит взаимосвязанные модели различных связанных с терроризмом подобластей: террористические организации, виды оружия, объекты инфраструктуры и т. д. [Chmielewski и др., 2009].

Онтология петербургских исследователей описывает весьма ограниченную предметную область: в данной работе под терроризмом понимается только международный терроризм, а за источник опасности принимаются террористические группировки, при этом отдельные террористы представляют опасность в той степени, в которой они связаны с террористической группировкой [Galjano, 2009, с. 234]. Ввиду того, что онтология направлена на узкую задачу (мониторинг объектов, которые являются потенциальными целями террористов), она содержит всего восемь концептов, которые при этом обладают большим количеством свойств.

Нечеткая онтология терроризма разрабатывается на базе анализа неструктурированных данных из новостных текстов. Релевантные по мнению разработчиков объекты группируются для получения более широких концептов. На следующем этапе разработчики определяют свойства каждого концепта и проверяют структуру онтологии, удаляя ненужные концепты и свойства. Особенностью данной онтологии является наличие в ней нечетких множеств и отношений [Инуаев и др., 2010].

Разработка онтологии AIT (Adversary—Intent—Target) базируется на анализе моделирующего утверждения о террористической ситуации на естественном языке, которое определяет структуру знаний в онтологии: «*Террористический акт*

происходит, когда злоумышленник, имея намерение и возможности, использует оружие против цели» [Turner, 2011, с. 13]. Все ключевые слова: акт, злоумышленник, намерение и др. — анализируются экспертами под руководством куратора — специалиста в области представления знаний. Например, анализ слова злоумышленник позволяет выявить концепт террористической организации и обратный концепт нетеррористической (легальной) организации. Посредством такого анализа проводится идентификация базовых терминов и концептов предметной области и отношений между ними. В онтологии АИТ содержится 15 отношений и множество концептов, полученных путем конкретизации понятий из моделирующего утверждения. Разработка онтологии осуществляется итеративно, поскольку определение отношений между концептами зачастую ведет к переопределению самих концептов и, в свою очередь, переосмыслению некоторых отношений [Там же].

Верхний и основной концепт онтологии терроризма, разрабатываемой пермскими учеными, — ТЕРРОРИЗМ, подразделяющийся затем на концепты, описывающие СУБЪЕКТ, ОБЪЕКТ, РЕЗУЛЬТАТЫ и ПОСЛЕДСТВИЯ террористической деятельности. В качестве источника для извлечения концептов разработчики используют псевдопараллельный корпус текстов русско- и англоязычных интернет-новостей, а также словарные дефиниции и тексты российских законов [Мишланова, 2012].

Исследователи Х. Бенахмед и Н. Ассули применяют подход к разработке онтологии на основе моделирующего утверждения, предложенный М. Тернером, а также повторно используют концепты из других онтологий терроризма и баз данных, в частности Global Terrorism Database. Основные концепты онтологии: PERPETRATOR, INDIVIDUAL, TERRORISTORGANIZATION, MOTIVATION, EFFECT и TARGET. Данная онтология содержит 62 концепта [Benahmed, 2017].

Процесс разработки онтологии Risk-Track выглядит следующим образом: сначала экспертами проводится концептуализация предметной области, затем повторно используются и расширяются концепты из других онтологий терроризма. Основные концепты онтологии могут быть отнесены к трем категориям:

- 1) концепты, описывающие профиль и сообщения пользователя соцсети;
- 2) индикаторы радикализации;

3) концепты, связанные с радикализацией и терроризмом (концепт TERRORIST ORGANIZATION взят из онтологии AIT и дополнен рядом подконцептов и экземпляров; также введен концепт TERRORIST PERSONALITY для репрезентации известных лиц, связанных с террористическими группами) [Barhamgi и др., 2018].

В качестве источников знаний для разработки онтологии CONNEXIONs авторы использовали данные, собранные из текстовых онлайн-источников (видимая, глубокая и теневая сеть, социальные сети), полицейских отчетов и устройств интернета вещей. Онтология призвана охватить широкий круг связанных с терроризмом аспектов. В ней сочетаются концепты, заимствованные из других семантических моделей (например, концепты USER ACCOUNT, WEB DOMAIN из модели репрезентации пространственно-временного контекста), и концепты, созданные авторами самостоятельно (например, AGENT, SENSING DEVICE) [Koufakis и др., 2020].

В ходе разработки онтологии для проекта OFEDR автор Л. Вендельберг выделяет основные параметры анализа, которые затем использует для создания верхнего уровня концептов (CRIMINAL PROFILES, BEHAVIOR, BIOCHEMISTRY, BRAIN, INDIVIDUAL CHARACTERISTICS, GROUP, ENVIRONMENT), после чего детализирует каждый концепт [Wendelberg, 2021]. Поскольку данная онтология рассматривает террористическую деятельность с позиций поведенческой психологии и биохимии, в ней почти отсутствуют концепты, характерные для других анализируемых онтологий (способы осуществления террористических актов, оружие, организации и т. п.).

Экземпляры. Все представленные онтологии включают экземпляры. Особенно важную роль экземпляры играют в онтологии PiT, поскольку значительная часть онтологии описывает индивидуальные характеристики отдельных террористов и их связи друг с другом и с террористическими организациями [Mannes, 2007]. Роль экземпляров в AIT менее значима, поскольку, по мнению разработчиков, отдельные террористические акты, в особенности совершенные в отдаленном прошлом, не могут помочь в решении актуальных задач контртеррористической деятельности [Turner, 2011, с. 24]. В пермской онтологии терроризма экземпляры имеют иллюстративную функцию. Следует отметить, что не все концепты онтологии должны быть представлены в виде экземпляров: так, в онтологии Risk-Track

экземпляры присутствуют только для некоторых концептов: например, TERRORIST ORGANIZATION, TERRORIST PERSONALITY [Barhamgi и др., 2018].

Способ представления онтологии. Наиболее распространенным способом представления знаний среди проанализированных онтологий является формализация на языке OWL [Galjano, 2009; Inyаem и др., 2010; Koufakis и др., 2020; Mannes, 2005; Turner, 2011; Wendelberg, 2021], также используются языки SWRL [Barhamgi и др., 2018] и RDF [Chmielewski и др., 2009]. Пермская онтология терроризма представлена в виде иерархической структуры фреймов со слотами, вершиной которой является концепт ТЕРРОРИЗМ [Жданова и др., 2012]. Онтология авторства алжирских ученых также представлена в виде графа [Benahmed, 2017].

Проведенный анализ работ, в которых рассматриваются принципы разработки онтологий предметной области «Терроризм», позволяет сделать следующие выводы. Во-первых, цели и задачи разработки онтологий терроризма могут существенно различаться от проекта к проекту, что непосредственным образом влияет на параметры разрабатываемой онтологии. Во-вторых, в ходе разработки онтологии исследователям приходится решать как общие проблемы концептуализации, касающиеся, в частности, ширины и глубины детализации онтологии, так и специфические для данной предметной области задачи структурирования знаний с учетом соображений реалистичности выполнения поставленных задач.

ВЫВОДЫ ПО ГЛАВЕ 1

В первой главе рассмотрены основные вопросы моделирования многоязычного интеллектуального контент-анализа и установлены теоретико-методологические основы исследования, определившие дальнейшую практическую работу.

Целью диссертационного исследования является разработка прикладной компьютерной модели, воспроизводящей исследовательскую процедуру контент-анализа с применением интеллектуальных технологий (интеллектуального контент-анализа), под которым понимается извлечение из неструктурированных, возможно, разноязычных текстов ограниченной предметной области контента,

соответствующего информационному запросу пользователя, числовая обработка такого контента, его интерпретация и представление в удобной для пользователя форме. Под контентом в настоящей работе подразумевается то, что содержится в тексте в явном виде или может быть логически выведено на основе эксплицитной текстовой информации и фоновых знаний исследователя.

В результате анализа современных исследований по теме настоящей работы определены требования к разрабатываемой модели многоязычного интеллектуального контент-анализа новостных сообщений предметной области «Терроризм»:

1. Модель должна быть ориентирована на предметную область, что предполагает анализ лингвистического материала данной предметной области.

2. Алгоритм модели должен включать в себя процедуры интеллектуального контент-анализа в том объеме и той последовательности, которые требуются для полноценного решения задачи контент-анализа.

3. Исследование концептуальной структуры предметной области и формализация экспертных знаний должны проводиться на основе онтологического анализа, что предполагает создание онтологии предметной области.

Выбор онтологии в качестве способа формализации экспертных знаний обусловлен тем, что онтологии позволяют проводить глубокий семантический анализ текста, что в совокупности с традиционными поверхностно-статистическими характеристиками способно обеспечить качественные результаты контент-анализа.

С учетом цели и задач диссертационного исследования требуется построение независимой от конкретного языка онтологии, предназначенной для обработки нескольких естественных языков в ограниченной предметной области «Терроризм» и связанной с какой-либо существующей онтологией верхнего уровня. Проведенный анализ онтологий терроризма, разработанных другими исследователями, позволяет заключить, что содержание такой онтологии напрямую зависит от поставленной задачи и может различаться от проекта к проекту весьма значительно.

Для разработки онтологии предметной области «Терроризм» и модели многоязычного интеллектуального контент-анализа в целом требуется провести корпусный анализ лингвистического материала предметной области.

На основании изложенного предлагается методика построения модели многоязычного интеллектуального контент-анализа в ограниченной предметной области, предусматривающая выполнение трех стадий разработки с применением ручных и автоматизированных методов:

1. Анализ лингвистического материала предметной области, который предусматривает сбор корпусов текстов предметной области на языках исследования; анализ ограничений и закономерностей языка подязыка; формирование концептуальных классов лексики, отражающих концептуальную структуру предметной области вне зависимости от конкретного естественного языка.

2. Создание исходной модели многоязычного интеллектуального контент-анализа, в которое входит разработка исходной онтологии предметной области на основе выделенных концептуальных классов, создание онтолексиконов; создание ономастиконов посредством выделения релевантных для предметной области имен собственных; наполнение базы экземпляров; установление связей онтологии с онтолексиконами, ономастиконами и базой экземпляров; создание правил логического вывода, создание исходного алгоритма онтологического анализа.

3. Совершенствование модели многоязычного интеллектуального контент-анализа, в рамках которого выполняется сбор дополнительных корпусов небольшого объема для проверки покрываемости базой знаний новых текстов предметной области. Проводится онтологический анализ части дополнительных корпусов, включающий формальную разметку текстов тегами концептов и при необходимости выполняемое вручную устранение неоднозначности разметки. Релевантные лексические единицы, оставшиеся неразмеченными, вносятся в базу знаний, при необходимости с внесением изменений в состав и содержание концептов. Части дополнительных корпусов последовательно подвергаются онтологическому анализу, в результате которого вносятся изменения в базу знаний. Этапы обработки дополнительных корпусов представляют собой цикл, который завершается при условии получения удовлетворительного результата (в новых текстах количество не покрываемых базой знаний релевантных лексических единиц минимально). После получения оптимальной базы знаний проводится анализ концептуальной

неоднозначности и способов ее устранения. Разрабатываются способы представления результатов контент-анализа, завершается разработка алгоритма модели.

В последующих главах диссертации автор применяет описанную методику для разработки модели многоязычного интеллектуального контент-анализа предметной области «Терроризм».

ГЛАВА 2.

КОРПУСНЫЙ АНАЛИЗ ПОДЪЯЗЫКА НОВОСТНЫХ СООБЩЕНИЙ ПРЕДМЕТНОЙ ОБЛАСТИ «ТЕРРОРИЗМ» НА МАТЕРИАЛЕ РУССКОГО, АНГЛИЙСКОГО И ФРАНЦУЗСКОГО ЯЗЫКОВ

2.1 Предварительные замечания

В последние несколько десятилетий живой интерес исследователей в области прикладной лингвистики вызывают подъязыки, т. е. «языки» конкретных предметных областей, обладающие ограничениями на уровне лексики, синтаксиса, семантики и пр. Неоднократно подтверждено [Сидорова, 2006; Venamara, 2004; Chandio, 1976], что ориентация модели, системы или программного продукта на ограниченный подъязык позволяет значительно повысить качество автоматического анализа и (или) синтеза текста на естественном языке. Это положение имеет непосредственное отношение к настоящему исследованию, нацеленному на обработку языка ограниченной предметной области «Терроризм».

Представление о подъязыке, существующем в рамках естественного языка, но обладающего при этом характерными ограничениями, было введено американским структуралистом З. Харрисом, который определил подъязык (sublanguage) как некоторое правильное (proper) подмножество предложений, замкнутое относительно некоторых операций, определенных для языка в целом (language as a whole) [Harris, 1968, с. 152]. Позже использованный З. Харрисом в этом определении термин «язык в целом» был подвергнут критике Дж. Лербергером, который предложил использовать альтернативный термин «стандартный язык», поскольку в подъязыке могут иметь место грамматические конструкции, не являющиеся нормативными для «языка в целом», и, следовательно, подъязык является не подмножеством «языка в целом», а вариацией стандартного языка [Lehrberger, 1982]. В качестве характерного примера подъязыка, в рамках которого существуют ненормативные для стандартной грамматики конструкции, можно привести высокорегламентированный подъязык телеграмм [Fitzpatrick, 1986].

В отечественной лингвистике термин «подъязык» был впервые использован Н. Д. Андреевым, согласно которому подъязыком считается «совокупность языковых единиц, выделенных тематически однородной областью социального функционирования языка» [Андреев, 1971, с. 36].

Следует отметить, что термин «подъязык» в отечественной и зарубежной лингвистике главным образом используется при характеристике языка научно-технической сферы вследствие высокого формализма ее лексической системы и синтаксической структуры [Смирнова, 1984]. Закономерно возникает вопрос об оправданности использования термина «подъязык» в отношении текстов новостных сообщений, которые не относятся к высокорегламентированным жанрам, как отмечают исследователи К. Озгювен и Д. Цудзии [Ozguven, 1992]. При этом авторы применяют термин «подъязык» по отношению к новостным текстам с некоторыми оговорками, отмечая, что подъязык новостных сообщений не столь четко определен, как подъязыки технической сферы, поскольку многие новостные сообщения содержат цитаты и прочие малоформализуемые элементы [Там же, с. 107]. В свою очередь Л. В. Кудрина отмечает, что новостные тексты являются самым «строгим» с лингвистической точки зрения жанром журналистики ввиду экономии языковых средств и, как следствие, высокой клишированности лексики» [Кудрина, 2020, с. 6].

Мы считаем, что лингвистический инвентарь новостных сообщений отличается ограничениями двойственной природы, определяемыми соответственно жанром и тематикой сообщений. Эта двойственность четко прослеживается в текстах предметной области «Терроризм», которые обладают системой языковых средств, отличающих их от текстов стандартного языка, и, следовательно, составляющих отдельный подъязык, который характеризуется специфической структурой текста, а также морфосинтаксическими и семантическими особенностями лексики.

2.2 Методика анализа подъязыка

Анализ подъязыка новостных сообщений предметной области «Терроризм» проведен последовательно на материале русского, английского и французского

корпусов текстов по единой трехэтапной методике с элементами автоматической и ручной обработки материала. Анализ ориентирован на выявление тех специфических характеристик подязыка, которые могут быть использованы при моделировании многоязычного интеллектуального контент-анализа. Сначала проведен анализ русского корпуса, так как это наиболее флективный язык из представленных (как следствие, заведомо более трудоемкий для анализа), менее изученный в аспекте проводимого исследования, и, наконец, он является родным языком автора.

В первую очередь корпуса были изучены посредством структурного анализа на уровне структуры новостного сообщения, т. е. общего принципа организации текста (структуры релевантности), и его суперструктуры [Van Dijk, 1988].

Традиционно выделяют два основных принципа организации информации в новостных текстах: перевернутая пирамида и повествование [Добросклонская, 2008; Brooks и др., 2008]. В новостном тексте, построенном по принципу перевернутой пирамиды, информация представлена в порядке уменьшения ее релевантности для реципиента. Это наиболее распространенная стратегия создания новостного текста, превалирующая в прессе с начала XX века [Brooks и др., 2008; Mindich, 1998; Röttker, 2003]. Второй принцип организации новостного текста, повествование, рассматривается как «представление фактов и жизненных наблюдений в напряженной манере, характерной для романа или рассказа» [Wise и др., 2009, с. 534], события часто излагаются в хронологическом порядке [Labov, 1967]. Некоторые авторы также выделяют принцип, совмещающий в себе черты перевернутой пирамиды и повествования — «песочные часы» [Колесниченко, 2008].

Суперструктура в свою очередь представляет собой «абстрактную схему, состоящую из условных категорий, которые задают общую функцию тем текста» [Van Dijk, 1988, с. 114]. Основными категориями суперструктуры, или информационными частями текста, являются «Заголовок», «Вводка», «Основная часть», «Фоновая информация» и «Реакция общественности». Помимо перечисленных, в новостных текстах могут присутствовать другие информационные части, например раздел комментариев, разделы «История», «Прошлые события» и т. п. Некоторые информационные части (в первую очередь «Заголовок») обязательны, другие

факультативны. «Заголовок» и «Вводка» (при наличии) имеют характерное форматирование (например, жирный шрифт), в то время как остальные части, как правило, визуально не выделяются. Важно, что информационные части представлены в тексте не непрерывной последовательностью, а появляются эпизодически в разном порядке, поэтому суперструктуры разных текстов могут не совпадать [Там же].

Второй этап анализа подязыка состоял в исследовании морфосинтаксических характеристик лексики на уровне отдельных частей речи (существительных, прилагательных, глаголов, наречий) и лексических групп (именных, глагольных, предложных). В целях автоматизации получения данных для анализа использовался универсальный автоматический экстрактор лексических групп LanaKey [Sheremetyeva, 2012]. Программа может быть настроена на извлечение грамматически правильных лексических групп длиной от 1 до 4 компонентов на материале различных языков и предметных областей. Полученные группы могут быть автоматически отсортированы по частоте, в алфавитном и обратном алфавитном порядке. В отличие от статистических экстракторов, LanaKey может работать на небольших текстах и позволяет получать на их основе достоверные результаты. В данном исследовании программа была настроена на обработку текстов предметной области «Терроризм» на русском, английском и французском языках.

Исследование морфосинтаксических характеристик лексики подязыка проводилось следующим образом. Сначала с помощью экстрактора LanaKey формировались частотные списки однокомпонентных словоформ, после чего автоматически выполнялась лемматизация и получение списков уникальных лексем с указанием частоты встречаемости, разделенных по частям речи. Затем аналогичные операции выполнялись для получения частотных списков многокомпонентных лексических групп длиной от 2 до 4 компонентов (программное ограничение экстрактора). Более длинные (до 10 компонентов) релевантные для предметной области лексические группы извлекались из корпуса посредством ручного поиска лексических групп с меньшим числом компонентов, являющихся частью более крупных лексических групп. Все полученные частотные списки анализировались на предмет грамматических ограничений и лексико-грамматических корреляций.

Особое внимание при анализе морфосинтаксических характеристик было уделено глаголам, поскольку глагольная лексика играет важную роль в организации структуры предложения и текста в целом. Кроме того, глаголы в подъязыках нередко обнаруживают лексико-грамматические корреляции [Шереметьева, 1994].

На третьем этапе анализа подъязыка из перечня всех лексических единиц были отобраны единицы, отражающие контент предметной области «Терроризм», в целях чего была использована прескриптивно-дескриптивная методика. Сначала интуитивно, на основе имеющихся знаний автора и данных, полученных при анализе существующих онтологий предметной области «Терроризм», были прескриптивно заданы первичные концептуальные классы¹ лексики предметной области. Далее извлеченные из корпусов лексические единицы были разнесены по концептуальным классам с использованием компонентного и контекстного анализа (последний выполнялся с помощью программы [Anthony, 2020]). Затем оставшиеся пустыми первичные концептуальные классы были удалены, а для единиц, которые представляют релевантную для предметной области информацию, но не были отнесены ни в один из классов, были созданы новые классы. Процедура формирования концептуальных классов применялась итеративно к каждому языку. В конце анализа был получен окончательный список концептуальных классов, подходящий для репрезентации концептуальной структуры текстов предметной области на любом из трех языков исследования, независимый от конкретного языка.

Как и семантический класс, концептуальный класс может содержать лексические единицы разных частей речи и пересекаться с другими концептуальными классами, образуя смежные синкретичные пространства. В результате между концептуальными классами и содержащимися в них лексическими единицами обнаруживаются связи вида 1:N («один ко многим») и N:1 («многие к одному»). В работе проводится анализ этих связей для каждого исследуемого языка.

¹ Под концептуальными классами в настоящей работе понимаются семантические классы, содержащие лексические единицы конкретной предметной области. В свою очередь семантический класс (поле) рассматривается как «ономасиологическая и семантическая группировка слов, иерархическая организация, объединенная одним родовым значением» [Жеребило, 2010, с. 334].

2.3 Результаты анализа русского корпуса

Исследование проведено на материале корпуса новостных сообщений предметной области «Терроризм» на русском языке за 2014–2020 гг. общим объемом 201 732 словоупотребления. Корпус собран на основе сайтов русскоязычных информационных агентств и газет: «Интерфакс» (<https://interfax.ru/>), «Лента» (<https://lenta.ru/>), «НТВ» (<http://www.ntv.ru/>), «РИА Новости» (<https://ria.ru/>), «Российская газета» (<https://rg.ru/>), «ТАСС» (<https://tass.ru/>) и др.

С точки зрения структуры релевантности все проанализированные сообщения русского корпуса организованы по принципу перевернутой пирамиды. Суперструктура новостных сообщений содержит такие информационные части, как «Заголовок», «Вводка», «Основная часть», «Фоновая информация» и «Реакция общестественности», причем последние две части приводятся в конце текста в произвольной последовательности и являются факультативными. Пример новостного сообщения предметной области «Терроризм» на русском языке с указанием суперструктуры приведен на рис. 2.1.

<i>В Афганистане шесть силовиков погибли в результате взрыва, сообщили СМИ</i>	}	Заголовок
<i>МОСКВА, 2 сен — РИА Новости. По меньшей мере шесть сотрудников сил безопасности погибли в результате взрыва в городе Кундуз на севере Афганистана, передает телеканал Tolo News со ссылкой на источники.</i>		
<i>Как сообщают источники, шесть сотрудников сил безопасности погибли в результате взрыва, прогремевшего на окраине города Кундуз.</i>	}	Основная часть
<i>Телеканал ITV сообщил со ссылкой на МВД страны, что 15 мирных жителей и четыре силовика получили ранения при взрыве.</i>		
<i>Пока ни одна группировка не взяла на себя ответственность за взрыв.</i>	}	Фоновая информация
<i>В Афганистане идет противостояние правительственных сил и боевиков радикального движения «Талибан», которое ранее овладело значительной территорией в сельских районах страны и развернуло наступление на крупные города. Национальные силы обороны и безопасности Афганистана проводят по всей стране совместные операции по борьбе с терроризмом.</i>		

Рисунок 2.1. Суперструктура типичного русскоязычного новостного сообщения предметной области «Терроризм»²

² В Афганистане шесть силовиков погибли в результате взрыва, сообщили СМИ // РИА НОВОСТИ. — Режим доступа: <https://ria.ru/20190902/1558164415.html>, свободный. — Загл. с экрана (дата обращения: 07.03.2021).

Результаты частотного анализа на уровне частей речи приведены в таблице 2.1. Среди знаменательных частей речи наибольшая доля словоупотреблений в русскоязычном корпусе приходится на имена существительные (из них порядка 20 тысяч словоупотреблений — имена собственные), за которыми следуют глаголы и прилагательные; наречия встречаются в корпусе значительно реже. Высокая частота словоупотреблений прочих частей речи обеспечивается в основном частотой предлогов (например, частота встречаемости предлога *в* составляет 13 112).

Таблица 2.1. Общие результаты частотного анализа русского корпуса новостных сообщений предметной области «Терроризм» на уровне частей речи

Часть речи	Частота словоупотреблений		Частота лексем	
	абсолютная	относительная	абсолютная	относительная
Существительные	93 505	46 %	7 160	64 %
Глаголы	31 140	15 %	1 522	14 %
Прилагательные	16 444	8 %	1 653	15 %
Наречия	4 014	2 %	341	3 %
Прочие	56 640	28 %	448	4 %
Всего	201 743	100 %	11 124	100 %

Результат подсчета уникальных лексем явно показывает доминирующее положение существительных: на них приходится 64 % всех лексем, при этом глаголы, прилагательные и наречия занимают суммарно всего треть лексического состава. Это свидетельствует о том, что основной контент предметной области «Терроризм» в русскоязычном корпусе отражен в именах существительных.

Наиболее частотные лексемы по частям речи приведены в таблице 2.2.

Таблица 2.2. Фрагменты частотных списков знаменательных частей речи в русском корпусе новостных сообщений предметной области «Терроризм»

Существительные	Глаголы	Прилагательные	Наречия
человек (2508)	быть (2497)	террористический (670)	ранее (386)
взрыв (2251)	погибнуть (1604)	исламский (440)	более (361)
теракт (2079)	произойти (1210)	местный (429)	менее (205)
результат (1289)	сообщать (748)	взрывной (417)	рядом (191)
город (1024)	получить (886)	сирийский (330)	утром (132)
полиция (971)	сообщить (653)	меньший (267)	сегодня (106)

(Продолжение таблицы 2.2)

Существительные	Глаголы	Прилагательные	Наречия
боевик (851)	взять (590)	последний (217)	позже (87)
год (832)	пострадать (471)	мирный (204)	сейчас (72)
ранение (803)	прогреметь (460)	другой (189)	недалеко (72)
террорист (747)	стать (457)	российский (175)	накануне (62)

В русском корпусе отсутствуют значимые закономерности с точки зрения морфологии имен существительных, прилагательных и наречий, однако глаголы имеют ограниченную парадигму по сравнению со стандартным русским языком (см. таблицу 1 в Приложении А). Большая часть глаголов в русском корпусе (71,62 %) употреблена в личной форме в изъявительном наклонении, из них больше всего глаголов использовано в прошедшем времени (56,45 % от общего числа глагольных словоупотреблений). Глаголов в форме настоящего времени существенно меньше (13,72 %), глаголы в будущем времени почти отсутствуют (1,45 %). Среди неличных форм наиболее частотной является причастие (20,95 %), которое главным образом используется в виде страдательного причастия прошедшего времени в краткой форме (10,35 %). Деепричастия используются редко (1,18 %), частота деепричастий настоящего и прошедшего времени почти одинакова. В незначительном количестве представлены глаголы в повелительном наклонении (0,09 %).

Отдельные глаголы в русском корпусе обладают еще меньшим количеством возможных словоформ: например, глагол *погибнуть* (1604)³ используется только в формах прошедшего времени (1601) и инфинитива (2), глагол *сообщать* (748) — в формах настоящего (653), прошедшего времени (90) и инфинитива (5). Частота словоформ десяти самых частотных глаголов дана в таблице 2 (Приложение А).

Частотный анализ русскоязычного корпуса на уровне лексических групп свидетельствует о высокой частоте предложных (ПГ) и именных групп (ИГ), в то время как глагольные группы (ГГ) встречаются несколько реже. Фрагмент результирующего частотного списка лексических групп приведен в таблице 2.3.

³ Здесь и далее в круглых скобках справа от лексической единицы приводится абсолютная частота данной лексической единицы в соответствующем корпусе.

Таблица 2.3. Наиболее частотные лексические группы в русском корпусе новостных сообщений предметной области «Терроризм»

Именные группы	Глагольные группы	Предложные группы
РИА Новости (376)	взять на себя ответственность (230)	в результате (512)
взрывное устройство (354)	запретить в РФ (144)	по данным (366)
Исламское государство (340)	запретить в России (47)	в результате взрыва (277)
мирный житель (145)	начинить взрывчаткой (38)	в РФ (179)
число погибших (120)	говориться в сообщении (37)	в России (162)
террористический акт (97)	говориться в заявлении (24)	по словам (153)
предварительные данные (95)	пострадать в результате взрыва (23)	в том числе (147)
мощный взрыв (82)	получить ранения (20)	на месте (144)
террористическая организация (78)	взять ответственность за теракт (17)	в действие (139)

Следует отметить, что списки лексических групп всех типов включают многокомпонентные единицы длиной до 10 компонентов: например, именные группы *152-миллиметровый артиллерийский осколочно-фугасный снаряд (1)*, *иракские и иранские силы безопасности (1)*; глагольные группы *взять на себя ответственность за инцидент (4)*, *заниматься вербовкой людей и промывкой мозгов (1)*, *находиться в критическом или крайне тяжелом состоянии (1)*, *отомстить за порабощение миллионов европейцев мусульманами (1)*; предложные группы *в 90 километрах к северу от иракской столицы (4)*, *в первые дни священного для мусульман месяца поста Рамадан (1)*, *возле офиса правящей партии и администрации провинции (1)*, *на территории комплекса мавзолея суфийского философа и поэта (1)*.

На уровне лексических групп морфологические ограничения также наблюдаются только среди глагольных групп: в частности, многие ГГ употреблены в текстах в единственной безличной форме (*говорится в сообщении, говорится в заявлении*), некоторые изменяются только по родам (*запрещен, -а, -ы в РФ*).

В результате анализа русского корпуса на морфосинтаксическом уровне были выявлены лексико-грамматические корреляции между значением и формой отдельных глаголов. В настоящем времени в основном употреблены глаголы со значением высказывания: *говорится, отмечает, передает, пишет, приводятся, считает, сообщает, указывают* и пр. В прошедшем времени использованы глаголы, указывающие на действие или результат, часто террористической или контртеррористической деятельности: *взорвался, задержали, наехал, напали, обвинили,*

погибли, поджег, произошел, пострадали, прогремел, разыскивались, сработало, убит, умер. Среди глаголов в будущем времени наиболее частотные использованы при обращении автора новостного сообщения к читателю: *напомним* (208), *сообщим* (13); при этом у глаголов с более низкой частотой не обнаружено лексико-грамматических корреляций. Глаголы первого лица используются в прямой речи представителями контртеррористической деятельности (*наблюдаем, поймаем, наказываем, рассматриваем*), террористами (*нападу, подохну*), другими лицами для выражения мнения о произошедшем (*выражаем, думаю, надеюсь, осуждаю, считаем*), а также авторами новостных сообщений (*напомним, подчеркнем*).

Объединенные списки одно- и многокомпонентных лексических единиц, из которых были удалены нерелевантные для предметной области единицы, были использованы для образования концептуальных классов с помощью компонентного и (в ряде случаев) контекстного анализа. Например, один из основных концептуальных классов «Террорист» со значением ‘человек, пользующийся методами террора для достижения своих целей’ был образован из следующих лексических единиц с гиперсемой <террорист>:

- *террорист* <террорист>;
- *боевик* <террорист + участник боевых действий>;
- *террористка* <террорист + женщина>;
- *смертник* <террорист + орудие теракта>;
- *смертница* <террорист + орудие теракта + женщина>;
- *талиб* <террорист + член организации «Талибан»>;
- *игиловец* <террорист + член ИГИЛ> и др.

Контекстный анализ использовался для отнесения в концептуальные классы тех лексических единиц, которые не могли быть разложены на компоненты или разложение которых на компоненты не позволяло отнести их к какому-либо релевантному для предметной области концептуальному классу. Например, на основе лексем *человек, подросток* <человек + подростковый возраст>, *женщина* <человек + женский пол>, *мужчина* <человек + мужской пол> может быть образован концептуальный класс «Человек», однако он является слишком широким и

не представляет существенного интереса в рамках предметной области. При этом контекстное употребление перечисленных лексем позволяет отнести их к более узким классам: «Последствия», «Объект теракта» и «Террорист». На рис. 2.2 показан фрагмент конкорданса для лексемы *мужчина*, где в примерах 1–6 этой лексемой обозначается террорист, а в примерах 7–8 — жертва (класс «Последствия»).

(1)	из-за	нападения	вооруженных	мужчин	на	полицейское	училище
(2)	по	московскому	времени	мужчина	открыл	стрельбу	в
(3)	а	также	второго	мужчину ,	бежавшего	к	кабине
(4)	что	увидел	высокого	мужчину	с	длинными	волосами
(5)	у	следствия	данным,	мужчина	открыл	стрельбу	по
(6)	были	задержаны	двое	мужчин.	Некоторые	СМИ	сообщили
(7)	Женщина	и	двое	мужчин	были	ранены	злоумышленниками
(8)	Нападавший	убил	56-летнего	мужчину	и	серьезно	ранил

Рисунок 2.2. Фрагмент конкорданса для лексемы *мужчина*

Отметим, что отнесение лексической единицы к концептуальным классам во всех случаях выполнялось на основании контекста того предложения, в котором встречалась данная единица, без учета общего контекста новостного сообщения.

В результате компонентного и контекстного анализа лексических единиц было образовано 25 концептуальных классов (см. Приложение Б). В таблице 2.4 представлен фрагмент списка концептуальных классов с примерами лексических единиц на русском языке.

Таблица 2.4. Концептуальные классы с примерами на русском языке (фрагмент)

Концептуальный класс	Лексические примеры
Объект теракта	авиабаза, американский солдат, верующий, выпускник полицейской академии, высотное здание
Последствия	арестовать, без жертв, без сознания, безнаказанный, большое число жертв, в голову, в реанимации, в розыске, в тяжелом состоянии, взорваться, взять в плен
Средство теракта	автомат, автомобиль, автомобильный таран, беспилотный летательный аппарат, бесхозный предмет
Террорист	автор нового террористического эпизода, алжирский террорист, афганский талиб, боевик-экстремист, Brenton Tarrant, вооруженный радикал, даишист

(Продолжение таблицы 2.4)

Концептуальный класс	Лексические примеры
Террористическая организация	Аль-Каида, Аль-Шабаб, Ан-Нусра, Боко Харам, вооруженное формирование Исламского государства, Джабхат Фатх аш-Шам, радикальная группировка, террористическая группировка
Тип теракта	автомобильный взрыв, автомобильный теракт, акт терроризма, атаковать, бойня, взрыв автобуса, варварская атака, взорвать, взрыв автобуса, взрыв велосипеда со взрывчаткой, взять в заложники

Следует отметить, что в процессе разделения лексики по концептуальным классам были обнаружены лексические единицы, которые в общем употреблении не несут в себе значения, связанного с терроризмом, однако в контексте новостного сообщения предметной области «Терроризм» обозначают предметы или явления, имеющие с ним непосредственную связь. Например, существительное *взрыв*, отнесенное в концептуальный класс «Теракт», не связывается напрямую с терроризмом в словарях [Ефремова, 2001; Толковый словарь русского языка], однако в анализируемом подязыке используется именно в значении ‘террористический акт с использованием взрывного устройства’, что, возможно, вызвано отсутствием в русском языке иного краткого обозначения для такого типа теракта. Лексемы *женщина*, *мужчина*, *человек*, *неизвестный*, а также *подросток*, *девочка* и *полицейский* в соответствующем контексте могут обозначать террористов: «*В ФРГ подросток дважды пытался взорвать бомбу на рождественской ярмарке*»⁴, «*Пояс со взрывчаткой был закреплен на теле 8-летней девочки*»⁵, «*Убийца российского посла в Турции был полицейским в Анкаре*»⁶.

⁴ В ФРГ подросток дважды пытался взорвать бомбу на рождественской ярмарке // НТВ. — 2016. — Режим доступа: <https://www.ntv.ru/novosti/1733180/>, свободный. — Загл. с экрана (дата обращения: 09.03.2021).

⁵ В Дамаске 8-летняя девочка-смертница взорвала полицейский участок // НТВ. — 2016. — Режим доступа: <https://www.ntv.ru/novosti/1733259/>, свободный. — Загл. с экрана (дата обращения: 09.03.2021).

⁶ Убийца российского посла оказался бывшим сотрудником полиции Турции // РБК. — 2016. — Режим доступа: <https://www.rbc.ru/politics/19/12/2016/5858228c9a7947bd7b2be2e4>, свободный. — Загл. с экрана (дата обращения: 09.03.2021).

Подобная ситуация характерна не только для именных групп; общеупотребительная глагольная лексика также может иметь террористическое значение: например, глаголы *прогреметь*, *врезаться* в проанализированных текстах предметной области однозначно указывают на террористическую тематику текста.

В ходе распределения лексических единиц по концептуальным классам были обнаружены явления концептуальной однозначности, концептуальной неоднозначности и концептуального синкретизма. Концептуально однозначными являются лексические единицы, которые могут быть отнесены только к одному концептуальному классу (например, лексема *взрывпакет* соотносится с единственным концептуальным классом «Средство теракта»). Концептуально синкретичными в свою очередь являются лексические единицы, одновременно реализующие значения, которые позволяют отнести их в несколько концептуальных классов. Концептуальный синкретизм имеет место в следующих случаях:

1. Каждый компонент многокомпонентной лексической единицы может быть отнесен в отдельный концептуальный класс, при этом лексическая единица не может быть разделена из-за тесной смысловой или структурной связи между компонентами. Например, в трехкомпонентной именной группе *силы безопасности Афганистана* лексема *Афганистан* относится к концептуальному классу «Страна», а *силы безопасности* — к концептуальному классу «Контртерроризм».

2. Лексическая единица содержит несколько релевантных семантических компонентов с возможностью их реализации вне указанного контекста. Примером является лексема *подозреваемый*, содержащая семантические компоненты <действующее лицо + подозрение>, поскольку *подозреваемый* — это по определению тот, кто предполагается виновным в совершении преступления, а значит, одновременно относится к концептуальным классам «Террорист» и «Предположение».

3. В конкретном контексте лексическая единица может быть отнесена к двум концептуальным классам, которые не противоречат друг другу. Например, в предложении «*Правоохранительные органы Колумбии задержали подозреваемого*»

в организации теракта 39-летнего Рикардо Андреса Карвахалья»⁷ глагол *задержали* может быть отнесен одновременно к двум релевантным концептуальным классам: «Контртерроризм» и «Последствия».

В отличие от концептуального синкретизма, концептуальная неоднозначность представляет собой явление, при котором лексические единицы в зависимости от контекста могут быть отнесены к различным противоречащим друг другу концептуальным классам и в каждом конкретном случае реализуют принадлежность только к одному из этих классов, как, например, лексема *солдат*:

«По данным агентства, исполнитель взрыва в Германии был одним из <i>солдат</i> исламистов». ⁸	Террорист
«Сообщается о гибели десяти <i>солдат</i> ». ⁹	Последствия
« <i>Солдаты</i> выстрелили и нейтрализовали нападавшего». ¹⁰	Контртерроризм
«По информации пресс-служб ЦАХАЛа, <i>солдаты</i> заметили террориста, бежавшего в их сторону». ¹¹	Объект теракта

Корпусный анализ подъязыка русскоязычных новостей предметной области «Терроризм» позволил выявить несколько причин концептуальной неоднозначности: частеречная омонимия, лексическая и синтаксическая неоднозначность, собственно концептуальная неоднозначность и экстралингвистический контекст.

⁷ Теракты 2019 года, список и краткое описание // POSREDI.RU. — 2019. — Режим доступа: <https://posredi.ru/terakty-2019-goda.html>, свободный. — Загл. с экрана (дата обращения: 12.02.2021).

⁸ «Исламское государство» взяло ответственность за взрыв в Ансбахе // Новая газета. — 2016. — Режим доступа: <http://novayagazeta.ee/articles/8246/>, свободный. — Загл. с экрана (дата обращения: 12.02.2021).

⁹ Атака на военную базу в Сомали: есть жертвы // Корреспондент. — 2016. — Режим доступа: <https://korrespondent.net/world/3713902-ataka-na-voennuu-bazu-v-somaly-est-zhertvy>, свободный. — Загл. с экрана (дата обращения: 12.02.2021).

¹⁰ Еще одна попытка теракта в Хевроне. Террористка арестована // NEWSru.co.il. — 2019. — Режим доступа: https://www.newsru.co.il/israel/12mar2019/hebron_002.html, свободный. — Загл. с экрана (дата обращения: 12.02.2021).

¹¹ Попытка теракта в Хевроне, террорист ликвидирован // NEWSru.co.il. — 2019. — Режим доступа: https://www.newsru.co.il/israel/12mar2019/tkifa_706.html, свободный. — Загл. с экрана (дата обращения: 12.02.2021).

Частеречная омонимия. В случае частеречной омонимии концептуальная неоднозначность возникает на уровне словоформ, при совпадении по крайней мере одной словоформы у разных лексем. Концептуальной неоднозначности по причине частеречной омонимии в русскоязычном корпусе подвержены только однокомпонентные лексемы. Например, лексема *задержанный* функционирует в корпусе в виде причастия и субстантивированного причастия. В первом случае данная лексема относится к концептуальному классу «Контртерроризм», поскольку представляет собой контртеррористическое действие, во втором случае лексема относится к концептуальному классу «Террорист», так как обозначает лицо. Еще одним примером является словоформа *после*, которая может быть интерпретирована как предлог либо как форма предложного падежа существительного *посол*. В первом случае данная словоформа относится к концептуальному классу «Время теракта», во втором — к концептуальным классам «Объект теракта» и «Источник».

Лексическая неоднозначность — возможность двух и более интерпретаций лексической единицы в конкретном контексте, вызываемая полисемией или омонимией [Зализняк, 2004; Поляков, 2004]. В русском корпусе концептуальной неоднозначности по причине лексической неоднозначности также подвержены только однокомпонентные лексемы. Например, лексема *огонь* встречается в русском корпусе в следующих значениях: 1) ‘стрельба’; 2) ‘процесс горения’. В первом случае данную лексему следует отнести в концептуальный класс «Тип теракта», во втором — в концептуальный класс «Последствия». Еще один пример — лексема *следствие*, также имеющая два значения: 1) ‘результат чего-либо’; 2) ‘расследование обстоятельств преступления’. При употреблении в первом значении лексема репрезентует концептуальный класс «Последствия», в то время как использование ее во втором значении относится к концептуальному классу «Контртерроризм».

Синтаксическая неоднозначность — неоднозначность синтаксической структуры словосочетания, вызванная синтаксическими средствами языка, а именно различным делением высказывания на синтагмы, порядком слов и валентностью [Иорданская, 1967]. В русском корпусе синтаксическая неоднозначность крайне редко является причиной концептуальной неоднозначности;

обнаружены только единичные примеры: именные группы *обстрел террористов* и *обстрел бандформирований*, отнесенные к категориальным классам «Теракт» и «Контртерроризм». В приведенных выше примерах неоднозначность возникает за счет неопределенности типа связи между компонентами именной группы, поскольку между ними возможна как субъектная, так и объектная связь. Субъектная связь в именной группе *обстрел террористов* реализуется в предложении: «*Расположенные рядом с демилитаризованной зоной населенные пункты на севере Сирии практически ежедневно подвергаются минометным обстрелам террористов*»¹², при этом в предложении «*Украинские силовики начали авиационный и артиллерийский обстрел террористов*»¹³ представлена объектная связь. При субъектной связи между компонентами именной группы она должна быть отнесена к концептуальному классу «Теракт», иначе — к классу «Контртерроризм».

Собственно концептуальная неоднозначность, т. е. принадлежность к некоторому множеству концептуальных классов при единстве лексического значения. Лексические единицы, идентичные по форме и словарному значению, могут быть отнесены в разные концептуальные классы в зависимости от контекста. Например, лексема *водитель* может быть отнесена к концептуальным классам «Террорист», «Последствия» и «Объект теракта»:

- | | |
|--|-----------|
| 1. «По данным агентства, взрывное устройство было заложено в автомобиле, <i>водитель</i> которого протаранил здание Министерства труда и социальных вопросов, расположенное в административном центре Могадишо». ¹⁴ | Террорист |
|--|-----------|

¹² СМИ: в сирийской Хаме из-за обстрела террористов погиб мирный житель // РИА «Новости». — 2019. — Режим доступа: <https://ria.ru/20190406/1552443035.html>, свободный. — Загл. с экрана (дата обращения: 12.02.2021).

¹³ Украинские силовики начали авиационный и артиллерийский обстрел террористов. — ИС // НВ. — 2014. — Режим доступа: <https://nv.ua/ukraine/ukrainskie-siloviki-nachali-aviacionnyu-i-artilleriyskiy-obstrel-terroristov-is-1886.html>, свободный. — Загл. с экрана (дата обращения: 12.02.2021).

¹⁴ Боевики взорвали здание Министерства труда и соцвопросов в Могадишо // REGNUM. — 2019. — Режим доступа: <https://regnum.ru/news/2597241.html>, свободный. — Загл. с экрана (дата обращения: 12.02.2021).

2. «*Водитель* и несколько сотрудников академии погибли на месте». ¹⁵ Последствия
3. «По данным полиции, подозреваемый затем приблизился к другому *водителю* в красном автомобиле марки Prius, стреляя в него». ¹⁶ Объект теракта

Экстралингвистический контекст. В ряде случаев концептуальная неоднозначность является следствием различия точек зрения на какой-либо вопрос, например вопрос классификации некой организации как террористической. Так, аббревиатура *КСС* (*курдские силы самообороны*) может быть отнесена к концептуальным классам «Террористическая организация» и «Контртерроризм», поскольку, с одной стороны, КСС выступают против «Исламского государства» и, следовательно, их действия можно считать контртеррористическими, но, с другой стороны, они совершают террористические акты против турецких военных, из-за чего могут быть названы террористической организацией.

Все обнаруженные случаи концептуальной неоднозначности в русском корпусе, вызванной экстралингвистическими факторами, связаны с разделением организаций на террористические и нетеррористические, а также отдельных лиц на террористов и представителей контртеррористической деятельности.

Проблема влияния экстралингвистического контекста на концептуальную классификацию лексических единиц, таким образом, достаточно серьезна, поскольку каждый случай такой концептуальной неоднозначности требует тщательного рассмотрения: при выборе концептуального класса необходимо решить, на каком основании организация определяется как террористическая или нетеррористическая, а лицо — как террорист или нетеррорист. Решение может быть принято

¹⁵ Колумбия скорбит по жертвам теракта в Боготе. Погиб 21 человек // ТАСС. — 2019. — Режим доступа: <https://tass.ru/proisshestviya/6013393>, свободный. — Загл. с экрана (дата обращения: 12.02.21).

¹⁶ Двое человек убиты в результате стрельбы в американском Сиэтле // NEWS.UA. — 2019. — Режим доступа: <https://news.ua/world/dvoe-chelovek-ubity-v-rezultate-strelby-v-amerikanskom-sietle>, свободный. — Загл. с экрана (дата обращения: 12.02.21).

на основании контекста или какого-либо перечня лиц и организаций, признанных террористическими, однако оба варианта влекут за собой некоторые проблемы.

В первом случае точка зрения издания, в котором опубликовано новостное сообщение, может не совпадать с мнением мирового сообщества, и тогда организация, являющаяся террористической, может быть помечена как нетеррористическая.

Во втором случае затруднения вызывает выбор перечня террористов и террористических организаций, на который следует ориентироваться при классификации лексической единицы, учитывая, что в разных странах эти перечни могут быть различны. Например, в России ХАМАС на 2019 год не был признан террористической организацией [Федеральная служба безопасности РФ, 2019] в отличие от стран Евросоюза [European Council, 2019]. Кроме того, такие перечни периодически изменяются; следовательно, собранные таким образом данные устаревают. Поэтому при классификации лексики неоднозначные лексические единицы, обозначающие лица и организации, были отнесены во все подходящие концептуальные классы с допущением, что выбор правильного варианта будет осуществляться на этапе разметки текста в процессе интеллектуального контент-анализа.

Следует отметить, что концептуальная неоднозначность лексической единицы в русском подъязыке может быть обусловлена одновременно несколькими причинами. Так, слово *операция* было отнесено в концептуальные классы «Последствия», «Контртерроризм» и «Прочая террористическая деятельность», при этом оно обнаруживает два источника концептуальной неоднозначности: лексическую неоднозначность и собственно концептуальную неоднозначность.

С точки зрения лексической неоднозначности, в анализируемом подъязыке слово *операция* может обладать одним из двух лексических значений:

- 1) 'ряд военных действий, объединенных одной целью';
- 2) 'хирургическая лечебная помощь'.

В первом значении слово употребляется часто, причем при описании действий как представителей контртеррористической деятельности, так и террористов, то есть в указанном лексическом значении данное слово имеет также два концептуальных значения. Например, на основе значения в предложении «*В ответ*

на активизацию боевиков армия Буркина-Фасо проводит операции по выявлению и ликвидации баз террористов»¹⁷ слово *операция* должно быть отнесено в концептуальный класс «Контртерроризм», в то время как на основе предложения «В начале месяца в МИД РФ предупредили, что террористы готовят военную операцию в сирийском Идлибе»¹⁸ — в класс «Прочая террористическая деятельность».

Во втором лексическом значении слово *операция* употребляется редко и может быть отнесено в класс «Последствия»: «Пострадавший доставлен в больницу „Шаарей-Цедек“ в Иерусалиме. Ему предстоит пройти операцию»¹⁹.

Итак, результаты анализа русскоязычного корпуса свидетельствуют о том, что тексты данного корпуса характеризуются собственным ограниченным подязыком, ограничения которого проявляются на уровне структуры релевантности и суперструктуры текста, грамматики, лексики и семантики. Преобладающей структурой релевантности текста является перевернутая пирамида; текст подразделяется на пять информационных частей, три из которых являются обязательными, две — факультативными. Большая часть лексики предметной области представлена существительными и именными группами. Ограничения на уровне морфологии присутствуют у глаголов и глагольных групп, парадигма которых значительно сокращена по сравнению со стандартным русским языком; преобладают глаголы в прошедшем времени. Многие глаголы и глагольные группы, в том числе высокочастотные, представлены в корпусе в единственной грамматической форме, что свидетельствует о клишированности средств языкового выражения. Кроме того, в некоторых формах русские глаголы демонстрируют корреляцию между грамматической формой и лексическим значением.

¹⁷ FP: в Буркина-Фасо при нападении на церковь погибли 14 человек // ТАСС. — 2019. — Режим доступа: <https://tass.ru/mezhdunarodnaya-panorama/7239641>, свободный. — Загл. с экрана (дата обращения: 12.02.21).

¹⁸ В результате двух взрывов в Идлибе погибло не менее 15 человек // LIFE. — 2019. — Режим доступа: <https://life.ru/p/1194727>, свободный. — Загл. с экрана (дата обращения: 12.02.21).

¹⁹ Теракт в поселении Бейт-Эль: 7-летний мальчик получил огнестрельное ранение // Вести. — 2019. — Режим доступа: <https://www.vesty.co.il/articles/0,7340,L-5484404,00.html>, свободный. — Загл. с экрана (дата обращения: 12.02.21).

Релевантное для предметной области содержание корпуса может быть представлено в виде 25 широких концептуальных классов, которые могут быть использованы в качестве основы онтологии предметной области «Терроризм».

2.4 Результаты анализа английского корпуса

Исследование проведено на материале корпуса новостных сообщений на английском языке за 2014–2020 гг. общим объемом 200 608 словоупотреблений. Корпус собран на основе сайтов англоязычных информационных агентств и газет: The Associated Press (<https://apnews.com>), BBC News (<https://bbc.com/news>), Bloomberg (<https://bloomberg.com>), The Daily Mail (<https://dailymail.co.uk>), The Defense Post (<https://thedefensepost.com>), Reuters (<https://reuters.com>), Sky News (<https://news.sky.com>), Xinhua News Agency (<http://news.cn/english>) и др.

В английском корпусе, как и в русском, приоритетной структурой новостного сообщения является перевернутая пирамида с выделением информационных частей «Заголовок», «Вводка», «Основные события», «Фоновая информация» и «Реакция общественности». Последние две части факультативны и дискретны, могут быть расположены в любой последовательности. Пример англоязычного новостного сообщения с указанной структурой приведен на рис. 2.3.

Некоторые новостные сообщения в английском корпусе обладают смешанной структурой, в которой основной, как правило, довольно небольшой по объему текст построен по принципу перевернутой пирамиды, а дополнительные сведения (реакция общества, информация о взятии ответственности за теракт, обновления о количестве раненых и погибших и пр.) отражаются в тексте по мере поступления к журналистам (рис. 2.4). Особенность новостных сообщений с этой структурой состоит в том, что они не статичны до исчерпания информационного повода и представляют информацию в обратном хронологическом порядке, из-за чего их нельзя отнести к структуре «песочные часы». Поскольку подобная структура встречается в корпусе довольно редко, в настоящем исследовании ею можно пренебречь.

<i>Egypt police officer killed defusing bomb at Coptic church</i>	}	Заголовок
<i>A police officer was killed and two wounded in an explosion outside a Coptic church near Cairo. The attack comes less than two days before Egypt's Christians celebrate Coptic Christmas.</i>		Вводка
<i>An Egyptian police officer has been killed attempting to defuse an improvised explosive device near a Coptic church on Saturday, a security source said.</i>	}	Основная часть
<i>The bomb had been placed inside a bag and put on a nearby roof in the western Cairo district of Ezzbet El-Hagana, Egypt's Al-Ahram newspaper said.</i>		
<i>Two other police officers and a bystander were injured when the device detonated, security sources added.</i>		
<i>There was no immediate claim of responsibility.</i>	}	Фоновая информация
<i>Egyptian security forces have increased their presence around churches and monasteries around New Year's Eve and Coptic Christmas, which is celebrated on January 7, in a bid to prevent jihadist attacks.</i>		
<i>Copts, who make up around 10 percent of the Egyptian population, have often been attacked by Islamist militants, who view them as infidels. The rise in attacks prompted authorities to place armed guards outside churches and monasteries.</i>		

Рисунок 2.3. Типичное англоязычное новостное сообщение о теракте, организованное по принципу перевернутой пирамиды²⁰

<i>5 terrorists, 3 jawans killed in gunbattle at Air Force base in Punjab</i>	}	Заголовок
<i>Fresh firing reported in Pathankot during combing operation, more terrorists suspected to be hiding, says Defence sources.</i>		Вводка
<i>In a pre-dawn attack, a group of heavily-armed terrorists, suspected to belong to Jaish-e-Mohammed outfit, struck at the Air Force base, leading to a fierce gunbattle in which three security personnel were killed along with five attackers.</i>	}	Основная часть
<i>Here are the live updates:</i>		}
<i>8:30 p.m.: Operation completed, five terrorists killed</i>		
<i>7:45 p.m.: Home Minister Rajnath Singh confirms that all five terrorists have been killed in Pathankot.</i>		
<i>7:43 p.m.: Enemies of humanity who can't see India progress attacked Pathankot, says Prime Minister Narendra Modi.</i>		

Рисунок 2.4. Фрагмент англоязычного новостного сообщения с обновлениями²¹

²⁰ Egypt police officer killed defusing bomb at Coptic church // DW. — 2019. — Режим доступа: <https://p.dw.com/p/3B5Go>, свободный. — Загл. с экрана (дата обращения: 07.06.2019)

²¹ 5 terrorists, 3 jawans killed in gunbattle at Air Force base in Punjab // The Hindu. — 2016. — Режим доступа: <https://www.thehindu.com/news/national/5-terrorists-3-jawans-killed-in-gunbattle-at-Air-Force-base-in-Punjab/article13977090.ece>, свободный. — Загл. с экрана (дата обращения: 07.06.2019).

Результаты частотного анализа английского корпуса, представленные в таблице 2.5, показывают, что среди знаменательных частей речи наиболее высокой частотой встречаемости обладают существительные, за которыми следуют глаголы и прилагательные; частота наречий при этом составляет 1 %. При этом прочие части речи занимают существенную долю от всех словоупотреблений в корпусе — 39 %, что обусловлено наличием артиклей в английском языке.

Таблица 2.5. Общие результаты частотного анализа английского корпуса новостных сообщений предметной области «Терроризм» на уровне частей речи

Часть речи	Частота словоупотреблений		Частота лексем	
	абсолютная	относительная	абсолютная	относительная
Существительные	70 340	35 %	6 859	64 %
Глаголы	35 768	18 %	1 399	13 %
Прилагательные	13 611	7 %	1 584	15 %
Наречия	2 968	1 %	311	3 %
Прочие	77 921	39 %	570	5 %
Всего	200 608	100 %	10 723	100 %

Наиболее частотные лексемы по частям речи приведены в таблице 2.6.

Таблица 2.6. Фрагменты частотных списков знаменательных частей речи в английском корпусе новостных сообщений предметной области «Терроризм»

Существительные	Глаголы	Прилагательные	Наречия
attack (2425)	be (3445)	Islamic (435)	around (215)
people (1178)	say (2532)	least (258)	outside (133)
police (1127)	kill (1955)	military (315)	inside (120)
terrorist (991)	have (765)	last (245)	now (104)
group (756)	injure (505)	local (242)	only (97)
security (595)	take (430)	dead (229)	nearby (96)
force (586)	wound (358)	Iraqi (204)	back (94)
city (553)	tell (355)	new (197)	deadly (73)
state (539)	claim (348)	Afghan (142)	reportedly (71)
terror (457)	shoot (336)	official (137)	early (66)

Результаты анализа морфологических характеристик однокомпонентных лексем указывают на отсутствие в англоязычном корпусе значимых закономерностей с точки зрения морфологии существительных, прилагательных и наречий, при этом глаголы обладают ограниченным числом словоформ. В связи с тем, что английский язык является в большей степени аналитическим и образует сложные времена и пассивные формы с помощью вспомогательных глаголов, анализ морфологических характеристик, проведенный исключительно на основе списка однокомпонентных словоформ, был бы неполным. Поэтому из корпуса также были извлечены биграммы и триграммы вида *am / are / is / have been / has been / had been* + причастие прошедшего / настоящего времени и пр.

Результаты анализа морфологической репрезентации глаголов в английском корпусе приведены в таблице 3 (Приложение А). В английском корпусе преобладают формы прошедшего времени с общей частотой 48,34 %, при этом на формы настоящего времени приходится 20,99 %. Глаголы в будущем времени встречаются менее чем в 1 % случаев. Среди неличных форм глагола наиболее частотны простое причастие настоящего времени (11,46 % от общего количества глагольных словоупотреблений), причастие прошедшего времени (9,87 %), а также простой инфинитив (8,23 %). Кроме того, в корпусе обнаружено незначительное количество глаголов в повелительном наклонении (0,09 %).

Отдельные глаголы в английском корпусе имеют еще меньшее количество словоформ: например, глагол *kill* ‘убить’ (1955) встречается в формах простого настоящего (66), совершенного настоящего (130), простого прошедшего (548), прошедшего совершенного времени (24), причастия настоящего времени (184), совершенного причастия настоящего времени (2), причастия прошедшего времени (972), инфинитива (29); глагол *wound* ‘ранить’ (358) — в формах простого настоящего (5), простого прошедшего (160), прошедшего совершенного времени (4), причастия настоящего времени (26), совершенного причастия настоящего времени (2), причастия прошедшего времени (161); глагол *read* ‘читать’ (15) используется только в форме повелительного наклонения. Частота встречаемости словоформ десяти самых частотных глаголов английского корпуса дана в таблице 4 (Приложение А).

Частотный анализ англоязычного корпуса на уровне лексических групп свидетельствует о высокой частоте именных групп, в особенности двухкомпонентных; глагольные и предложные группы имеют более низкую частоту встречаемости. Фрагмент частотного списка лексических групп приведен в таблице 2.7.

Таблица 2.7. Наиболее частотные лексические группы в английском корпусе новостных сообщений предметной области «Терроризм»

Именные группы	Глагольные группы	Предложные группы
Islamic State (331)	claim responsibility (160)	on Friday (122)
terrorist attack (354)	take place (134)	on Saturday (102)
terror attack (270)	say in a statement (97)	in the attack (91)
security forces (207)	open fire (64)	as a result (83)
police officer (139)	shoot dead (50)	on Monday (78)
suicide bomber (124)	condemn the attack (50)	on Tuesday (73)
explosive device (83)	kill in terror attack (23)	on Thursday (63)
Prime Minister (92)	kill in the attack (22)	last year (63)
news agency (75)	wound in the attack (17)	in Afghanistan (63)
car bomb (66)	injure in the attack (17)	at the scene (61)

Следует отметить, что в английском корпусе было выявлено большое количество многокомпонентных лексических групп всех типов: например, именные группы *bus interior strewn with shattered glass* (1), *checkpoint at the entrance to Ramallah* (1), *department of Health Administration of Stockholm* (1); глагольные группы *claim responsibility for the terrorist attack* (1), *cause material damage to the property* (1), *put in action by means of remote control* (2); предложные группы *140 km to the north-west of the Iraqi capital* (1), *from Monday night into the early hours Tuesday* (3), *in predominantly Shiite neighborhoods of the Iraqi capital* (1).

На уровне лексических групп морфологическими ограничениями обладают только глагольные группы: в частности, высокочастотные глагольные группы *said in a statement*, *killed in terror attack*, *killed in the attack*, *wounded in the attack*, *injured in the attack* встречаются в корпусе только в одной (приведенной выше) форме.

Выраженных лексико-грамматических корреляций не было обнаружено.

Полученные в результате частотного анализа релевантные для предметной области одно- и многокомпонентные лексические единицы были подвергнуты компонентному и контекстному анализу, а затем распределены по выделенным при анализе русского корпуса концептуальным классам.

В ходе классификации лексики были отмечены некоторые единицы, не несущие в общем употреблении связанного с терроризмом значения, но в контексте рассматриваемых новостных сообщений о террористических актах обозначающие предметы или явления, непосредственно связанные с терроризмом. Например, англоязычные толковые словари [Cambridge Dict.; Merriam—Webster Dict.] не приводят значений слов *militant* ‘боец’ и *fighter* ‘боец’, которые были бы связаны с терроризмом. Однако дистрибутивный анализ подтверждает гипотезу о том, что данные лексемы встречаются в подязыке в аналогичных контекстах, что и слово *terrorist*, при условии, что в тексте сообщается о боевых действиях террористов. При этом слова *militant* и *fighter*, как правило, не используются для обозначения представителей контртеррористической деятельности. Эти факты свидетельствуют в пользу того, что по крайней мере в рассматриваемом подязыке лексемы *militant* и *fighter* можно трактовать как обозначения террористов. Среди глагольной лексики подобная ситуация имеет место со словом *hit* ‘бить’, который в проанализированном корпусе во всех случаях употребления сигнализирует о теракте.

Пример наполнения концептуальных классов лексическими единицами английского языка приведен в таблице 2.8.

Таблица 2.8. Концептуальные классы с примерами на английском языке (фрагмент)

Концептуальный класс	Лексические примеры
Объект теракта	administration officer, Afghan base, airplane, airport, Al-Noor Mosque, ancient temple, asylum seeker, busy fish market, cadet school, Canadian priest, checkpoint
Последствия	aftermath, alive, among the dead, apprehend, arrest, blow themselves up, broken glass shards, bullet wound, burnt stretcher, get seriously damaged, knife

(Продолжение таблицы 2.8)

Концептуальный класс	Лексические примеры
Средство теракта	82-mm mortar, ammonium nitrate, anti-tank mine, armed drone, bomb-laden vehicle, car ram, cartridge clip, cellphone-detonated IED, detonation device, dynamite, hexogen, homemade radio-controlled bomb
Террорист	al Qaeda terrorist, Albanian extremist, alleged perpetrator, armed gang, arsonist, attacker, Daesh supporter, Dozo hunter, female suicide bomber
Террористическая организация	Ahrar al-Sham, al Nusra, al Qaeda, al-Shabaab, Ansar Allah, Aum Shinrikyo, Basque separatist group ETA, cell of the Islamic group, Hezbollah, Islamic State of Iraq and al-Sham, Khorasan, PKK
Тип теракта	airport shooting, armed attack, arson, artillery strike, attempted hijacking, blast, blow himself up, bomb explosion, carnage, detonate suicide vests, direct hit of a mine, exploded car, fire, gunfight, hostage taking, indiscriminate shooting, lone-wolf attack, truck attack

В английском корпусе также были обнаружены случаи концептуального синкретизма и концептуальной неоднозначности. Причинами последней являются частеречная омонимия, лексическая неоднозначность, принадлежность ко множеству концептуальных классов при единстве лексического значения и экстралингвистический контекст. Синтаксическая неоднозначность как причина концептуальной неоднозначности в английском корпусе не выявлена.

Примеры частеречной омонимии приведены в таблице 2.9.

Таблица 2.9. Частеречная омонимия в английском корпусе

Словоформа	Часть речи	Лексема	Концептуальный класс
fire	существительное	fire ‘пожар’	Последствия
	глагол	fire ‘стрелять’	Тип теракта
knife	существительное	knife ‘нож’	Средство теракта
	глагол	knife ‘ударить ножом’	Тип теракта
may	существительное	May ‘май’	Время теракта
	глагол	may ‘мочь’	Предположение

(Продолжение таблицы 2.9)

Словоформа	Часть речи	Лексема	Концептуальный класс
mine	существительное	mine ‘мина’	Средство теракта
	глагол	mine ‘минировать’	Тип теракта
suspect	существительное	suspect ‘подозреваемый’	Предположение + Террорист
	глагол	suspect ‘подозревать’	Предположение

Примеры лексической неоднозначности приведены в таблице 2.10.

Таблица 2.10. Лексическая неоднозначность в английском корпусе

Лексема	Лексическое значение	Концептуальный класс
body	1) труп	Последствия
	2) группа людей	Контртерроризм
directed	1) направленный на какой-либо объект (о террористических актах)	Направление теракта
	2) действующий под руководством кого-л. (о террористах)	Прочая террористическая деятельность
Khorasan	1) регион на Ближнем Востоке	Место теракта
	2) филиал Исламского государства, расположенный в регионе Хорасан	Террористическая организация
	3) предполагаемая группа членов Аль-Каиды, действующая в Сирии	
located	1) расположенный	Место теракта
	2) обнаруженный (о бомбе)	Контртерроризм
release	1) освободить (о заложниках)	Контртерроризм
	2) обнародовать (о заявлениях)	Заявление
station	1) вокзал	Место теракта
	2) радиостанция	Источник
underground cell	1) небольшая группа, действующая в составе более крупной террористической организации	Террорист
	2) камера в тюрьме, расположенная под землей	Контртерроризм

В качестве примера принадлежности лексической единицы ко множеству концептуальных классов при единстве лексического значения приведем лексему *child* ‘ребенок’ в форме множественного числа *children* в контексте следующих предложений:

- | | |
|---|----------------|
| <p>1. Six people were martyred, 47 got injured. most of whom students, <i>children</i> and women. ²²
 ‘Шестеро убиты, 47 ранены, большинство из которых студенты, дети и женщины’</p> | Последствия |
| <p>2. He has actively recruited Australian men, women and <i>children</i>. ²³
 ‘Он активно вербовал австралийских мужчин, женщин и детей’</p> | Террорист |
| <p>3. A “henna night” was underway in the streets, attended mostly by women and <i>children</i>. ²⁴
 ‘На улицах проходила «ночь хны», в которой участвовали в основном женщины и дети’</p> | Объект теракта |

Концептуальная неоднозначность, вызванная экстралингвистическим контекстом, имеет в английском корпусе те же причины, что и в русском. Например, именная группа *Dogon group* ‘группа догонов’ может быть отнесена к противоречащим друг другу концептуальным классам «Террорист» и «Контртерроризм», поскольку догоны считаются ответственными за несколько терактов против народа фулани в Мали, при этом сами фулани поддерживают исламистов, исходя из чего действия догонов могут быть расценены как антитеррористические.

В некоторых случаях концептуальная неоднозначность лексемы в английском корпусе может быть вызвана одновременно несколькими причинами. Например, словоформа *accused* может быть отнесена к различным концептуальным классам: если *accused* — это субстантивированное причастие, то оно может быть отнесено в концептуальные классы «Террорист» и «Предположение» (данные классы

²² Six martyred, 47 injured in terrorist attacks on neighborhoods in Aleppo // Syria Times. — 2016. — Режим доступа: <http://syriatimes.sy/index.php/don-t-miss/26711-six-martyred-47-injured-in-terrorist-attacks-on-neighborhoods-in-aleppo>, свободный. — Загл. с экрана (дата обращения: 15.02.2021).

²³ Australian IS recruiter Neil Prakash 'still alive' // BBC. — 2016. — Режим доступа: <https://www.bbc.com/news/world-australia-38100785>, свободный. — Загл. с экрана (дата обращения: 15.02.2021).

²⁴ Was wedding blast in Gaziantep revenge for IS defeats in Syria? // Al-Monitor. — 2016. — Режим доступа: <https://www.al-monitor.com/ar/originals/2016/08/turkey-isis-suicide-attack-on-kurdish-wedding.html>, свободный. — Загл. с экрана (дата обращения: 20.08.2016).

не являются взаимоисключающими, следовательно, это проявление концептуального синкретизма); если это глагол в форме простого прошедшего времени, то он может быть в зависимости от контекста отнесен в концептуальные классы «Контр-терроризм», «Предположение» или «Заявление»; если же это причастие прошедшего времени, то оно может быть отнесено в концептуальный класс «Последствия». Следовательно, причинами концептуальной неоднозначности словоформы *accused* являются частеречная омонимия и принадлежность к нескольким концептуальным классам при единстве лексического значения.

Итак, результаты анализа англоязычного корпуса позволяют сделать вывод, что тексты данного корпуса также характеризуются собственным подязыком с ограничениями на уровне структуры релевантности и суперструктуры текста, грамматики, лексики и семантики. В целом представленные результаты аналогичны результатам, полученным для русскоязычного корпуса, с небольшими различиями, вызванными несопадением структуры данных языков.

2.5 Результаты анализа французского корпуса

Исследование проведено на материале корпуса новостных сообщений на французском языке за 2014–2020 гг. общим объемом 201 467 словоупотреблений, собранного на основе сайтов информационных агентств и газет AFP (<https://afp.com>), L'Express (<https://lexpress.fr>), Le Figaro (<https://lefigaro.fr>), Libération (<https://liberation.fr>), Le Monde (<https://lemonde.fr>), RT France (<https://francais.rt.com>), Sputnik France (<https://sputniknews.com>) и др.

В проанализированном французском корпусе все новостные сообщения имеют структуру перевернутой пирамиды. Суперструктура в свою очередь состоит из информационных частей «Заголовок», «Вводка», «Основная часть», «Фоновая информация» и «Реакция общественности». Последние две части факультативны и дискретны, могут быть расположены в любой последовательности. В некоторых текстах присутствуют подзаголовки, необходимые либо для четкого разделения информационных частей, либо для выделения в них подразделов.

Пример новостного сообщения на французском языке приведен на рис. 2.5.

<i>Somalie : explosion d'une voiture piégée</i>	}	Заголовок
<i>Au moins 4 morts et une dizaine de blessés dans un attentat à la voiture piégée, perpétré mercredi dans la capitale somalienne, Mogadiscio, selon des témoins locaux.</i>		Вводка
<i>L'explosion a eu lieu le long de la rue Maka Al-Mukarama malgré la récente augmentation du nombre de check points policiers à Mogadiscio à la suite d'une recrudescence des attentats à la bombe commis par les insurgés islamistes shebab, liés à Al-Qaïda.</i>	}	Основная часть
<i>Le 28 mars, une voiture piégée avait foncé dans un restaurant tuant 15 personnes, quelques jours seulement après une attaque lancée par les shebab contre un complexe administratif à Mogadiscio, qui avait fait 11 morts, dont le vice-ministre du Travail.</i>		}
<i>Cet attentat n'a pas encore été revendiqué mais seraient l'œuvre des terroristes somaliens shebab.</i>		

Рисунок 2.5. Суперструктура типичного франкоязычного сообщения о теракте²⁵

Результаты частотного анализа французского корпуса на уровне частей речи, приведенные в таблице 2.11, показывают, что среди знаменательных частей речи наиболее высокая частота наблюдается у имен существительных, за которыми следуют глаголы, при этом частота прилагательных и наречий довольно низкая (6 % и 1 % соответственно). Частота встречаемости прочих частей речи составляет 42 %.

Таблица 2.11. Общие результаты частотного анализа французского корпуса новостных сообщений предметной области «Терроризм» на уровне частей речи

Часть речи	Частота словоупотреблений		Частота лексем	
	абсолютная	относительная	абсолютная	относительная
Существительные	61 594	31 %	5 280	65 %
Глаголы	38 730	19 %	1 153	14 %
Прилагательные	12 844	6 %	1 031	13 %
Наречия	2 793	1 %	194	2 %
Прочие	85 506	42 %	405	5 %
Всего	201 467	100 %	8 063	100 %

²⁵ Somalie : explosion d'une voiture piégée // Sahel Intelligence. — 2019. — Режим доступа: <https://sahel-intelligence.com/14340-somalie-explosion-dune-voiture-piegee.html>, свободный. — Загл. с экрана (дата обращения: 09.03.2021).

Наиболее частотные лексемы по частям речи приведены в таблице 2.12.

Таблица 2.12. Фрагменты частотных списков знаменательных частей речи во французском корпусе новостных сообщений предметной области «Терроризм»

Существительные	Глаголы	Прилагательные	Наречия
attentat (1531)	être (4663)	dernier (347)	y (294)
attaque (1321)	avoir (1726)	local (309)	également (244)
personne (895)	faire (1682)	civil (269)	alors (189)
explosion (693)	blessier (1531)	militaire (264)	aussi (177)
police (636)	mourir (1375)	américain (240)	toujours (150)
terroriste (611)	tuer (880)	responsable (237)	ainsi (123)
groupe (556)	revendiquer (556)	islamique (216)	encore (116)
forces (495)	déclarer (404)	sud (205)	notamment (77)
AFP (437)	piéger (388)	afghan (184)	déjà (65)
sécurité (432)	indiquer (354)	somalien (170)	pendant (61)

Результаты анализа французского корпуса на уровне частей речи показывают отсутствие значимых закономерностей с точки зрения морфологии существительных, прилагательных и наречий, при этом глагольные парадигмы являются ограниченными. Отметим, что французский, как и английский, является в значительной степени аналитическим языком и имеет сложные времена (например, *passé composé*, *plus-que-parfait*), которые образуются при помощи глаголов *avoir* ‘иметь’ и *être* ‘быть’ в соответствующих формах²⁶, а также формы страдательного залога, образующиеся по следующей модели: *avoir* в нужной форме + *été* + причастие прошедшего времени. Поэтому для полноценного анализа морфологической репрезентации глаголов были извлечены биграмммы вида *avoir / être* (в соответствующей форме) + смысловой глагол в форме причастия прошедшего времени, а также триграммы вида *avoir* (в нужной форме) + *été* + причастие прошедшего времени.

Результаты анализа морфологической репрезентации глаголов во французском корпусе представлены в таблице 5 (Приложение А). В отношении рассматриваемого корпуса текстов можно сделать вывод, что обширная временная система

²⁶ В частности, для образования формы *passé composé* глагол *avoir* (*être*) используется в форме *présent simple*, для образования формы *plus-que-parfait* — в форме *imparfait* и т. д.

французского глагола ограничена в нем двумя формами инфинитива, тремя формами причастия, шестью формами изъявительного наклонения, двумя формами условного наклонения и двумя формами сослагательного наклонения. При этом наиболее высокую частоту имеет форма сложного прошедшего времени *passé composé*, на которую приходится 25,67 % от общего количества глагольных словоупотреблений. Вторая по частоте встречаемости личная форма — форма простого настоящего времени *présent* (20,22 %). Кроме того, высокую частоту встречаемости имеет причастие прошедшего времени (24,31 %). При этом глаголы в будущем времени почти отсутствуют (0,72 %). Бóльшая часть глагольных словоформ (56,84 %) согласуется с местоимениями третьего лица *il(s)*, *elle(s)*. Глаголы первого лица встречаются исключительно в цитатах (1,79 %), примеры глаголов второго лица единичны. Глаголов в повелительном наклонении не обнаружено.

Отдельные французские глаголы также имеют меньшее число словоформ: например, глагол *mourir* ‘умирать’ (1375) встречается только в форме настоящего времени (1), сложного прошедшего (79), плюсквамперфекта (3), условного прошедшего (4), инфинитива (13) и причастия прошедшего времени (1275). Частота встречаемости словоформ десяти самых частотных глаголов французского корпуса приведена в таблице 6 (Приложение А).

Частотный анализ французского корпуса на уровне лексических групп демонстрирует высокую по сравнению с другими типами групп частоту глагольных групп, а также преобладание трех- и четырехкомпонентных групп. Фрагмент частотного списка лексических групп представлен в таблице 2.13.

Во французском корпусе также были обнаружены лексические группы всех типов длиной свыше четырех компонентов: именные группы *agence de presse kurde ANHA* (1), *ancien responsable des services secrets afghans* (2), *groupe jihadiste sunnite Etat Islamique* (1); глагольные группы *adresser un message de soutien* (1), *masquer leur défaite sur le terrain des combats* (1), *plonger dans le Bosphore* (2); предложные группы *à 200 km de la frontière algérienne* (2), *à la fin d'un concert de la chanteuse américaine* (1), *dans la zone de récupération des bagages* (1) и т. д.

В ходе анализа французской лексики была выявлена следующая лексико-грамматическая корреляция на уровне глаголов: инверсивная форма глагола вида *a-t-il (elle, on)* + причастие прошедшего времени во всех случаях используется для введения цитат (*a-t-il ajouté, a-t-on appris, a-t-il dit*).

Таблица 2.13. Наиболее частотные лексические группы во французском корпусе новостных сообщений предметной области «Терроризм»

Именные группы	Глагольные группы	Предложные группы
attentat à la bombe (92)	faire état (152)	sur les réseaux sociaux (45)
chef de la police (52)	avoir lieu (131)	en Nouvelle-Zélande (42)
explosion d'une voiture piégée (45)	ouvrir le feu (58)	dans un état critique (28)
véhicule piégé (44)	coûter la vie (25)	dans l'explosion d'une voiture (25)
porte-parole de la police (43)	être la cible (24)	dans un attentat suicide (25)
Union africaine en Somalie (31)	perdre la vie (20)	en Turquie (22)
opération de guérilla (31)	revendiquer l'attaque (20)	dans un double attentat (21)
Gardiens de la révolution (28)	faire mort (13)	à Paris (20)
forces de sécurité (27)	blessés grave (11)	dans l'explosion d'une bombe (18)

Полученные одно- и многокомпонентные лексические единицы, релевантные для отражения контента предметной области, были подвергнуты компонентному и контекстному анализу, после чего распределены по выделенным ранее концептуальным классам.

В процессе распределения французской лексики по концептуальным классам также были обнаружены релевантные для предметной области единицы, которые в стандартном французском языке не несут террористического значения. В частности, высокочастотные существительные *attaque* ‘атака’ и *attentat* ‘нападение’, под которыми в стандартном языке понимается любое нападение [Dictionnaires français Larousse], в анализируемом корпусе всегда используются в значении ‘терракт’ (вероятно, как сокращения от *attaque terroriste* ‘террористическая атака’ и *attentat terroriste* ‘террористическое нападение’). Кроме того, глагол *frapper* ‘ударять’ однозначно указывает на теракт, а глагол *revendiquer* ‘требовать’ в рассматриваемом корпусе всегда означает ‘взять на себя ответственность (за теракт)’.

Пример наполнения концептуальных классов предметной области «Терроризм» французскими лексическими единицами приведен в таблице 2.14.

Таблица 2.14. Концептуальные классы с примерами на французском языке (фрагмент)

Концептуальный класс	Лексические примеры
Объект теракта	acheteur, aéroport saoudien, avion, base du contingent tchadien, centre d'Istanbul, citoyen, civil innocent, chiite, groupe de personnes, jeune, piéton, représentants de l'occident honni, acheteur, gens ordinaires, groupe de soldats israéliens, ambassadeur de Russie en Turquie
Последствия	activer sa charge explosive, arrestation, assassiné, atteint à l'abdomen, blessé au visage et à l'abdomen, blessure légère, brûlé, cadavre, carrosserie endommagée, causer la mort, colonne de fumée noire, corps martyrisé
Средство теракта	arme à feu, arme de poing, attaque à la voiture bélier, balle, bombe humaine, camion grue, charge explosive, engin explosif de fabrication artisanale, mitraillette
Террорист	agresseur, assaillant terroriste, auteur matériel, bombe humaine, combattant, conducteur, délinquant radicalisé, djihadiste ouzbek, jeune homme, soldat, tireur, type
Террористическая организация	Boko Haram, Etat islamique en Irak et au Levant, ex-branche syrienne d'al-Qaïda, groupe islamiste, mouvements d'extrême gauche ou anarchistes grecs
Тип теракта	abattre, acte de terrorisme, acte terroriste, action kamikaze, attaque à la camionnette, carnage, double attentat-suicide à la bombe, égorger, frapper, fusillade, percuter, premier coup d'éclat, prise d'otages, raid macabre, s'être fait exploser, s'être fait sauter

При распределении французской лексики по концептуальным классам обнаружены концептуальный синкретизм и концептуальная неоднозначность ряда лексических единиц. Причинами концептуальной неоднозначности во французском корпусе являются частеречная омонимия, лексическая неоднозначность, принадлежность ко множеству концептуальных классов при единстве лексического значения и экстралингвистический контекст; синтаксическая неоднозначность как причина концептуальной неоднозначности в корпусе не выявлена.

Примеры частеречной омонимии в корпусе приведены в таблице 2.15.

Таблица 2.15. Частеречная омонимия во французском корпусе

Словоформа	Часть речи	Лексема	Концептуальный класс
armée	существительное	armée ‘армия’	Контртерроризм
	глагол	armer ‘вооружить’	Обладание средством теракта
attaquant	существительное	attaquant ‘нападавший’	Террорист
	глагол	attaquer ‘атаковать’	Тип теракта
juge	глагол	juger ‘осуждать’	Заявление
	существительное	juge ‘судья’	Контртерроризм
marché	существительное	marché ‘рынок’	Место теракта + Объект теракта
	глагол	marcher ‘шествовать’	Реакция общественности

Примеры лексической неоднозначности приведены в таблице 2.16.

Таблица 2.16. Лексическая неоднозначность во французском корпусе

Лексема	Лексическое значение	Концептуальный класс
chargé	1) начиненный (об автомобиле, начиненном взрывчаткой)	Обладание средством теракта
	2) отвечающий (указание на область ответственности человека или организации)	Взятие ответственности
site	1) место	Место теракта
	2) веб-сайт	Источник

В качестве примера собственно концептуальной неоднозначности приведем лексему *personne* ‘человек’:

1. Au moins 11 policiers ont été tués et 78 *personnes* blessées, vendredi.²⁷

Последствия

‘По меньшей мере 11 полицейских были убиты и 78 человек получили ранения в пятницу’

2. Cette absence de revendication a de quoi surprendre de la part de Daesh, toujours prompt à prendre la responsabilité d'attentats commis en son nom, y compris par des *personnes* qui n'ont parfois jamais mis les pieds en Syrie ou en Irak.²⁸

Террорист

²⁷ Un poste de police attaqué à la voiture piégée dans l'est de la Turquie // France 24. — 2016. — Режим доступа: <https://www.france24.com/fr/20160826-turquie-cizre-attaque-voiture-piegee-attentat-policiers-pkk>, свободный. — Загл. с экрана (дата обращения: 17.06.2019).

²⁸ Attentat de Gaziantep : la Turquie, une cible à part pour Daesh? // CNEWS. — 2016. — Режим доступа: <https://www.cnews.fr/monde/2016-08-22/attentat-de-gaziantep-la-turquie-une-cible-part-pour-daesh-736845>, свободный. — Загл. с экрана (дата обращения: 17.06.2019).

‘Такое отсутствие заявлений удивительно для ДАИШ, которая всегда быстро берет на себя ответственность за нападения, совершенные от ее имени, в том числе людьми, которые подчас никогда не ступали на территорию Сирии или Ирака’

3. Un assaillant armé d'un fusil d'assaut surgit devant la boîte de nuit Reina, au coeur d'Istanbul, et ouvre le feu sur les *personnes* qui se trouvent à l'entrée.²⁹

Объект теракта

‘Нападавший, вооруженный автоматом, появляется перед ночным клубом Reina в центре Стамбула и открывает огонь по людям у входа’

4. 350 *personnes* restent mobilisées sur le terrain.³⁰

Контртерроризм

‘350 человек по-прежнему задействованы на месте событий’

Примером концептуальной неоднозначности, вызванной экстралингвистическими факторами, является обнаруженная во французском корпусе лексема *Hezbollah*, которая может быть отнесена в классы «Террористическая организация» и «Контртерроризм», поскольку, несмотря на то что организация, действующая под этим именем, во многих странах считается террористической, в незначительном числе контекстов в исследуемом корпусе эта лексема приводится в ряду лексем, которые репрезентуют представителей контртеррористической деятельности.

Во французском корпусе также отмечены случаи, когда концептуальная неоднозначность лексической единицы вызвана несколькими причинами. Например, неоднозначность словоформы *marché* обусловлена частеречной омонимией: это может быть причастие от глагола *marcher* ‘шествовать’, которое используется для указания на реакцию общественности, или существительное со значением ‘рынок’, которое может представлять как место теракта, так и его конкретный объект.

²⁹ Attentat à Istanbul : le point sur la situation // TV5Monde. — 2017. — Режим доступа: <https://information.tv5monde.com/info/attentat-istanbul-le-point-sur-la-situation-146649>, свободный. — Загл. с экрана (дата обращения: 17.06.2019).

³⁰ Fusillade à Strasbourg : trois morts, la France en "urgence attentat" // France Culture. — 2018. — Режим доступа: <https://www.franceculture.fr/emissions/journal-de-6h30/journal-de-6h30-du-mercredi-12-decembre-2018>, свободный. — Загл. с экрана (дата обращения: 17.06.2019).

Итак, результаты анализа франкоязычного корпуса в целом аналогичны результатам, полученным для русско- и англоязычного корпусов, поскольку франкоязычный корпус также характеризуется собственным подъязыком с ограничениями на уровне структуры релевантности и суперструктуры новостного сообщения, его грамматики, лексики и семантики. Незначительные различия, наблюдаемые по результатам анализа корпусов, объясняются несовпадениями в структуре языков.

ВЫВОДЫ ПО ГЛАВЕ 2

Описанный в настоящей главе анализ подъязыка новостных сообщений предметной области «Терроризм» на материале русского, английского и французского корпусов текстов подтверждает двойственную природу ограничений подъязыка.

На уровне структуры текста проявляются ограничения подъязыка, обусловленные жанром новостных сообщений. Рассмотренные новостные сообщения в подавляющем большинстве случаев вне зависимости от конкретного языка имеют структуру релевантности типа «перевернутая пирамида», в которой наиболее релевантная для реципиента информация приводится в заголовке, и релевантность дальнейшей информации последовательно уменьшается. В английском корпусе, в отличие от русского и французского, обнаружены единичные новостные сообщения, обладающие другой структурой релевантности, что может быть обусловлено культурными различиями носителей этих языков и, на наш взгляд, требует более подробного исследования в русле сравнительно-сопоставительного языкознания.

Новостные сообщения также обладают характерной суперструктурой, определяющей присутствие в тексте таких информационных частей, как «Заголовок», «Вводка», «Основная часть», «Фоновая информация» и «Реакция общественности», причем последние две информационные части факультативны и дискретны, т. е. могут отсутствовать в новостном сообщении, а в случае их наличия могут быть расположены в разной последовательности и в том числе могут быть включены одна в другую. В большинстве новостных сообщений информационные части

(кроме заголовка и в некоторых случаях вводки) не имеют формального выделения и могут быть обнаружены только посредством содержательного анализа.

Кроме того, к ограничениям, обусловленным жанром, можно отнести ограничения глагольной лексики на морфологическом уровне. Большая часть глаголов во всех трех языках представлена в формах прошедшего времени, что можно объяснить направленностью новостных сообщений на описание прошедших событий.

Ограничения и закономерности, заданные предметной областью, проявляются на лексико-семантическом уровне. Большая часть контента в рассмотренных языках отражена в существительных, причем конкретные пропорции частотности уникальных лексем почти совпадают, в то время как пропорции частотности словоупотреблений различаются (рис. 2.6). Это, по всей вероятности, обусловлено различиями в грамматике русского, английского и французского языков. Как видно из приведенных на рис. 2.6 диаграмм, показывающих относительную частоту словоупотреблений разных частей речи в каждом из корпусов, основное различие состоит в доле прочих частей речи: числительных, предлогов, союзов, артиклей и т. п. В английском и французском языках, в отличие от русского, есть артикли; при этом в английском и французском языках отличаются продуктивные модели образования словосочетаний (ср. англ. *bomb attack* и фр. *attaque à la bombe*), что также увеличивает долю показателя «Прочие части речи» для французского языка.

Помимо существительных и именных групп, контент предметной области отражается в глаголах и глагольных группах, предложных группах, прилагательных и — в меньшей степени — в наречиях. Все контентно-релевантные лексические единицы можно отнести к одному из 25 базовых концептуальных классов. Набор концептуальных классов для корпусов текстов на каждом из языков совпадает, однако их лексическое наполнение может различаться с точки зрения частей речи. Например, в английском корпусе видна тенденция для обозначения некоторых типов терактов использовать именную лексику (*truck attack, lorry attack*), тогда как в русском и французском корпусах для описания таких типов чаще используются глагольные группы: *врезаться во что-л./кого-л., въехать во что-л./кого-л.*,

протаранить что-л./кого-л. и *foncer sur qqch/qq.* *foncer dans qqch/qqn.* Тем не менее такие различия, вероятно, не вызовут проблем при разработке онтологии.



Рисунок 2.6. Сравнение результатов частотного анализа русского, английского и французского корпусов на уровне частей речи

Содержащиеся в концептуальных классах лексемы обнаруживают явления концептуальной неоднозначности и концептуального синкретизма, которые могут повлиять на точность автоматической концептуальной разметки в процессе интеллектуального контент-анализа. Было выявлено, что причины концептуальной неоднозначности имеют как лингвистический, так и экстралингвистический характер. Лингвистические причины включают в себя частеречную омонимию, лексическую и синтаксическую неоднозначность; к экстралингвистическим относятся контекстуально обусловленная множественность концептуальных значений при единстве лексического значения и экстралингвистический контекст. В русском корпусе были обнаружены все пять типов концептуальной неоднозначности, в то время как

в английском и французском корпусах отсутствуют примеры синтаксической неоднозначности, влекущие за собой концептуальную неоднозначность.

Отметим, что выявленные ограничения подъязыка не являются столь строгими, как, например, ограничения подъязыка телеграмм или подъязыков технической сферы, и допускают некоторую вариативность лингвистических средств.

Данные, полученные в результате корпусного анализа подъязыка новостных сообщений предметной области «Терроризм», были применены для создания базы знаний модели интеллектуального контент-анализа и базы знаний платформы концептуального аннотирования, которые описываются в следующей главе. Кроме того, некоторые из полученных результатов могут быть при необходимости использованы для дизамбигуации в процессе концептуального аннотирования.

ГЛАВА 3.

ПОСТРОЕНИЕ МОДЕЛИ МНОГОЯЗЫЧНОГО ИНТЕЛЛЕКТУАЛЬНОГО КОНТЕНТ-АНАЛИЗА НОВОСТНЫХ СООБЩЕНИЙ ПРЕДМЕТНОЙ ОБЛАСТИ «ТЕРРОРИЗМ»

3.1 Методологические основы построения модели

Многоязычный интеллектуальный контент-анализ представляет собой извлечение из неструктурированных разноязычных корпусов текстов ограниченной предметной области контента, релевантного информационному запросу пользователя, его числовую обработку, интерпретацию, при необходимости сравнение и представление в удобной для пользователя форме. Специфика поставленной задачи (ориентация на неструктурированные входные данные, необходимость извлечения релевантной информации, требования простоты и удобства представления результатов анализа) определяет методологию настоящего исследования.

В качестве основного метода достижения сформулированной задачи выбрано представление концептуальной структуры (концептуальная разметка) текстов. В разработанной нами модели определение концептуальной структуры основано на онтологическом анализе с использованием специально построенной многоязычной предметно-ориентированной лингвистической онтологии. Многоязычность онтологии в данном случае трактуется как возможность применения одной онтологии к обработке текстов на различных языках.

Данные положения определяют как методологию разработки, так и архитектуру модели интеллектуального контент-анализа, которая состоит из двух компонентов: лексико-онтологической базы знаний и алгоритмических процедур интеллектуального контент-анализа. Архитектура модели интеллектуального контент-анализа изображена в виде блок-схемы на рис. 3.1.

В следующих параграфах данной главы подробно излагается процедура разработки каждого компонента модели, описывается его содержание и принцип действия. В конце главы приводятся примеры использования разработанной модели

многоязычного интеллектуального контент-анализа для решения задач анализа текстов новостных сообщений о террористической деятельности.

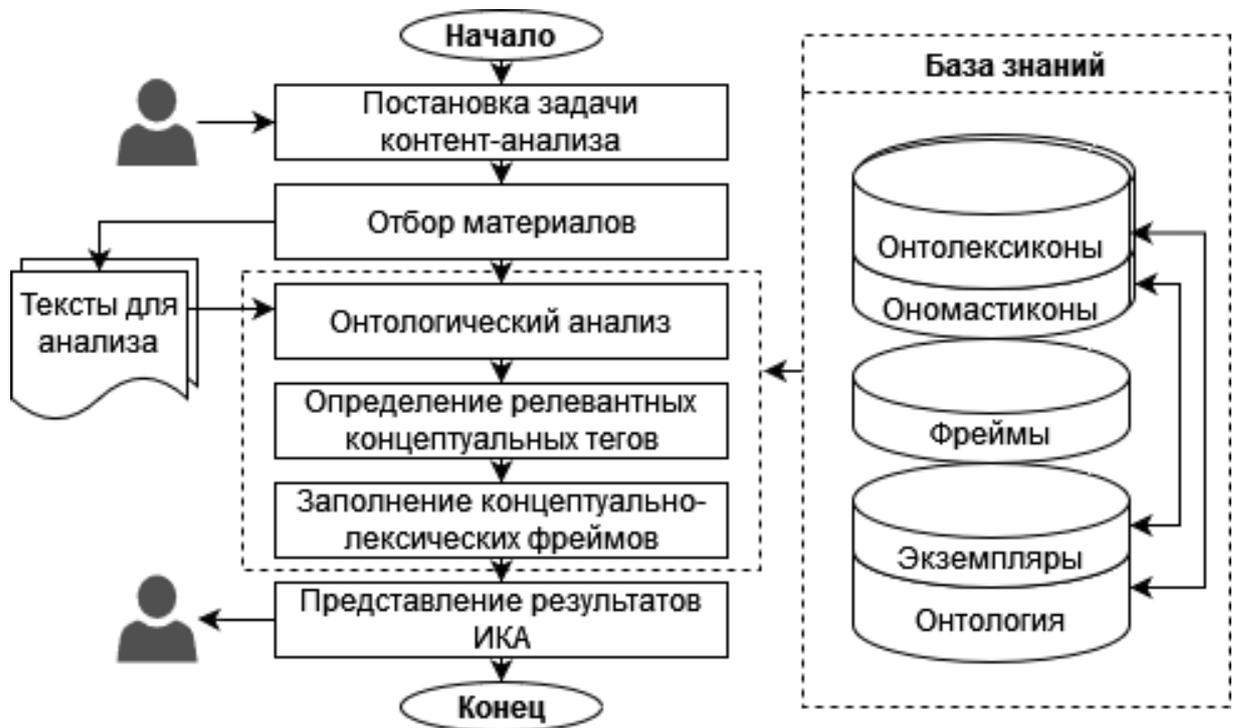


Рисунок 3.1. Архитектура модели интеллектуального контент-анализа

3.2 База знаний модели

База знаний является ключевым компонентом модели многоязычного интеллектуального контент-анализа, наиболее сложным для построения. В базу знаний нашей модели входят следующие компоненты:

- многоязычная онтология предметной области «Терроризм»;
- база экземпляров онтологии;
- онтолексиконы и ономастиконы, содержащие релевантные для предметной области лексемы, в том числе имена собственные, на русском, английском и французском языках;
- правила онтологического анализа и логического вывода;
- правила формирования концептуально-лексических фреймов и оформления результатов работы модели.

В пунктах ниже подробно рассматривается каждый из этих компонентов.

3.2.1 Онтология предметной области «Терроризм»

Многоязычная предметная онтология является основным ресурсом базы знаний модели многоязычного интеллектуального контент-анализа. В соответствии с задачей разработки онтологии к ней предъявлялись следующие требования:

1. Разрабатываемая онтология должна представлять собой независимый от конкретного естественного языка ресурс с возможностью многократного использования для экономии исследовательских усилий.

2. Поскольку «знания о предметной области не изолированы от общих знаний о мире» [Moreno, 2011, с. 233], разрабатываемая онтология предметной области должна быть включена в систему концептов онтологии верхнего уровня.

3. Знания предметной области могут быть получены на основе анализа многоязычных псевдопараллельных (сопоставимых) корпусов текстов с использованием комбинированной методики, включающей в себя методы корпусной лингвистики и неформальные интуитивные (эвристические) методы.

В целях реализации первого требования построение онтологии проводилось на материале трех языков, имеющих расхождения на грамматическом и лексико-семантическом уровнях, как уже было отмечено ранее в главе 2.

Для соответствия второму требованию была выбрана онтология верхнего уровня, задачи которой схожи с задачами настоящего исследования, а именно онтология проекта MikroKosmos (μK) [Viegas и др., 1999; Mahesh, 1996; 2000; Nirenburg и др., 1996; Nirenburg, 2004], изначально разработанная в целях машинного перевода и, как следствие, ориентированная на многоязычие. Один из основных принципов онтологии μK — разделение онтологических и лексических знаний, причем количество лексем многократно превышает количество концептов, что обеспечивает независимость онтологии от конкретного естественного языка.

Соответствие третьему требованию обеспечивалось использованием корпусов новостных сообщений предметной области «Терроризм». В качестве исходного материала для создания онтологии были использованы концептуальные классы, выделенные в ходе корпусного анализа и описанные в главе 2.

Разработанная онтология предметной области «Терроризм» далее подробно описана по пунктам: формализм, концепты, отношения, атрибуты, аксиомы. Отдельно описывается база экземпляров, поскольку, как и в онтологии μK , в онтологии терроризма она представлена отдельным ресурсом, связанным с онтологией.

В качестве языка представления знаний в онтологии предметной области «Терроризм» был принят формализм онтологии μK с незначительными изменениями. В онтологии существует три типа концептов: объекты (OBJECT), события (EVENT) и свойства (PROPERTY), последние при этом подразделяются на отношения (RELATION) и атрибуты (ATTRIBUTE). Концепты организованы в иерархию посредством родовидовых отношений IS-A, но также взаимосвязаны другими отношениями. Таким образом, онтология представлена в виде запутанного дерева концептов.

Все концепты онтологии обладают фреймовой структурой. Фрейм каждого концепта содержит обязательные служебные слоты LABEL (название) и DEFINITION (определение). Названия концептов представляют собой условные обозначения, составленные из слов английского языка и написанные через дефис, при этом содержание концепта обусловлено не названием, а свойствами и определением.

Слоты фрейма могут быть заполнены различными элементами: другими концептами, экземплярами, скалярными величинами, примитивными строковыми литералами, элементами, блокирующими наследование (*nothing*) и т. п. Тип конкретного элемента определяется фасетом, например VALUE, SEM, DEFAULT, NOT. Фасет VALUE ограничивает перечень возможных элементов-заполнителей до фактических значений: экземпляров, литералов, чисел или других концептов (для слотов IS-A, SUBCLASSES, INVERSE, INSTANCE-OF, INSTANCE). Фасет SEM используется для указания других концептов, литералов, чисел или числовых диапазонов. Элементы типа SEM наследуются дочерними концептами и выступают в качестве ограничений на свойства концептов. Фасет DEFAULT описывает ограничение слота по умолчанию. Фасет NOT необходим для исключения из слота конкретного элемента.

Концепты ветви PROPERTY содержат также слоты DOMAIN и RANGE, которые указывают на возможность связывания концептов определенным отношением или

возможность описания концепта определенным атрибутом. Как правило, концепты ветви RELATION требуют указания обратного отношения в слоте INVERSE.

В качестве исходного материала для разработки онтологии терроризма использовались 25 концептуальных классов предметно-релевантной лексики, выделенных на основе анализа корпусов текстов и описанных в главе 2. На основе концептуальных классов были созданы соответствующие концепты онтологии терроризма типов ОБЪЕСТ и EVENT, при этом отдельные концептуальные классы для удобства связи с концептами верхнего уровня онтологии μ К и более точного отражения предметной области были преобразованы в несколько концептов онтологии.

В частности, концептуальный класс «Контртерроризм», в который входили как события (собственно контртерроризм), так и объекты (представители и средства контртерроризма), был преобразован в концепты COUNTERTERRORISM-EVENT, COUNTERTERRORISM-AGENT и COUNTERTERRORISM-INSTRUMENT.

Широкий концептуальный класс «Последствия» был аналогично разделен с образованием трех концептов типа EVENT, обозначающих последствия для жертв (CONSEQUENCES-FOR-PEOPLE), террористов (CONSEQUENCES-FOR-TERRORIST) и ущерб, нанесенный неодушевленным объектам (DAMAGE), и трех концептов типа ОБЪЕСТ, обозначающих жертв (VICTIM), террористов (AFFECTED-TERRORIST) и неодушевленные объекты (AFFECTED-OBJECT), которые пострадали в результате теракта.

В таблице 3.1 дан фрагмент списка концептов типов ОБЪЕСТ и EVENT с определениями на русском языке, полученных на этом этапе.

Таблица 3.1. Базовые объекты и события онтологии терроризма (фрагмент)

Концепт	Определение
CONSEQUENCES-FOR-PEOPLE	Положительные и отрицательные последствия террористического акта для его жертв: погибшие, раненые, выжившие и т. п.
COUNTERTERRORISM-AGENT	Лица и организации, занимающиеся контртеррористической деятельностью
COUNTERTERRORISM-EVENT	Меры по профилактике терроризма, борьбе с терроризмом и контрмеры в ответ на террористические действия
HUMAN-TARGET	Лица, являющиеся объектом (целью) террористического акта

(Продолжение таблицы 3.1)

Концепт	Определение
LOCATION	Место, где произошел террористический акт
NATION	Совокупность людей, организованная под единым правительством; отдельная страна, рассматриваемая вместе с ее общественно-политическими структурами
OBJECT-TARGET	Предметы, на которые направлен террористический акт и которые могут быть отнесены к движимому имуществу, например автомобили
PROPERTY-TARGET	Объекты инфраструктуры и частная собственность, на которые направлен террористический акт
SOURCE	Источник сообщения о террористическом акте и его последствиях: газеты, телеканалы, власти, свидетели
TERROR-ATTACK	Нападение, совершаемое террористом или группой террористов для устрашения населения для достижения политических или иных целей
TERROR-ATTACK-PLANNING-EVENT	Действия по планированию террористического акта
TERRORISM-AGENT	Лицо или группа лиц, занимающиеся террористической деятельностью либо в составе террористической организации, либо самостоятельно и использующие методы террора для достижения своих целей
WEAPON	Оружие или подобные объекты (например, грузовик), используемые для совершения террористического акта, а также функциональные элементы такого оружия

Таким образом, сначала на основе 25 концептуальных классов предметно-релевантной лексики был создан исходный набор концептов онтологии, состоявший из 33 концептов ветвей OBJECT и EVENT. Далее с помощью комбинации известных методов были определены свойства (отношения и атрибуты) концептов предметной области. В частности, для определения отношений был использован метод, предложенный А. Морено [Moreno, 2011], суть которого состоит в анализе связей между лексемами, репрезентующими определенные концепты и находящимися в непосредственной близости друг от друга в корпусе. Например, слово *теракт* часто встречается в русском корпусе рядом с лексемами, обозначающими средство теракта (*взрывное устройство, грузовик*), место происшествия (*метро, стадион*), пострадавших (*человек, мирное население*), исполнителей теракта (*боевик*,

террорист), организаторов теракта (*террористическая группировка*). Данный метод был также применен для анализа английского и французского корпусов. В результате были идентифицированы отношения INSTRUMENT, LOCATED, DIRECTION, AGENT, являющиеся, по существу, семантическими валентностями концепта TERROR-ATTACK.

Отметим, что отношение AGENT и обратное ему отношение AGENT-OF могут быть использованы для описания связи концепта TERROR-ATTACK как с концептом TERRORISM-AGENT, так и с концептом TERRORIST-ORGANIZATION, однако фактически степень участия в теракте террориста и террористической организации различается. Ранее нами был также выделен концепт CLAIM-RESPONSIBILITY, относящийся к ветви EVENT и выражающийся на лексическом уровне с помощью лексем *взять ответственность за / claim responsibility for* (англ.) / *revendiquer* (фр.), который косвенно указывает на TERRORIST-ORGANIZATION как на агенса TERROR-ATTACK. Тем не менее этот концепт отражает не столько тот факт, что организация действительно ответственна за какой-либо теракт, сколько ее намерение убедить в этом общественность. Это подтверждается примерами из корпуса: *«Ответственность за происшедшее взяли на себя две группировки: „Хайет Тахрир аш-Шам“ (бывшая „Джебхат ан-Нусра“, запрещена в РФ) и „Джейш ан-Насер“ (находится в составе „Сирийской свободной армии“))»*³¹, *«Сразу две террористические организации поспешили обозначить свою причастность к обстрелу»*³².

Одним из вариантов решения этой проблемы было связывание концепта CLAIM-RESPONSIBILITY с концептом TERRORIST-ORGANIZATION отношением AGENT, а концептов TERROR-ATTACK и CLAIM-RESPONSIBILITY — отношением вида «причина — следствие», однако это привело бы к отсутствию прямой связи между

³¹ В Сирии боевики сбили российский штурмовик Су-25 // АЕХ.RU. — 2018. — Режим доступа: <https://www.aex.ru/news/2018/2/5/180824>, свободный. — Загл. с экрана (дата обращения: 03.04.2021).

³² Радикальные меры: сразу две группировки взяли на себя ответственность за покушение на главу Пентагона в Кабуле // RT. — 2017. — Режим доступа: <https://russian.rt.com/world/article/434147-afghanistan-terror-taliby-mattis>, свободный. — Загл. с экрана (дата обращения: 03.04.2021).

концептами TERRORIST-ORGANIZATION и TERROR-ATTACK, что лишено логики. Поэтому в онтологию было введено отношение RESPONSIBILITY-CLAIMED-BY и обратное ему отношение CLAIMED-RESPONSIBILITY-FOR, при этом возможность связи концептов TERRORIST-ORGANIZATION и TERROR-ATTACK отношением AGENT была оставлена для тех случаев, когда организатор теракта точно известен (в частности, для репрезентации именной группы *ISIS attack* и ей подобных).

Кроме того, были установлены причинно-следственные связи между концептами TERROR-ATTACK-PLANNING-EVENT, TERROR-ATTACK и CLAIM-RESPONSIBILITY, поскольку террористическая организация может взять на себя ответственность за теракт только в том случае, если он произошел, теракт в свою очередь происходит в том случае, если он запланирован. В онтологии μK уже есть подходящие отношения между событиями: PRECONDITION и PRECONDITION-OF.

Компонентный анализ словосочетаний *исламистский терроризм, fundamentalist terrorism, racist terrorism, terrorisme d'extrême-gauche* <терроризм + идеология> позволил выделить отношение HAS-IDEOLOGY. Аналогичным образом были выделены отношения HAS-TYPE (тип теракта), HAS-GOAL (цель теракта), HAS-RELIGION (религиозная принадлежность террориста или жертвы теракта), HAS-NATIONALITY (национальность террориста или жертвы теракта).

Всего между концептами онтологии было установлено 28 отношений. Примеры выделенных на данном этапе отношений приведены в табл. 3.2 (полный перечень отношений онтологии терроризма представлен в Приложении В).

Таблица 3.2. Отношения онтологии терроризма (фрагмент)

Отношение	Обратное отношение	Определение
AGENT	AGENT-OF	Субъект, совершающий или побуждающий действие
CAPITAL-OF	HAS-CAPITAL	Связывает столицу и государство
INSTRUMENT	INSTRUMENT-OF	Указывает на средство совершения теракта
LOCATED	LOCATION-OF	Указывает на место совершения теракта
MEMBER-OF	HAS-MEMBER	Связывает террориста и организацию
RESPONSIBILITY-CLAIMED-BY	CLAIMED-RESPONSIBILITY-FOR	Связывает теракт с террористической организацией или (реже) террористом, которые заявили, что этот теракт был совершен ими

Определение атрибутов концептов было выполнено посредством компонентного анализа лексем и метода оппозиций. Например, анализ оппозиций *террорист / terrorist / terroriste* — *террористка / female terrorist / femme terroriste* <террорист + женский пол>, *мужчина / man / homme* <человек + мужской пол> — *женщина / woman / femme* <человек + женский пол> позволил идентифицировать атрибут GENDER с возможными значениями MALE и FEMALE, присущий концепту онтологии μ K HUMAN и всем его дочерним концептам, используемым в онтологии терроризма.

Оппозиции *девочка / girl / fille* — *женщина / woman / femme*, *мальчик / boy / garçon* — *мужчина / man / homme*, *ребенок / child / enfant* — *подросток / teenager / adolescent* свидетельствуют о наличии у концепта HUMAN наследуемого атрибута AGE, который может быть заполнен численной величиной больше нуля.

Кроме того, в онтологию был введен атрибут NEGATION, необходимый для корректного извлечения информации, поскольку в текстах нередко встречаются формулировки вида «**Никакая** организация **не** взяла на себя ответственности за теракт», «**Никто не погиб**», «**Полицейские не разглашают** сведения о личности террориста» и пр. В случае отсутствия атрибута NEGATION подобные формулировки могли бы привести к искажению извлеченной из текста информации.

Выделенные атрибуты онтологии терроризма представлены в таблице 3.3 (полный перечень атрибутов приведен в Приложении В).

Таблица 3.3. Атрибуты онтологии терроризма (фрагмент)

Атрибут	Определение	Диапазон
GENDER	Пол человека	MALE, FEMALE
AGE	Возраст человека	> 0
SCALE-OF-ATTACK	Масштабность теракта	TRUE, FALSE
NEGATION	Отрицание	TRUE, FALSE

Полученные по результатам описанных этапов концепты типов OBJECT, EVENT, RELATION и ATTRIBUTE были объединены в группы на основе общих признаков и соотнесены с верхними уровнями онтологии μ K. Фрагмент ветви OBJECT представлен на рис. 3.2 (пунктирная линия обозначает наличие промежуточных концептов верхнего уровня онтологии μ K, не представленных на схеме).

Далее базовые концепты онтологии были детализированы с помощью метода текстовых паттернов — повторяющихся языковых конструкций, представляющих то или иное отношение [Hearst, 1992]. Так, родовидовые отношения X IS-A Y отражены в следующих конструкциях в русском языке: X — это (такой) Y , X является Y , Y такие как X , такие Y как X , X и другие Y ; в английском языке: X is a Y , X and other Y , Y such as X ; во французском языке: X est Y , X et autres Y , Y comme X .

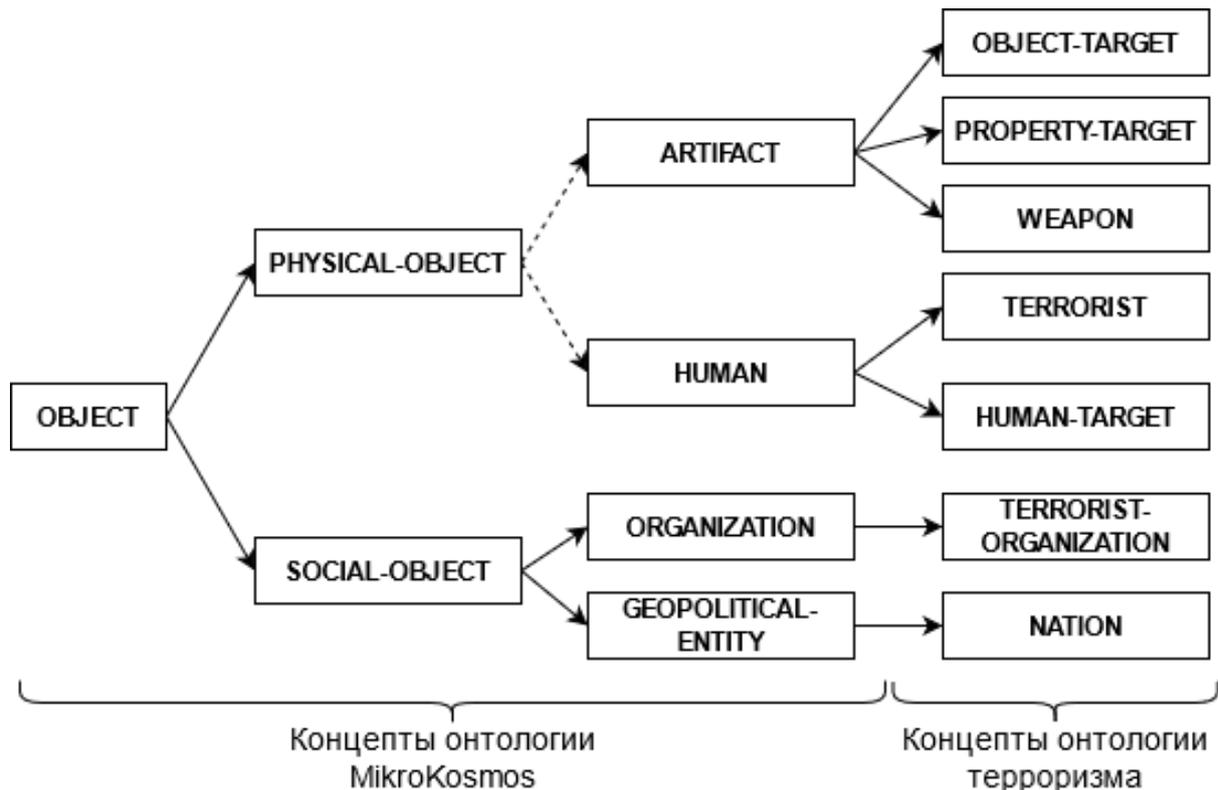


Рисунок 3.2. Фрагмент ветви ОБЪЕКТ в онтологии терроризма после соединения с онтологией верхнего уровня μ К

Помимо отношения IS-A аналогичные текстовые паттерны были также использованы для поиска концептов, связанных отношениями INSTRUMENT и DIRECTION. В частности, названия видов оружия, используемого для совершения теракта, можно обнаружить в тексте при помощи следующих паттернов (X — любое слово, выражающее концепт TERROR-ATTACK, Y — название оружия): X с использованием Y , X с применением Y ; X using Y , X with Y , X involving Y , Y X ; X avec Y , X au moyen de Y , X au (à la) Y . При этом в английском языке для обозначения типа теракта в зависимости от оружия чаще использовался именно четвертый паттерн (*truck attack*, *bomb attack*), однако таким образом могут быть также обозначены

темпоральные параметры (*Saturday's attack, 9/11 attacks*), локации (*Istanbul attack, nightclub attack*), организатор (*Taliban attack, ISIS-claimed attack*), количество взрывов (*double attack, triple attack*). Использование второго и третьего паттерна также не всегда показывает высокую точность, в то время как первый паттерн во всех случаях указал именно на отношение INSTRUMENT.

При поиске во французском корпусе наибольшую точность показал третий паттерн, однако в некоторых случаях он использовался для обозначения отношения LOCATED (ср. *attentat au camion-bélier* и *attentat au Reina*).

Паттерны для русского языка позволили обнаружить отношение INSTRUMENT во всех случаях, однако при поиске необходимо было обращать внимание на дистантное расположение элементов паттерна, характерное только для русского корпуса, например (элементы паттерна выделены жирным шрифтом): «...*серию резонансных диверсионно-террористических акций в городе Москве и городе Санкт-Петербурге с применением автоматического оружия и самодельных взрывных устройств большой мощности*»³³ и «...*совершить теракты в местах массового пребывания людей с применением автоматического оружия и самодельных взрывных устройств*»³⁴. Для обнаружения таких паттернов использовались маски вида $X * с применением Y$, где «*» обозначает любое количество символов.

Применение паттернов, выражающих отношение INSTRUMENT, позволило детализировать концепт WEAPON, для которого были выделены следующие дочерние концепты: BIOLOGICAL-WEAPON (биологическое оружие), CHEMICAL-WEAPON (химическое оружие), GUN (огнестрельное оружие), EXPLOSIVE-DEVICE (взрывное устройство), INCENDIARY-WEAPON (зажигательное оружие), KNIFE (холодное оружие), RAMMING-VEHICLE (автомобиль-таран), ROCKET (ракетное оружие). Отметим, что в ходе анализа паттернов были обнаружены названия и других видов оружия

³³ В России силовики пресекли деятельность террористической группы // Интерфакс. — 2016. — Режим доступа: <https://www.interfax.ru/russia/536703>, свободный. — Загл. с экрана (дата обращения: 03.04.2021).

³⁴ Глава ФСБ РФ Александр Бортников: фигуранты дела о взрыве в петербургском метро планировали новые теракты // Коммерсантъ. — 2017. — Режим доступа: <https://www.kommersant.ru/doc/3305841>, свободный. — Загл. с экрана (03.04.2021).

с более узким значением, например названия различных видов взрывных устройств, различающихся конструктивно (*бомба с часовым механизмом, взрывное устройство с поражающими элементами, pipe bomb, nail bomb, pressure cooker bomb*), однако они не были включены в онтологию как отдельные концепты, поскольку настоящее исследование не требует столь глубокой детализации.

При помощи паттернов, выражающих отношение DIRECTION, был выполнен поиск дочерних концептов HUMAN-TARGET (человек, являющийся объектом теракта) и PROPERTY-TARGET (недвижимое имущество, являющееся объектом теракта).

Паттерны имеют следующий вид:

- для русского языка: TERROR-ATTACK *против* Y, TERROR-ATTACK *направлен** на Y, TERROR-ATTACK *нацелен** на Y, Y *подверг*с** TERROR-ATTACK;
- для английского языка: TERRORIST-ORGANIZATION / TERRORIST-AGENT * *target** Y, TERROR-ATTACK *is directed at* Y;
- для французского языка: TERRORIST-ORGANIZATION / TERRORIST-AGENT * *cibl** Y и т. д. (Y — лексема, указывающая на объект теракта).

Таким образом было установлено, что целями террористов могут являться такие лица, как военные (MILITARY-TARGET), полицейские (POLICE-TARGET), мирные жители (CIVIL-TARGET), политики (POLITICAL-TARGET), представители определенных конфессий (RELIGIOUS-TARGET) и такие объекты инфраструктуры, как правительственные (GOVERNMENTAL-INSTITUTION), полицейские (POLICE-FACILITY), военные (MILITARY-INSTITUTION), учебные (EDUCATIONAL-INSTITUTION), медицинские (MEDICAL-INSTITUTION) и религиозные (RELIGIOUS-INSTITUTION) учреждения, развлекательные (ENTERTAINMENT-FACILITY) и транспортные (TRANSPORT-FACILITY) объекты, а также частная собственность (CIVIL-PROPERTY), исходя из чего в онтологию были введены соответствующие подконцепты.

Каждый внесенный в онтологию концепт был оформлен в виде фрейма. Примеры фреймов онтологических концептов, относящихся к ветвям OBJECT, EVENT, RELATION и ATTRIBUTE, приведены в таблицах 3.4–3.7 соответственно.

Таблица 3.4. Фрейм концепта WEAPON

LABEL	EXPLOSIVE-DEVICE	VAL
DEFINITION	Предмет, снабженный разрывным зарядом и взрываемый с помощью взрывателя, путем удара или иным образом	VAL
IS-A	WEAPON	VAL
SUBCLASSES	CAR-BOMB, SUICIDE-VEST	VAL
INSTRUMENT-OF	BOMB-ATTACK	SEM
WEAPON-OF	TERRORISM-AGENT, ARTIFACT	SEM

Таблица 3.5. Фрейм концепта TERROR-ATTACK

LABEL	TERROR-ATTACK	VAL
DEFINITION	Нападение, совершаемое террористом или группой террористов для устрашения населения для достижения политических или иных целей	VAL
IS-A	CRIMINAL-ACTIVITY	VAL
SUBCLASSES	ARMED-ASSAULT, ARSON, BIOLOGICAL-ATTACK, BOMB-ATTACK, CHEMICAL-ATTACK, DISTANT-ATTACK, HIJACKING, HOSTAGE-TAKING, ASSASSINATION, SUICIDE-ATTACK, UNARMED-ASSAULT, VEHICLE-RAMMING-ATTACK, CYBER-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, OBJECT-TARGET, PROPERTY-TARGET	SEM
INSTRUMENT	WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL

Таблица 3.6. Фрейм концепта RESPONSIBILITY-CLAIMED-BY

LABEL	RESPONSIBILITY-CLAIMED-BY	VAL
DEFINITION	Связывает теракт с террористической организацией или (реже) террористом, которые заявили, что теракт был совершен ими	VAL
IS-A	EVENT-OBJECT-RELATION	VAL
DOMAIN	TERROR-ATTACK	VAL
RANGE	TERRORISM-AGENT, TERRORIST-ORGANIZATION	VAL
INVERSE	CLAIMED-RESPONSIBILITY-FOR	VAL

Таблица 3.7. Фрейм концепта GENDER

LABEL	GENDER	VAL
DEFINITION	Пол человека	VAL
IS-A	LITERAL-ATTRIBUTE	VAL
DOMAIN	HUMAN	VAL
RANGE	MALE, FEMALE	VAL

В онтологию терроризма также был введен ряд утверждений (аксиом), которые связывают различные элементы онтологии и позволяют делать различные умозаключения. Ниже приведены примеры аксиом онтологии:

$$\forall T \text{ located}(T, La) \cap \text{part of}(La, Lc) \rightarrow \text{located}(T, Lc), \quad (1)$$

где T — TERROR-ATTACK, La — CITY, Lc — COUNTRY; т. е. для всех терактов верно, что, если теракт произошел в некоем городе и нам известно, что этот город находится в некоей стране, то, следовательно, теракт произошел в этой стране. В этой аксиоме концепт CITY может быть заменен на концепты CAPITAL-CITY, CITY-AREA, COUNTRY-AREA и SPECIFIC-LOCATION, а концепт COUNTRY — на TERRITORY.

$$\forall T \text{ instrument}(T, Cc) \rightarrow \text{type}(T, Taa), \quad (2)$$

где T — TERROR-ATTACK, Cc — GUN, Taa — GUN-ATTACK; т. е. для всех терактов верно, что если теракт был совершен с использованием огнестрельного оружия, то типом теракта является стрельба. В этой аксиоме концепт GUN может быть заменен на любой дочерний концепт WEAPON, а концепт GUN-ATTACK — на любой дочерний концепт TERROR-ATTACK соответственно.

$$\forall A \text{ member of}(A, U) \cap \text{has ideology}(U, Wa) \rightarrow \text{has ideology}(A, Wa), \quad (3)$$

где A — TERRORISM-AGENT, U — TERRORIST-ORGANIZATION, Wa — ISLAMISM (дочерний концепт IDEOLOGY); т. е. для всех исполнителей терактов верно, что если исполнитель является членом некоей террористической организации и такая террористическая организация проповедует исламизм, то, следовательно, этот исполнитель является исламистом. В этой аксиоме концепт TERRORISM-AGENT может быть заменен на свой дочерний концепт TERRORIST-GROUP, а концепт ISLAMISM — на любой другой дочерний концепт IDEOLOGY.

В конечном варианте онтология терроризма состоит из 113 концептов ветвей OBJECT и EVENT на трех уровнях иерархии, 28 концептов ветви RELATION, пяти концептов ветви ATTRIBUTE и 32 аксиом. Перечень всех концептов онтологии приведен в Приложении В «Иерархическая структура концептов онтологии терроризма». Фреймы концептов представлены в Приложении Г «Фреймы концептов онтологии терроризма». Аксиомы перечислены в Приложении Е.

3.2.2 База экземпляров

Помимо описанных в предыдущем пункте концептуальных знаний, онтология терроризма также содержит блок эпизодических знаний, представленный экземплярами отдельных концептов. В целях наполнения этого блока из разнесенных по концептуальным классам лексических единиц были отобраны имена собственные, большая часть из которых относилась к следующим концептуальным классам: «Страна» (например, *Алжир, Германия, Россия*), «Место теракта» (*аэропорт Нарита, вокзал Виктория*), «Террористическая организация» (*РПК, ИГИЛ, Боко Харам*), «Источник» (*Аль-Джазира, Би-би-си, Бинали Йылдырым*), «Террорист» (*Абу Бакр аль-Багдади, Брентон Таррант*).

При этом более или менее постоянным характером и, как следствие, релевантностью для предметной области обладают только те эпизодические знания, которые относятся к концептуальным классам «Страна», «Место теракта», «Террористическая организация» и «Источник», при этом имена существенной части террористов (за исключением глав организаций) не имеют большого значения и редко упоминаются в новостях в рамках более чем одного информационного повода. В связи с этим в онтологии терроризма концепты только четырех ветвей представлены экземплярами: NATION, LOCATION, TERRORIST-ORGANIZATION и SOURCE.

Как и концепты, экземпляры представлены в виде фреймов. В таблице 3.8 показан пример фрейма для экземпляра TERRORIST-ORGANIZATION-3, обозначающего террористическую организацию «Ястребы свободы Курдистана».

Фрейм содержит слот DEFINITION с кратким определением на естественном языке; слот INSTANCE-OF, соединяющий экземпляр с концептом TERRORIST-ORGANIZATION; слот HAS-IDEOLOGY, в котором указана информация об основной идеологии организации, а также слоты DEFAULT-NAME-RU, DEFAULT-NAME-EN, DEFAULT-NAME-FR, заполненные наиболее употребительными (частотными) языковыми репрезентациями данного экземпляра на русском, английском и французском языках, взятыми из соответствующего ономастикона.

Таблица 3.8. Фрейм экземпляра TERRORIST-ORGANIZATION-3

LABEL	TERRORIST-ORGANIZATION-3	VAL
DEFINITION	«Соколы свободы Курдистана», или «Ястребы свободы Курдистана» — курдская националистическая вооруженная группировка, которая борется за независимое курдское государство в юго-восточной Турции	VAL
INSTANCE-OF	TERRORIST-ORGANIZATION	VAL
HAS-IDEOLOGY	NATIONALISM, SEPARATISM	VAL
DEFAULT-NAME-RU	Ястребы свободы Курдистана	VAL
DEFAULT-NAME-EN	Kurdistan Freedom Hawks	VAL
DEFAULT-NAME-FR	Les Faucons de la liberté du Kurdistan	VAL

Для заполнения фреймов в базе экземпляров были использованы данные из анализируемых корпусов, в случае недостаточности таких данных использовалась информация из открытых источников, в частности сайта Wikipedia.org.

Фреймы некоторых других экземпляров даны в Приложении Д.

3.2.3 Онтолексиконы и ономастиконы

Второй компонент базы знаний — лексиконы релевантных для предметной области одно- и многокомпонентных лексических единиц, соотнесенных с онтологией, т. е. онтолексиконы. Каждый онтолексикон содержит свыше 20 000 единиц, снабженных концептуальными и морфосинтаксическими признаками, а также перекрестными ссылками на соответствующие лексические единицы остальных языков исследования. При этом между единицами каждого из лексиконов и

концептами онтологии наблюдаются отношения N:1 и 1:N, что обусловлено разной концептуальной природой лексических единиц, т. е. их возможной концептуальной однозначностью, концептуальной неоднозначностью и концептуальным синкретизмом. Подробно эти явления были рассмотрены в главе 2.

Следует отметить, что процесс формирования лексиконов был итеративным. Для первоначального формирования лексиконов использовались лексические единицы, извлеченные из исходных корпусов и разнесенные по концептуальным классам, а затем соединенные с концептами разработанной онтологии терроризма. Поскольку подязык новостных сообщений предметной области «Терроризм» не обладает строгими ограничениями на лексическом уровне, полученные лексиконы, вероятно, не обеспечили бы исчерпывающего покрытия каких-либо новых текстов. Поэтому далее были собраны новые корпусы небольшого объема, из которых были извлечены релевантные лексические единицы. Полученные списки лексических единиц были сопоставлены с наполнением лексиконов, и единицы, отсутствовавшие в лексиконах, были в них добавлены и соотнесены с концептами онтологии. Данная процедура повторялась циклически до достижения удовлетворительной покрываемости новых текстов. Отметим, что в ходе данной процедуры возможно не только пополнение лексиконов, но и корректировка концептов онтологии, однако в нашем случае корректировка не потребовалась.

Перекрестные онтолексиконы создавались посредством сопоставления извлеченных из корпусов лексических единиц, в случае отсутствия эквивалента в списках предлагался возможный перевод. Если проверка через интернет показывала существование такого перевода, он сохранялся. Например, для англоязычной именной группы *explosive-laden vehicle* не было найдено соответствия в русском корпусе, однако переводческая гипотеза о существовании именной группы *начиненное взрывчаткой транспортное средство* подтвердилась поиском в Google, поэтому в онтолексиконе указанной англоязычной именной группе был поставлен в соответствие именно этот русскоязычный эквивалент. В ряде случаев лексической единице на одном языке соответствовало несколько эквивалентов, в этом случае такая единица вносилась в онтолексикон несколько раз.

В качестве примера единицы онтолексикона приведем англоязычную именную группу *car bomb*. На концептуальном уровне данная лексическая единица связана с единственным концептом CAR-BOMB (дочерний концепт EXPLOSIVE-DEVICE, который в свою очередь является дочерним концептом WEAPON). Следует отметить, что, если лексическая единица относится к более узкому концепту, она не имеет прямой связи с более широким, т. е. единица *car bomb* не входит в список единиц, связанных с концептами EXPLOSIVE-DEVICE и WEAPON. Таким образом, непосредственную связь с наиболее крупными концептами онтологии имеют только те лексические единицы, которые не удалось распределить по дочерним концептам. Лексическая единица *car bomb* в лексиконе имеет один эквивалент на французском языке (*voiture piégée*) и шесть эквивалентов на русском: *автомобиль со взрывчаткой, автомобильная бомба, бомба на колесах, джихадмобиль, заминированная машина, начиненный взрывчаткой автомобиль*.

В онтолексиконе выделяется особый блок — ономастикон, который содержит языковые репрезентации экземпляров онтологических концептов. Этот ресурс необходим для систематизации языковых репрезентаций, поскольку многие экземпляры имеют несколько вариантов именования и написания (последнее особенно характерно для арабских названий, как правило, имеющих множественные варианты передачи на языки с латинскими и кириллическими алфавитами). Например, экземпляр концепта онтологии TERRORIST-ORGANIZATION с условным наименованием TERRORIST-ORGANIZATION-3 в соответствующих одноязычных лексиконах связан с единицами *ТАК, Ястребы свободы Курдистана, Соколы свободы Курдистана (рус.), Kurdistan Freedom Hawks, Kurdistan Freedom Falcons (англ.), Faucons de la liberté du Kurdistan (фр.)*, которые представляют собой синонимы. В свою очередь экземпляр TERRORIST-ORGANIZATION-8 в русском ономастиконе имеет варианты написания *Хезболла, Хезболлах, Хизбалла, Хизбаллах*, в английском — *Hizbollah, Hezbolla, Hezbollah, Hisbollah, Hizbu'llah* и *Hizb Allah*. Языковые репрезентации экземпляров содержатся в ономастиконе в виде именных и предложных групп.

Три лексемы для каждого концепта и экземпляра являются их названиями по умолчанию на соответствующем языке и указываются в слотах DEFAULT-NAME-

RU, DEFAULT-NAME-EN и DEFAULT-NAME-FR для удобства восприятия результатов интеллектуального контент-анализа пользователем, который является носителем одного из указанных языков. В случае экземпляров это наиболее частотные лексемы (*Хезболла* для TERRORIST-ORGANIZATION-8), в случае концептов — наиболее полных описывающие (*последствия* для CONSEQUENCES).

3.2.4 Правила онтологического анализа

Онтологический анализ (онтоанализ) представляет собой процесс извлечения знаний о сущностях определенной предметной области [Façanha, 2019]. На практике онтологический анализ заключается в отображении лексических единиц текста на концепты онтологии с последующей формализацией и интерпретацией результатов такого отображения в зависимости от конкретной задачи. Отображение лексических единиц на концепты выполняется с помощью разметки лексических единиц кодами (тегами), присвоенными концептам онтологии.

Данная процедура онтоанализа обозначается синонимичными терминами «концептуальная разметка» и «концептуальное аннотирование», на английском языке — conceptual annotation [Sheremetyeva, 2020], concept labeling [Viju, 2018]. Концептуальная разметка представляет собой не что иное, как частный случай семантической разметки, т. е. обогащения текста семантической информацией.

В отличие от семантической разметки, концептуальная разметка выполняется на основе онтологии или иной концептуальной модели и ориентирована на ограниченную предметную область. Например, в Национальном корпусе русского языка лексеме *ребенок* присвоены семантические метки ‘предметное имя’, ‘лицо’ [Нац. корп. рус. яз.], а в исследуемом нами корпусе текстов предметной области «Терроризм» данная лексема может быть размечена тегом концепта CONSEQUENCES-FOR-PEOPLE, так как ребенок может стать жертвой теракта.

Исходя из вышеизложенного, в настоящей диссертационной работе мы используем термин «концептуальная разметка», подразумевая отдельный вид

семантической разметки текстов с помощью концептуальных тегов предметной области, и отличаем его от более общего термина «семантическая разметка».

Правила онтологического анализа в рамках модели интеллектуального контент-анализа включают формальную разметку (кодирование) текстов тегами онтологических концептов и разрешение концептуальной неоднозначности. Под формальной концептуальной разметкой понимается присвоение лексическим единицам текста тегов всех концептов, с которыми они связаны в онтологии.

Для концептуальной разметки концептам онтологии терроризма, которые могут быть эксплицитно представлены в тексте, присвоены специальные теги: например, концепту CONSEQUENCES поставлен в соответствие тег Р, его дочерним концептам CONSEQUENCES-FOR-PEOPLE, CONSEQUENCES-FOR-TERRORIST и DAMAGE — теги P_a, P_b и P_c соответственно, а дочерним концептам CONSEQUENCES-FOR-PEOPLE — в свою очередь теги P_{aa} (PEOPLE-DYING), P_{ab} (PEOPLE-INJURING), P_{ac} (TAKEN-NOSTAGE), P_{ad} (VICTIM-SURVIVAL). Полный список тегов приведен в Приложении В.

В связи с отсутствием взаимно-однозначных соответствий между лексическими единицами и концептами онтологии формальная концептуальная разметка неизбежно вызывает присвоение некоторым лексическим единицам комбинаций двух и более тегов (мультитегов), не все из которых являются релевантными в конкретном контексте. Часть таких мультитегов обусловлена концептуальным синкретизмом лексических единиц, то есть наличием у единиц двух и более непротиворечивых концептуальных значений в контексте или вне его. Остальные случаи присвоения лексической единице мультитега являются проявлениями концептуальной неоднозначности. Для решения этой проблемы используется процедура разрешения концептуальной неоднозначности.

3.2.5 Проблема концептуальной неоднозначности при онтоанализе

В целях изучения возможных методов разрешения концептуальной неоднозначности был проведен тестовый анализ случайных выборок из каждого корпуса (подкорпусов) новостных сообщений предметной области «Терроризм» (объем

выборок указан в таблице 3.9). Отметим, что в рамках данного этапа исследования перед нами не стояла задача сопоставления репрезентации концептуальной неоднозначности в корпусах на разных языках, поэтому объем выборок и (в ряде случаев) проверяемые методы различаются.

Таблица 3.9. Объем выборок для анализа концептуальной неоднозначности

Корпус	Русский	Английский	Французский
Объем выборки			
словоупотреблений	~26 000	~25 000	~45 000
текстов	262	107	143

При анализе нами была использована упрощенная процедура разметки: для удобства подсчетов были взяты исходные крупные концепты онтологии, соответствующие концептуальным классам лексики (т. е. концепт CONSEQUENCES не был разделен, концепты, относящиеся к контртеррористической деятельности, объединены в один концепт COUNTERTERRORISM, концепты, обозначающие объект теракта, — в концепт TARGET). Им были присвоены следующие теги: А = TERRORISM-AGENT (исполнитель теракта), В = TIME (время теракта), С = WEAPON (средство теракта), CR = CLAIM-RESPONSIBILITY (взятие ответственности за теракт), D = DECLARE (заявление), DA = DIRECTION (направление теракта), E = TERRORIST-THREAT (террористическая угроза), G = RELIGION (религия), HC = HAS-WEAPON (владение средством теракта), I = ASSUME (предположение), K = TERRORIST-ATTACK-PLANNING-EVENT (планы террористов), L = LOCATION (место теракта), M = SCALE-OF-ATTACK (масштаб теракта), N = NATION (нация), O = OTHER-TERRORIST-EVENT (прочая террористическая деятельность), P = CONSEQUENCES (последствия), Q = MEMBER-OF (принадлежность к террористической организации), R = COUNTERTERRORISM (контртерроризм), S = SOURCE (источник сообщения), T = TERROR-ATTACK (тип теракта), U = TERRORIST-ORGANIZATION (террористическая организация), W = IDEOLOGY (идеология), X = GOAL-OF-ATTACK (цель теракта), Y = COMMUNITY-REACTION (реакция общественности), Z = TARGET (объект теракта).

Подкорпусы были формально размечены на концептуальном уровне посредством присвоения единицам текста тегов всех концептов, с которыми они связаны

в онтологии, после чего сотрудниками НОЦ «ЛИНТ» ЮУрГУ (в том числе автором диссертационного исследования) была вручную устранена концептуальная неоднозначность. Пример формальной и эталонной разметки изображен в таблице 3.10.

Таблица 3.10. Формальная и эталонная разметка новостного сообщения подкорпуса на русском языке

Формальная разметка	Эталонная разметка
{В Нигере}~L~N при {атаке}~T на {военный пост}~Z {погибли}~P {89}~Num {человек}~P~Z~R~A	{В Нигере}~L~N при {атаке}~T на {военный пост}~Z {погибли}~P {89}~Num {человек}~P
{89}~Num {военнослужащих}~P~Z~R {стали жертвами}~P {атаки}~T {неизвестных}~A {на военную базу}~L {в Республике Нигер}~L~N, {пишет}~D {Reuters}~S .	{89}~Num {военнослужащих}~P {стали жертвами}~P {атаки}~T {неизвестных}~A {на военную базу}~L {в Республике Нигер}~L~N, {пишет}~D {Reuters}~S .
{По данным издания}~S , {в настоящее время}~B {против}~DA {вооруженных сил республики}~Z {действуют}~O {боевики}~P~A~R, {связанные с}~Q {запрещенной в РФ международной террористической организацией Аль-Каида}~A~U и {радикальной исламистской группировкой Боко Харам}~A~U .	{По данным издания}~S , {в настоящее время}~B {против}~DA {вооруженных сил республики}~Z {действуют}~O {боевики}~A, {связанные с}~Q {запрещенной в РФ международной террористической организацией Аль-Каида}~U и {радикальной исламистской группировкой Боко Харам}~U .

В результате сравнительно-сопоставительного анализа формально размеченных и эталонных подкорпусов были выявлены некоторые количественные методы снятия концептуальной неоднозначности, а также разработаны продукционные правила снятия неоднозначности вида ЕСЛИ-ТО-ИНАЧЕ. Кроме того, были определены синкретичные мультитеги, которые не требуют разрешения неоднозначности. Сначала приведем описание количественных методов.

Ранжирование тегов. В общих чертах ранжирование тегов аналогично методу, описанному в работе [Рахилина и др., 2006], где было предложено отсортировать лексические значения лексем на основе корпусных данных вместо данных словарей и таким образом установить иерархию значений для разрешения семантической неоднозначности в НКРЯ. Этот метод может быть применен к разрешению концептуальной неоднозначности, если будет построена аналогичная иерархия тегов в мультитегах для высокочастотных лексических единиц, для которых такая иерархия может быть построена на основе корпусных данных. Результаты

ранжирования тегов высокочастотных лексических единиц на материале русского, английского и французского подкорпусов представлены в таблицах 3.11–3.13.

Таблица 3.11. Ранжирование тегов высокочастотных лексем русского корпуса

Лексема (частота)	Мультитег	Варианты разрешения (вероятность)
человек (184)	P~Z~R~A	P: 176 (0,956), Z: 7 (0,038), A: 1 (0,005)
полиция (80)	P~S~Z~R	R: 60 (0,750), S: 15 (0,188), Z: 5 (0,063)
автомобиль (48)	P~Z~C	C: 33 (0,688), Z: 11 (0,229), P: 4 (0,083)

Таблица 3.12. Ранжирование тегов высокочастотных лексем английского корпуса

Лексема (частота)	Мультитег	Варианты разрешения (вероятность)
kill (130)	P~T~R	P: 90 (0,692), P~T: 19 (0,146), P~R: 14 (0,008), T: 4 (0,031), R: 3 (0,023)
people (63)	P~S~Z	P: 57 (0,905), Z: 5 (0,079), S: 1 (0,016)
police (55)	P~S~Z~R	R: 29 (0,527), S: 23 (0,418), Z: 2 (0,036), P: 1 (0,018)

Таблица 3.13. Ранжирование тегов высокочастотных лексем французского корпуса

Лексема (частота)	Мультитег	Варианты разрешения (вероятность)
personne (197)	P~Z~R~A	P: 176 (0,893), A: 11 (0,056), Z: 7 (0,036), R: 3 (0,015)
police (56)	P~S~Z~R	S: 33 (0,590), R: 21 (0,375), Z: 2 (0,036)
autorité (45)	S~R	S: 36 (0,800), R: 9 (0,200)

Метод ранжирования тегов обладает двумя очевидными недостатками. Во-первых, значительное количество лексических единиц в подкорпусах имеют недостаточную частоту встречаемости для построения достоверной иерархии тегов. Во-вторых, в некоторых случаях частоты тегов, присвоенных одной лексической единице, равны либо очень близки, и поэтому ни одному из этих тегов нельзя отдать предпочтение. Следовательно, данный метод подходит только для ограниченного числа лексических единиц, у которых наблюдается значительный перевес частоты в пользу одного из тегов (можно определить основное концептуальное значение).

Во избежание проблем, связанных с низкой частотой встречаемости отдельных лексических единиц, возможно осуществить ранжирование тегов без учета

таковых. Метод ранжирования тегов без учета лексических единиц был испытан только на англоязычном подкорпусе, результаты для двух наиболее частотных мультитегов L~N и P~T~R представлены в таблице 3.14.

Таблица 3.14. Ранжирование тегов без учета лексики (английский корпус)

Мультитег (частота)	Варианты разрешения (вероятность)
L~N (235)	L~N: 213 (0,906), N: 22 (0,093)
P~T~R (203)	P: 104 (0,512), T: 35 (0,172), P~T: 27 (0,133), P~R: 27 (0,133), R: 10 (0,049)

Мультитег L~N в анализируемом подкорпусе в большинстве случаев проявляет синкретичный характер и, как следствие, не разрешается. Тем не менее в 9 % случаев L~N разрешается в пользу N и никогда — в пользу L. Это объясняется тем, что нация, национальность в отличие от местоположения является неотъемлемым атрибутом кого- или чего-либо. Мультитег P~T~R имеет пять вариантов разрешения, из которых два — в пользу синкретичных тегов P~T и P~R. Интересно, что мультитег P~T~R никогда не разрешается в пользу T~R, поскольку концепты, обозначаемые этими тегами, являются дизъюнктивными (какое-либо действие не может быть терактом и контртеррористическим действием одновременно).

Ранжирование тегов без учета лексики также обнаруживает свои недостатки по сравнению с методом, учитывающим лексические единицы. В ряде случаев необходимость и вероятность разрешения неоднозначности в пользу того или иного тега зависит от лексической единицы и даже словоформы, в которой она представлена. Например, для лексемы *Pakistan* процент разрешения неоднозначности L~N в пользу N составляет 5 %, в то время как для лексемы *Italy* — 75 % (т. е. в рассматриваемой выборке Пакистан значительно чаще, чем Италия, выступает как место совершения терактов). Мультитег P~T~R, присваиваемый страдательному причастию *shot* и действительному причастию *shooting*, в первом случае чаще всего разрешается в пользу P~R, во втором — в пользу T.

Таким образом, оба варианта этого метода не лишены недостатков, тем не менее использование данного метода при автоматизированном разрешении неоднозначности может упростить выбор тега для кодировщика.

Позиционный метод. В основе данного метода лежит классическая идея Г. Эдмундсона об идентификации ключевых слов по их расположению в тексте [Edmundson, 1969] и данные о структуре новостных текстов, полученные при анализе подязыка. Принимая во внимание, что бóльшая часть текстов исследуемых корпусов построена согласно принципу перевернутой пирамиды и имеет характерную суперструктуру, можно выдвинуть гипотезу, что доля лексических единиц, обозначающих различные концепты, зависит от позиции предложения в тексте.

Материалом для проверки данной гипотезы послужили эталонные подкорпусы, из которых были удалены лексические единицы и оставлены только теги. Количественные характеристики подкорпусов приведены в таблице 3.15.

Тексты каждого из подкорпусов были разделены на предложения, после чего на их основе были сформированы позиционные подкорпусы предложений (подкорпус первых предложений, подкорпус вторых предложений и т. д.) для каждого языка. Для каждого подкорпуса была посчитана относительная частота концептуальных тегов и построены графики, отражающие изменение доли концептуальных тегов в зависимости от позиции предложения в тексте.

Таблица 3.15. Количественные характеристики корпусов для проверки позиционного метода снятия концептуальной неоднозначности

Корпус	Кол-во текстов	Разброс длины текста (предложения)	Модальная длина текста (предложения)	Медианная длина текста (предложения)
Русский	262	3–27	8	9
Английский	107	3–44	7	10
Французский	143	3–43	14	14

Следует отметить, что в выборке наблюдается большой разброс длин новостных сообщений: максимальная длина сообщений достигает от 27 до 44 предложений в зависимости от корпуса, при этом медианная и модальная длины текстов значительно меньше. Количество необычно длинных текстов незначительно, т. е. данные о частоте встречаемости в позиционных подкорпусах с порядковым номером существенно выше медианной длины получены на материале небольшого

количества текстов. Следовательно, этими данными можно пренебречь, и далее учитываются только данные позиционных подкорпусов с порядковым номером не выше медианной длины текста в соответствующей выборке.

В результате анализа графиков были выделены следующие закономерности распределения относительной частоты разных тегов в позиционных корпусах:

- 1) изначально высокая частота с последующим снижением;
- 2) резкий рост частоты после заголовка с последующим снижением;
- 3) изначально низкая частота с последующим увеличением;
- 4) равномерное распределение (характерно для низкочастотных тегов).

Данные закономерности в основном схожи в выборках на всех трех языках. В целях разрешения неоднозначности могут быть полезны первые три закономерности. Например, на рис. 3.3 показано распределение тегов T (TERROR-ATTACK) и R (COUNTERTERRORISM) в английском подкорпусе: в то время как частота тега R увеличивается к концу среднестатистического новостного сообщения, частота тега T, напротив, уменьшается и в конечном счете становится меньше частоты тега R. Основываясь на этих данных, можно снимать неоднозначность мультитега T~R.

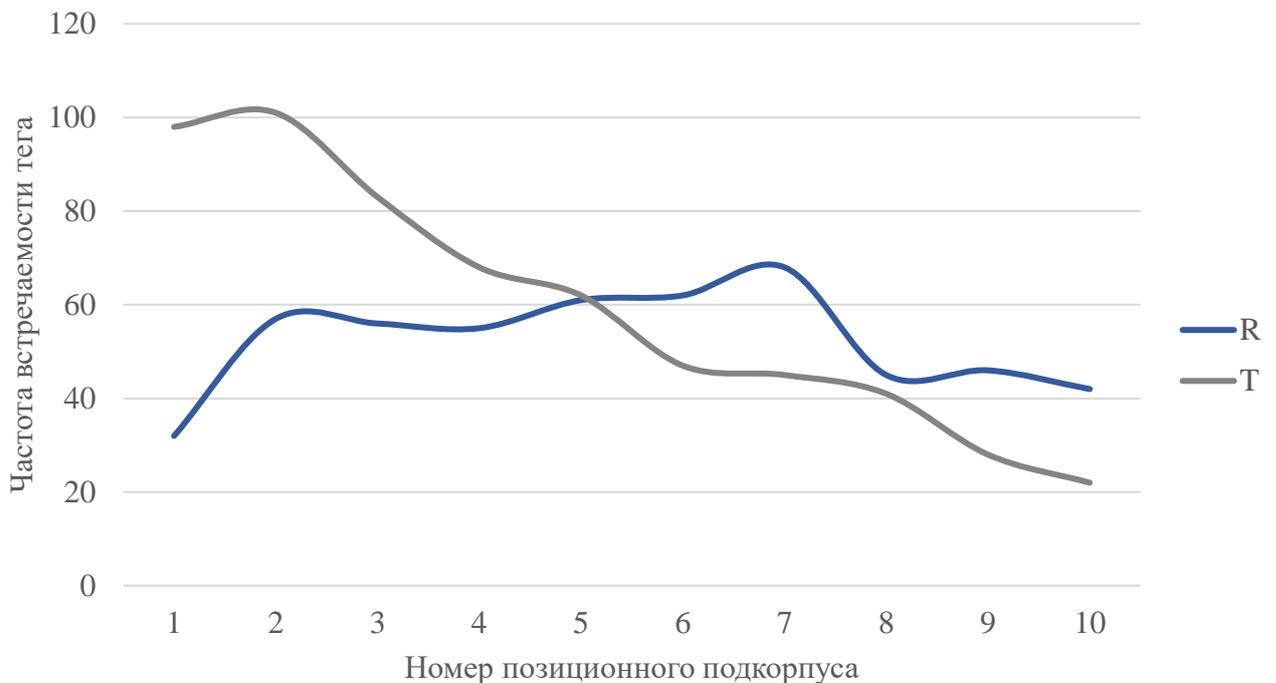


Рисунок 3.3. Распределение тегов T и R в английском подкорпусе

Подчеркнем, что из-за отсутствия в подязыке строгих ограничений на уровне структуры релевантности и суперструктуры позиционный метод целесообразнее использовать в совокупности с какими-либо другими методами.

Помимо количественных методов снятия концептуальной неоднозначности, на основе сравнительно-сопоставительного анализа формально размеченных и эталонных подкорпусов был также выведен ряд продукционных правил ЕСЛИ-ТО-ИНАЧЕ, которые затем были уточнены с учетом концептов всех уровней онтологии терроризма. Ниже приведены правила снятия неоднозначности для некоторых высокочастотных мультитегов, универсальные для всех языков исследования:

Правило 1.

ЕСЛИ за мультитегом $S \sim Z_a(x)^{35} \sim R_a$ следует тег D ,
ТО преобразовать $S \sim Z_a(x) \sim R_a$ в S .

Правило 2.

ЕСЛИ за мультитегом $A(x) \sim H_b$ следует тег $H_b(x)$,
ТО оставить $A(x) \sim H_b$ без изменений,
ИНАЧЕ преобразовать $A(x) \sim H_b$ в $A(x)$.

Правило 3.

ЕСЛИ перед мультитегом $A \sim U$ есть тег Q ,
ТО преобразовать $A \sim U$ в U .

Правило 4.

ЕСЛИ за мультитегом $Z_a(x) \sim A(x) \sim H_a \sim H_b \sim R_a$ следует тег U ,
ТО преобразовать $Z_a(x) \sim A(x) \sim H_a \sim H_b \sim R_a$ в $A(x)$.

Правило 5.

ЕСЛИ перед мультитегом $Z_a(x) \sim A(x) \sim H_a \sim H_b \sim R_a$ есть тег $P_b(x) \sim R_b$,
ТО преобразовать $Z_a(x) \sim A(x) \sim H_a \sim H_b \sim R_a$ в $A(x) \sim H_b$.

³⁵ Под $Z_a(x)$ понимается любой тег из списка: Z_{aa} , Z_{ab} , Z_{ac} , Z_{ad} , Z_{ae} .

Правило 6.

ЕСЛИ перед мультитегом $Za(x) \sim A(x) \sim Na \sim Nb \sim Ra$ есть теги Num и $Pa(x)$,
ТО преобразовать $Za(x) \sim A(x) \sim Na \sim Nb \sim Ra$ в Na .

Комбинации следующих тегов, как правило, обусловлены концептуальным синкретизмом и не требуют снятия концептуальной неоднозначности (примеры приведены на русском языке, однако данные не являются лингвоспецифичными):

1. Тег N (NATION) и теги $Za(x)$ (HUMAN-TARGET), Na (VICTIM), $A(x)$ (TERRORISM-AGENT), S (SOURCE). В следующем примере концепт, репрезентуемый тегом N, представляет собой заполнитель слота HAS-NATIONALITY концепта INDIVIDUAL-TERRORIST: *{28-летний австралиец Брентон Таррант}*~ $Aa \sim N-9$.

2. Теги $W(x)$ концептов ветви IDEOLOGY и теги $A(x)$, U. В следующем примере концепт, репрезентуемый тегом Wa , заполняет слот HAS-IDEOLOGY концепта TERRORISM-AGENT: *{исламистский террорист}*~ $A \sim Wa$.

3. Теги $G(x)$ концептов ветви RELIGION и теги $Za(x)$, Na , $A(x)$.

4. Тег M и теги $T(x)$ (TERROR-ATTACK). В примере далее тег M представляет атрибут SCALE-OF-ATTACK концепта TERROR-ATTACK: *{крупнейший теракт}*~ $T \sim M$.

5. Тег I (ASSUMPTION) и любые другие теги. В данном случае тег I указывает на предположительный характер помеченного фрагмента: *{подозреваемый}*~ $A \sim I$.

6. Теги AGE, GEN (атрибуты AGE, GENDER) при совместном употреблении с тегами $Za(x)$, Na , A указывают на возраст или пол объектов, жертв теракта или террористов соответственно.

7. Тег NEG и любые другие теги. Тег NEG указывает на обратный характер помеченного им текстового фрагмента: *{никто}*~ $Na \sim NEG$ *{не погиб}*~ $Pa \sim NEG$.

В заключение настоящего пункта отметим, что данные о возможных методах разрешения концептуальной неоднозначности, полученные на выборочных фрагментах исследуемых корпусов, представляют собой достаточно ценное знание для дальнейших исследований новостных сообщений предметной области «Терроризм». Полное же решение данной задачи не входит в объем настоящего исследования и является заделом на нашу будущую работу.

3.2.6 Прочие компоненты базы знаний

Прочие компоненты базы знаний включают правила логического вывода и правила формирования динамических концептуально-текстовых фреймов и оформления результатов работы модели интеллектуального контент-анализа.

Правила логического вывода используются при отсутствии в тексте лексических единиц, размеченных релевантными для поставленной задачи контент-анализа тегами, т. е. в случаях, когда информация не представлена эксплицитно на текстовом уровне. Правила логического вывода основаны на аксиомах онтологии.

Концептуально-лексические фреймы необходимы для первичного представления и количественной обработки извлеченной информации. Они создаются динамически на основе данных, введенных пользователем на этапе постановки задачи интеллектуального контент-анализа. Процедура создания и заполнения фреймов автоматизирована с помощью авторской программы-экстрактора, формирующей текстовый файл с возможностью преобразования в табличный формат.

Концептуально-лексический фрейм содержит слоты, соответствующие выявленным в тексте релевантным концептам и их экземплярам и заполняющиеся фрагментами текста, размеченными тегами данных концептов, с указанием частоты встречаемости концепта и конкретного фрагмента текста. Фрейм содержит идентификатор текста (заданное пользователем имя файла). При генерации фрейма могут использоваться символы «*», «?» и «!»: символом «*» обозначаются концепты, полученные посредством логического вывода, «?» — предположительные концепты (фрагмент текста помечен тегом I атрибута ASSUMPTION), «!» — концепты, наличие которых в тексте отрицается (фрагмент помечен тегом NEG атрибута NEGATION). Приведем пример фрейма для следующего фрагмента новостного сообщения: «В результате {теракта}~T {в Лондоне}~Laa-186 {никто}~Ha~NEG {не погиб}~Paa~NEG. {Подозреваемый}~A~I~Hb {арестован}~Pbd»:

AFFECTED-TERRORIST {1}

подозреваемый {1}

!CONSEQUENCES-FOR-PEOPLE {1}

—

!PEOPLE-DYING {1}	не погиб {1}
*COUNTRY-186 {1}	—
CITY {1}	—
CAPITAL-CITY-186 {1}	в Лондоне {1}
CONSEQUENCES-FOR-TERRORIST {1}	—
TERRORIST-CAPTURING {1}	арестован {1}
TERROR-ATTACK {1}	теракта {1}
?TERRORISM-AGENT {1}	подозреваемый {1}
!VICTIM {2}	никто {1}

Отметим, что даже если теги концептов онтологии первого уровня отсутствуют в тексте, их частота тем не менее подсчитывается во фрейме на основании частоты дочерних концептов или экземпляров (в примере выше так получены частоты CITY, CONSEQUENCES-FOR-PEOPLE и CONSEQUENCES-FOR-TERRORIST).

При формировании конечных результатов интеллектуального контент-анализа условные названия концептов для удобства русскоязычного пользователя заменяются на слова русского языка, содержащиеся в слоте DEFAULT-NAME-RU. Подвергнутые числовой обработке результаты оформляются в виде таблиц или графиков, например, средствами программы Microsoft Excel.

3.3 Алгоритм модели

Алгоритм модели многоязычного интеллектуального контент-анализа, в котором используется описанная в предыдущем разделе база знаний, представлен в центральной части блок-схемы архитектуры модели интеллектуального контент-анализа на рис. 3.1 (с. 101). Данный алгоритм включает в себя этап определения задачи интеллектуального контент-анализа и процедуры ее решения, выполняемые вручную или автоматически на основе авторских программ. На вход каждой из процедур подается информация, полученная на выходе предыдущей процедуры алгоритма. Последовательность основных процедур алгоритма приведена ниже:

1. Постановка задачи контент-анализа.
2. Отбор материала.
3. Онтологический анализ.
4. Определение релевантных концептуальных тегов.
5. Заполнение фреймов.
6. Представление результатов контент-анализа.

Постановка задачи контент-анализа требует четкой формулировки, поскольку именно на этом этапе определяются предпосылки для отбора текстового материала и идентификации релевантных концептуальных тегов.

После формулирования задачи интеллектуального контент-анализа первой выполняется процедура отбора релевантного текстового материала на одном или нескольких языках в зависимости от поставленной задачи. На вход модели могут поступать как корпуса текстов, так и отдельные тексты. Отбор материала производится методом критериальной выборки через интернет-агрегатор новостей по ключевым словам. Ключевыми считаются слова, которые входят в формулировку задачи интеллектуального контент-анализа. Репрезентативность выборки обеспечивается путем фильтрации текстового материала по следующим критериям:

1. Текст опубликован в открытом источнике, находящемся в свободном доступе и не требующем платной подписки.
2. Текст не является републикацией текста, уже отобранного в корпус.
3. Текст может содержать информацию о нескольких релевантных запросу событиях, если отдельных текстов о каждом событии не обнаружено.

Следующие процедуры алгоритма применяются последовательно к каждому отдельному тексту результирующего корпуса.

Сначала выполняется онтологический анализ, который состоит из:

- а) подпроцедуры разметки текста тегами концептов онтологии на основе онтолексиконов, выполняемой автоматически;
- б) подпроцедуры снятия концептуальной неоднозначности, выполняемой вручную или с частичной автоматизацией; при этом теги концептуально синкретичных лексем не требуют разрешения.

В ходе следующей процедуры определяется набор релевантных задаче контент-анализа концептуальных тегов. Эта процедура включает в себя:

- а) подпроцедуру определения фрагмента онтологии, содержащего релевантные для поставленной задачи концепты (и репрезентующие их теги);
- б) подпроцедуру логического вывода, которая актуализируется только в том случае, если в тексте отсутствуют лексемы, релевантные для поставленной задачи, т. е. если релевантная информация не представлена в тексте эксплицитно.

Подпроцедура логического вывода основана на анализе связей онтологических концептов и позволяет получить на выходе дополнительные релевантные для поставленной задачи теги, а также размеченные ими лексические единицы анализируемого текста. Например, концепт GUN (дочерний концепт WEAPON, связанного отношением INSTRUMENT-OF с концептом TERROR-ATTACK) связан отношением INSTRUMENT-OF с концептом GUN-ATTACK и может быть использован для логического вывода информации о типе теракта, если в тексте таковая отсутствует.

После идентификации релевантных тегов выполняется процедура извлечения информации и заполнения концептуально-лексического фрейма, формируемого динамически на основе вводных данных. Процедура выполняется автоматически с помощью разработанного автором экстрактора. Затем данные в заполненных фреймах обсчитываются и выдаются пользователю. Представление результатов исследования выполняется в форме таблиц или графиков.

3.4 Платформа концептуального аннотирования

С целью автоматизации онтологического анализа сотрудниками НОЦ «Лингво-инновационные технологии» Южно-Уральского государственного университета при непосредственном участии автора настоящего исследования была разработана платформа концептуального аннотирования (ПАНТ) для обработки текстов предметной области «Терроризм» на русском и английском языках. Кроме того, автором данной работы была выполнена подготовка платформы к обработке французского языка. В настоящем разделе приводится краткое описание

платформы (подробное описание с примерами на материале русского и английского языков представлено в работах [Шереметьева, 2020; Sheremetyeva, 2020]) и демонстрируется процесс и результат ее настройки на французский язык.

Платформа ПАНТ состоит из базы знаний и программной оболочки. База знаний платформы содержит три основных компонента: а) концептуальные знания предметной области, независимые от конкретного естественного языка, т. е. концепты онтологии терроризма, которым присвоены специальные теги; б) лингвоспецифические знания предметной области, которые представляют собой одноязычные взаимосвязанные посредством перекрестных ссылок лексиконы русского, английского и французского языков (таблица 3.16); в) сведения о связи между концептуальными и лингвоспецифическими знаниями предметной области.

Таблица 3.16. Фрагмент трехязычного лексического наполнения базового концепта онтологии терроризма TERRORISM-AGENT

Русский язык	Английский язык	Французский язык
боевик	militant	combattant
джихадист	jihadi	djihadiste
смертник	suicide bomber	bombe humaine
террорист	terrorist	terroriste
террористка-смертница	female suicide bomber	femme kamikaze
террорист-одиночка	lone-wolf terrorist	loup solitaire
шахид	shaheed	chahid

Вследствие концептуального синкретизма и концептуальной неоднозначности одна лексическая единица может быть связана в базе знаний с несколькими концептами, что отражено в таблице 3.17.

Таблица 3.17. Примеры лексем на русском, английском и французском языках, связанных с несколькими концептами

Лексема	Тег концепта							
	A	Cd	Le	N	Pa	Pb	Td	Zb
взрыв бомбы		*					*	
водитель автомобиля	*				*			
жертва взрыва					*		*	

(Продолжение таблицы 3.17)

Лексема	Тег концепта							
	A	Cd	Le	N	Pa	Pb	Td	Zb
российский турист				*	*			
airport			*					*
airport shooting suspect	*		*					
detainee	*					*		
night club attack			*				*	
attaque à la bombe								
bombe humaine	*	*						
hôpital			*					*
journaliste italien				*	*			

Программная оболочка ПАНТ состоит из модуля сбора и хранения знаний, который представляет собой адаптированный к контент-анализу электронный словарь TransDict [Sheremetyeva, 2017], и модуля концептуальной разметки (концептуального теггера). Оба модуля имеют пользовательский интерфейс на английском языке, который является основным языком международной коммуникации.

Модуль сбора и хранения знаний платформы представляет собой набор одноязычных электронных лексиконов с перекрестными ссылками между отдельными словарными статьями. Лингвистическая информация о каждой одноязычной единице лексикона формализована в виде дерева свойств следующим образом:

[концепт [язык [класс³⁶ [морфологические признаки [тег]]]]].

Словарная статья одноязычного лексикона предназначена для отражения одного концептуального значения лексической единицы в соответствующем корпусе текстов предметной области. Если единица является концептуально неоднозначной и (или) концептуально синкретичной, для каждого ее концептуального значения создается отдельная словарная статья. Словарная статья разделена на семантическую и морфосинтаксическую зоны: в первой представлена информация

³⁶ Лексико-грамматический класс.

о концептуальном классе лексической единицы (концепте), во второй иконически приведены словоформы, входящие в ее словоизменительную парадигму, при этом для каждой словоформы даны подробные сведения о части речи и грамматических признаках (роде, числе, падеже, залоге и т. п. — в зависимости от языка). Каждой словоформе соответствует супертег, кодирующий всю связанную с ней концептуальную и морфологическую информацию в виде дерева тегов.

Пример словарной статьи TransDict приведен на рис. 3.4.

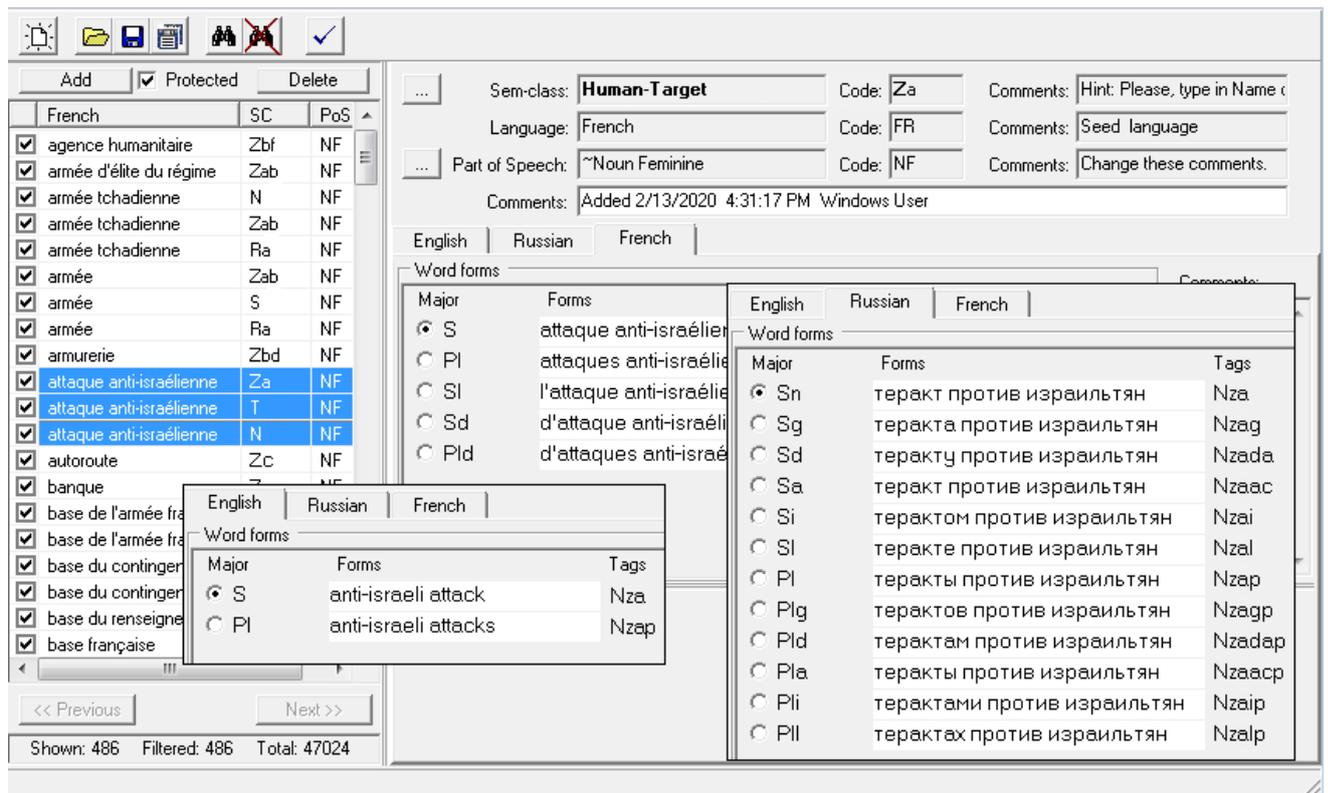


Рисунок 3.4. Интерфейс модуля сбора и хранения знаний ПАНТ с примерами словарных статей на французском, английском и русском языках

В левой части окна расположен фрагмент алфавитного перечня лексических единиц на выбранном языке (в данном случае французском), для каждой из которых указан тег связанного с ней концепта (например, N — NATION, Ra — COUNTERTERRORISM-AGENT, S — SOURCE, Za — HUMAN-TARGET) и тег лексико-грамматического класса (NF — нарицательное имя существительное женского рода). Некоторые лексические единицы дублируются (например, синим выделена единица *attaque anti-israélienne*, представленная в лексиконе трижды), что обусловлено их концептуальной неоднозначностью или концептуальным синкретизмом.

В правой части окна представлена словарная статья для описания лексической единицы *attaque anti-israélienne*, привязанной к концепту HUMAN-TARGET. В статье приведена информация о языке единицы, связанном с ней концепте и лексико-грамматическом классе. Ниже перечислены все словоформы лексической единицы с соответствующими им супертегами, в которых закодирована информация о категории определенности — неопределенности, числе и падеже (для других частей речи возможно кодирование других категорий, что будет подробно рассмотрено ниже). Помимо вкладки **French**, в которой приведены сведения о лексической единице французского языка, доступны вкладки **English** и **Russian**, на которых представлены английский и русский эквиваленты лексической единицы, к которым пользователь может получить доступ, кликнув по соответствующей вкладке.

Модуль сбора и хранения знаний ПАНТ снабжен рядом программных компонентов, которые позволяют в автоматизированном режиме выполнять сбор, администрирование и кодирование концептуально-лингвистических знаний [Sheremetyeva, 2020]. Из всех компонентов программы отметим морфологический генератор, в который нами были внесены изменения для французской базы знаний. Морфологический генератор автоматически генерирует словоформы для конкретной лексической единицы и заполняет поля морфологической зоны словарной статьи с автоматическим присвоением тегов, указанных в настройках конфигурации.

Концептуальный теггер ПАНТ получает на вход неразмеченный текст и выдает его с разметкой на уровне концептуальных тегов (рис. 3.5) или супертегов. Концептуальный теггер также состоит из нескольких компонентов, из которых в настоящей работе следует упомянуть блок снятия неоднозначности, который позволяет создавать, хранить и использовать продукционные правила вида ЕСЛИ... ТО для разрешения концептуальной неоднозначности.

В процессе подготовки платформы ПАНТ к обработке французского языка изменениям подверглись следующие компоненты модуля сбора и хранения знаний: набор лексико-грамматических классов, морфологическая зона словарной статьи и морфологический генератор.

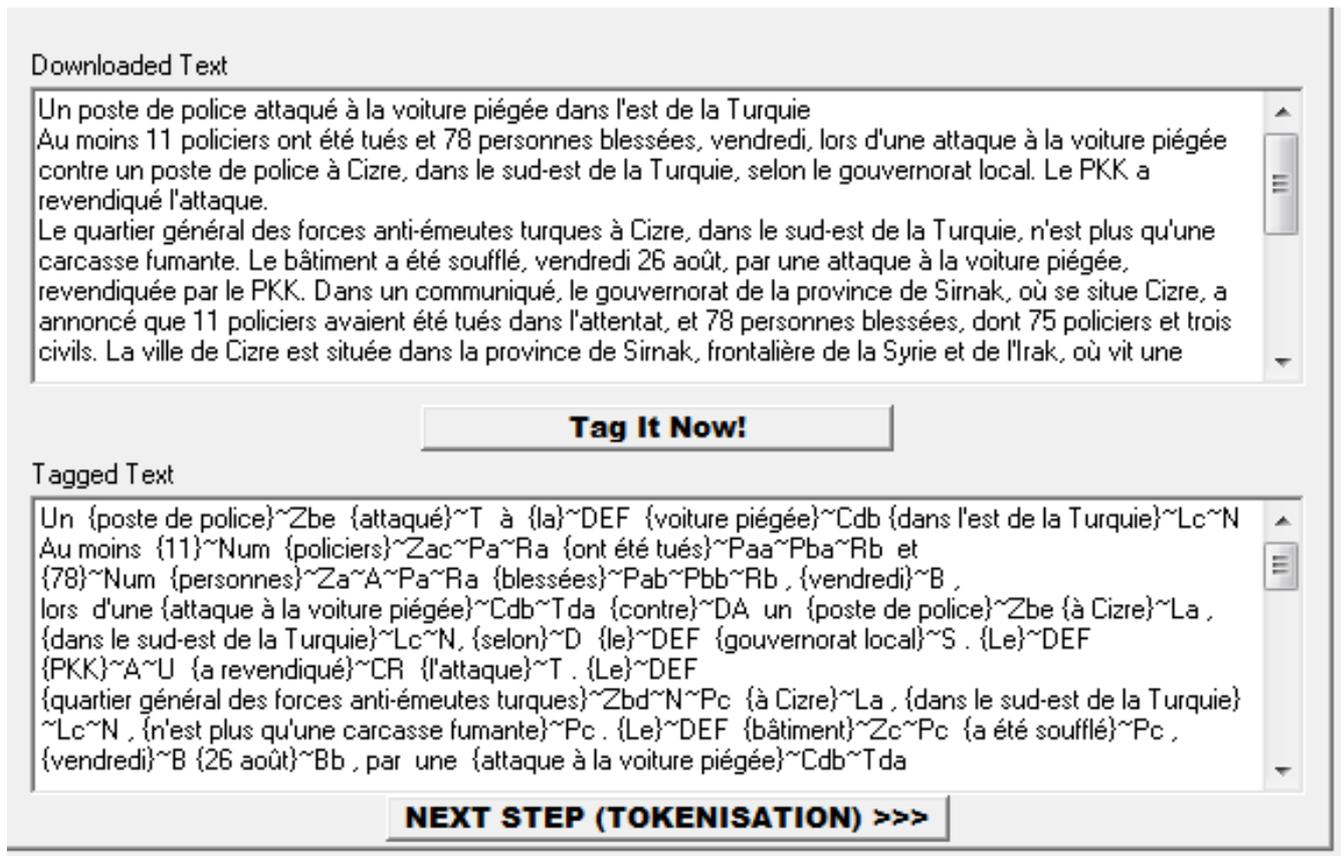


Рисунок 3.5. Интерфейс концептуального теггера платформы ПАНТ с примером размеченного текста на французском языке

Словарная статья лексической единицы французского языка, связанной с любым концептом онтологии, может быть отнесена к одному из девяти лексико-грамматических классов: имя собственное, существительное женского рода, существительное мужского рода, одушевленное существительное, прилагательное, наречие, предложная группа, глагол в активной форме и глагол в пассивной форме.

Морфологическая зона словарной статьи была изменена для всех лексико-грамматических классов кроме наречия и предложной группы, которые являются неизменяемыми и в морфологической зоне имеют только одно поле во всех языках.

Словоизменительная парадигма французских нарицательных имен существительных в ПАНТ отличается от таковой в традиционном понимании и содержит пять словоформ. Это обусловлено тем, что во французском языке артикль *le* и предлог *de* пишутся слитно с существительными, начинающимися с гласного звука (*l'attentat, d'explosion*). Следовательно, необходимо либо постоянно выполнять предобработку французского текста перед интеллектуальным контент-анализом,

отделяя элементы *l'* и *d'* от существительного, либо вносить существительные в базу знаний сразу с этими элементами. В связи с тем, что второй подход представляется нам более экономичным с точки зрения времени и усилий, в морфологическую зону французского нарицательного имени существительного было добавлено пять полей. На рис. 3.6 показаны морфологические зоны нарицательного имени существительного для английского, французского и русского языка, отражающие число словоформ в парадигме существительного в соответствующем языке.

English	Russian	French
Word forms		
Major	Forms	Tags
<input checked="" type="radio"/> S	attack	Nft
<input type="radio"/> Pl	attacks	Nftp

English	Russian	French
Word forms		
Major	Forms	Tags
<input checked="" type="radio"/> Sn	атака	Nftn
<input type="radio"/> Sg	атаки	Nftg
<input type="radio"/> Sd	атаке	Nftda
<input type="radio"/> Sa	атаку	Nftac
<input type="radio"/> Si	атакой	Nfti
<input type="radio"/> Sl	атаке	Nftl
<input type="radio"/> Pl	атаки	Nftnp
<input type="radio"/> Plg	атак	Nftgp
<input type="radio"/> Pld	атакам	Nftdap
<input type="radio"/> Pla	атаки	Nftacp
<input type="radio"/> Pli	атаками	Nftip
<input type="radio"/> Pll	атаках	Nftlp

English	Russian	French
Word forms		
Major	Forms	Tags
<input checked="" type="radio"/> S	attaque	Nft
<input type="radio"/> Pl	attaques	Nftp
<input type="radio"/> Sl	l'attaque	Nftsl
<input type="radio"/> Sd	d'attaque	Nftsd
<input type="radio"/> Pld	d'attaques	Nftpd

Рисунок 3.6. Морфологическая зона словарной статьи английской лексемы *attack* и ее эквивалентов на французском и русском языках

В тех случаях, когда существительное начинается на согласный звук, три дополнительных поля морфологической зоны не требуют заполнения.

Имена собственные, в отличие от нарицательных, как правило, не имеют формы множественного (или, в редких случаях, единственного) числа, поэтому для них число полей было сокращено до трех.

Французские прилагательные изменяются по родам (женский, мужской) и числам (единственное, множественное), поэтому количество полей в морфологической зоне для прилагательного было установлено равным четырем.

Наречия и предложные группы имеют по одной форме.

Состав морфологической зоны глаголов в активной и пассивной форме значительно отличается от словоизменительной парадигмы глагола в стандартном французском языке. Это вызвано тем, что, по результатам корпусного анализа,

подязык новостных сообщений предметной области «Терроризм» не имеет столь богатой системы глагольных словоформ, как соответствующий стандартный язык. Перечисление всех возможных словоформ времязатратно и, следовательно, в рамках нашей задачи не имеет смысла. Поэтому для французской версии ПАНТ были отобраны самые частотные словоформы с учетом изменения по лицам, числам и родам — 52 формы для активного глагола и 27 форм для пассивного глагола.

Далее нами был модифицирован морфологический генератор для автоматизации заполнения полей морфологической зоны активных и пассивных глаголов. Для формирования моделей генерации словоформ был проведен анализ глаголов французского корпуса, в результате чего были сделаны следующие выводы:

1) правильные французские глаголы относятся к одному из двух спряжений: первому (с окончанием *-er*) или второму (с окончанием *-ir*);

2) неправильные французские глаголы объединены в третье спряжение, представленное несколькими группами глаголов, которые спрягаются по разным моделям: например, по модели глагола *battre* ‘бить’ спрягаются глаголы *abattre* ‘сбивать’, *combattre* ‘бороться’ и *contre-battre* ‘сопротивляться’, по модели глагола *couvrir* ‘покрывать’ изменяются глаголы *découvrir* ‘раскрывать’ и *ouvrir* ‘открывать’, по модели глагола *conduire* ‘вести’ — глаголы *introduire* ‘вводить’ и *détruire* ‘разрушать’ и т. п.; отдельно выделяются модели спряжения глаголов *aller* ‘идти’, *être* ‘быть’, *avoir* ‘иметь’ (представленный список не является исчерпывающим);

3) в корпусе есть местоименные глаголы, образующиеся при помощи частицы *se* перед основой (*se faire exploser* ‘подорваться’); они образуют сложные времена с помощью глагола *être* и могут относиться к любому из трех спряжений;

4) глаголы могут образовывать сложные времена при помощи *être* или *avoir*;

5) глаголы могут начинаться на согласный или гласный звук; во втором случае в некоторых формах в начало необходимо будет добавлять элемент *d'* (в форме инфинитива с частицей *de*), *j'* (в форме первого лица единственного числа) и *n'* (в отрицательной форме);

б) основа глагола может заканчиваться на *g* или *s*; в первом случае ряд форм требует добавления буквы *e* перед регулярным окончанием для сохранения

правильного звучания *g* (*charger* ‘обвинять’ — *chargeant*); во втором случае аналогичные формы требуют изменения *c* на *ç* (*foncer* ‘бросаться’ — *fonçant*);

7) основа глагола может заканчиваться на *y*; в этом случае в некоторых формах требуется заменять *y* на *i* (*déployer* ‘развертывать’ — *déploie*);

8) в аналитических формах между вспомогательной и смысловой частью глагола может присутствовать отрицательная частица *pas* или наречие.

Отметим, что большое количество моделей спряжения глаголов третьей группы, которые также могут изменяться в соответствии с вышеприведенными правилами 3–8, может усложнить сценарий автоматической генерации словоформ для глагола в активной форме. Чтобы избежать этого, было принято решение создать алгоритм генерации не для каждой модели спряжения, а только для тех глаголов, модели которых имеют высокую частоту встречаемости в подязыке или которые могут образовывать сложные именные и глагольные сказуемые.

Например, в подязыке очень часто встречаются глаголы, спрягающиеся по модели *mettre* ‘класть’, в частности, глагол *commettre* ‘совершить’, широко использующийся в таких словосочетаниях, как *commettre l'acte de terrorisme* ‘совершить теракт’, *commettre l'attaque à l'arme blanche* ‘совершить нападение с холодным оружием’, *commettre un triple attentat suicide à la bombe* ‘совершить тройной самоподрыв’. Кроме того, высокочастотными являются глаголы *avoir* ‘иметь’, *être* ‘быть’, *faire* ‘делать’, *pouvoir* ‘мочь’, глаголы, спрягающиеся по модели глаголов *battre* ‘бить’ (*abattre* ‘сбивать’, *combattre* ‘бороться’, *contre-battre* ‘сопротивляться’), *prendre* ‘брать’, *rendre* ‘возвращаться’ (*défendre* ‘защищать’, *entendre* ‘слышать’, *suspendre* ‘приостанавливать’, *se rendre* ‘сдаваться’), *tenir* ‘держат’ (*intervenir* ‘вмешиваться’, *soutenir* ‘поддерживать’, *venir* ‘приходить’).

При этом, несмотря на то что глаголы *savoir* ‘знать’, *dire* ‘говорить’, *mourir* ‘умирать’, *voir* ‘видеть’ сами по себе имеют высокую частоту встречаемости в корпусе, в рассматриваемом подязыке нет других глаголов, спрягающихся по этим моделям; также они не образуют сложных сказуемых и, как следствие, должны быть внесены в словарь только единожды, поэтому создавать для них модель автоматической генерации словоформ нецелесообразно.

В результате анализа французских глаголов было описано 45 классов спряжения активного глагола и один класс спряжения пассивного глагола.

В рамках элементарно-комбинаторного подхода к морфологической генерации модель словоформы французского глагола можно представить следующим образом: (*префикс*) + *основа* + *окончание*. Под префиксами в данном случае подразумеваются любые элементы, стоящие перед основной глагола, а именно сокращенные формы местоимения *je*, отрицательной частицы *ne*, возвратной частицы *se*, частицы *de*, инвертированные формы *a-t-elle*, *a-t-il*, *a-t-on*, а также вспомогательные глаголы *être* и *avoir*. В зависимости от словоформы префикс может отсутствовать, основа является неизменяемой, окончание изменяется.

Например, для первой группы спряжения активных глаголов, основа которых начинается на гласный звук и заканчивается на *g*, фрагмент модели генерации словоформ выглядит следующим образом:

Participle présent	∅ + основа + <i>eant</i>
Infinitif	∅ + основа + <i>er</i>
Présent (1, ед. ч.)	<i>j'</i> + основа + <i>e</i>
Présent (1, мн. ч.)	∅ + основа + <i>eons</i>
Passé composé (3, мн. ч., жен.)	<i>ont</i> + основа + <i>é</i>
Imparfait (3, мн. ч.)	∅ + основа + <i>eaient</i>
Plus-que-parfait (3, ед. ч., муж.)	<i>avait</i> + основа + <i>é</i>
Subjonctif présent (3, ед. ч.)	∅ + основа + <i>e</i>
Инверсивная форма (безл.)	<i>a-t-on</i> + основа + <i>é</i>

Фрагмент модели генерации словоформ пассивных глаголов дан ниже:

Participle passé (ед. ч., муж.)	∅ + основа + <i>é</i>
Participle passé (ед. ч., жен.)	∅ + основа + <i>ée</i>
Plus-que-parfait (3, мн. ч., муж.)	<i>avaient été</i> + основа + <i>és</i>
Plus-que-parfait (3, мн. ч., жен.)	<i>avaient été</i> + основа + <i>ées</i>
Conditionnel présent (3, ед. ч., муж.)	<i>serait</i> + основа + <i>é</i>

На основе описанных моделей генерации словоформ активных и пассивных глаголов французского языка был разработан алгоритм автоматизированного заполнения полей морфологической зоны глаголов в модуле сбора и хранения знаний ПАНТ. При использовании алгоритма в первую очередь вручную определяется класс глагола: в случае пассивного глагола используется одна модель генерации, в случае активного глагола по формальным признакам (первая и последняя буква основы, наличие возвратной частицы и т. п.) выбирается одна из 45 моделей, после чего все поля заполняются основой и запускается алгоритм, который автоматически генерирует все необходимые синтетические и аналитические формы. Для внесения в поля с аналитическими формами наречий или частиц между вспомогательным и смысловым глаголом предусмотрена команда, которая автоматически вставляет требуемые элементы перед последним или предпоследним (в зависимости от конкретного поля) элементом.

Описанная платформа концептуального аннотирования ПАНТ может быть использована для полной автоматизации формальной концептуальной разметки текста в ходе онтологического анализа и частичной автоматизации снятия концептуальной неоднозначности. Кроме того, формализация базы знаний в виде электронного словаря обеспечивает возможность использовать заложенные в ней концептуальные и лингвистические знания в целях машинного перевода.

3.5 Примеры использования модели

В данном разделе поэтапно описаны примеры использования модели интеллектуального контент-анализа применительно к реальным задачам. Приведено три примера: анализ тенденций в отношении количества терактов и способов их осуществления на материале англоязычного корпуса, анализ характеристик исполнителей теракта на материале русскоязычного корпуса, а также анализ межъязыковой репрезентации одного теракта в различных СМИ.

3.5.1 Пример первый: анализ тенденций в англоязычном корпусе

Этап 1. Постановка задачи интеллектуального контент-анализа

Проследить изменение количества терактов и способов их осуществления в различных странах и на различных территориях мира в начале 2019 и 2020 гг.

Этап 2. Отбор материала

С учетом критериев отбора через интернет-агрегатор «Google Новости» по ключевым словам *terror attack*, *terrorist attack*, *act of terrorism* с ограничением по датам публикации 01.01.2019–31.03.2019 и 01.01.2020–31.03.2020 соответственно собрано два корпуса англоязычных сообщений о терактах объемом 45 и 32 новостных сообщения (разница в количестве текстов в корпусах обоснована тем, что именно такое количество текстов удовлетворило критериям выборки).

Этап 3. Онтологический анализ

Ниже приведен один из текстов корпуса на выходе процедуры онтологического анализа, включающего концептуальную разметку текста и разрешение концептуальной неоднозначности:

```
{Attempted stabbing terror attack}~Tab~K {near Hebron}~La-86,
{terrorist}~Hb~Aa {killed}~Pba~Rb
```

```
{An}~DEF {attempted stabbing terror attack}~Tab~K occurred {near
Hebron}~La-86 {on Monday}~B as {the}~DEF {terrorist}~Hb~Aa {was
shot}~Pba~Rb by {IDF forces}~Ra, {according to}~S {IDF Spokesperson's
Unit}~S.
```

```
{According to}~S {the}~DEF {report}~D, {the}~DEF {terrorist}~A
{attempted}~K to {stab}~Tab {soldiers}~Zab stationed at {an}~DEF {IDF
post}~Le {in Kiryat Arba}~La-102. {The}~DEF {force}~Ra {shot}~Rb and
{killed}~Pba~Rb {the}~DEF {terrorist}~Hb~A as he entered {a}~DEF
{building}~Le close by.
```

```
{No Israelis}~Ha~N-83~NEG {were injured}~Pab in {the}~DEF
{attempted attack}~T~K.
```

Этап 4. Определение релевантных концептуальных тегов

В соответствии с поставленной задачей требовалось выделить ветви дерева онтологии с вершинами LOCATION и TERROR-ATTACK, а теги узлов этих ветвей считать релевантными (рис. 3.7).

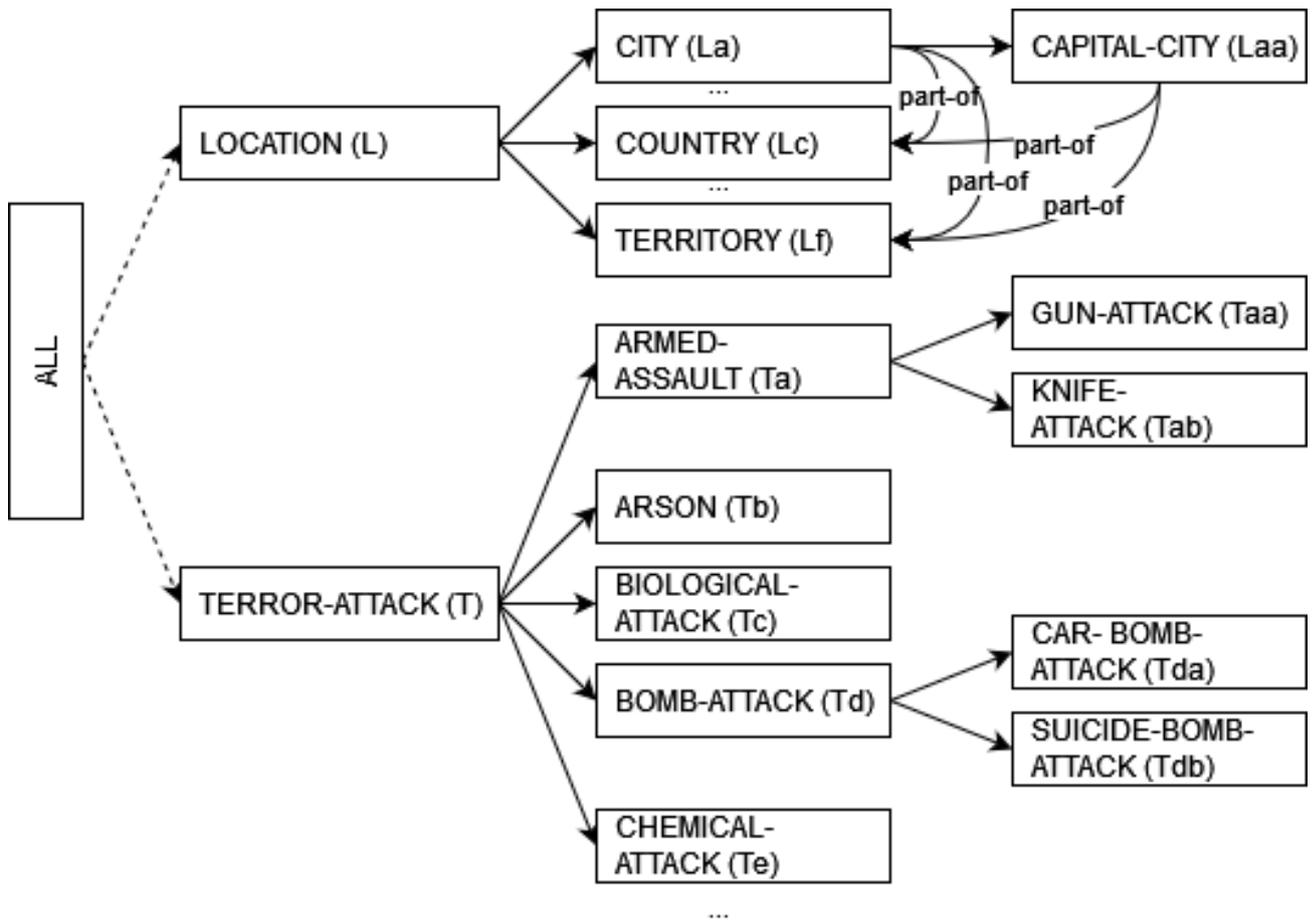


Рисунок 3.7. Фрагмент онтологии терроризма: ветви LOCATION и TERROR-ATTACK

В размеченном тексте приведенного выше примера отсутствуют лексические единицы, эксплицитно выражающие концепты COUNTRY или TERRITORY (т. е. в размеченном тексте нет тегов Lc и Lf), однако есть единицы, репрезентующие концепт CITY (La), который связан отношением PART-OF с концептами COUNTRY и (или) TERRITORY: это предложные группы *in Kiryat Arba* ‘в городе Кирьят-Арба’ и *near Hebron* ‘недалеко от Хеврона’, включенные в онтологию как экземпляры CITY-102 и CITY-86 соответственно и связанные отношением PART-OF с экземпляром TERRITORY-2 (Западный берег реки Иордан). Следовательно, на основании

приведенной в п. 3.2.1 аксиомы «Если некий теракт был совершен в некоем городе и этот город является частью некой территории, то этот теракт был совершен на этой территории» можно сделать вывод, что территорией, на которой был совершен описываемый теракт, является Западный берег реки Иордан.

Этап 5. Заполнение фреймов

Полученная на предыдущем этапе релевантная информация была извлечена из анализируемых текстов, после чего на ее основе были сформированы концептуально-лексические фреймы. Ниже дан фрейм приведенного ранее текста; символом «*» отмечены результаты логического вывода.

Идентификатор текста:	20190312JP.txt
*TERRITORY {3}	—
*TERRITORY-2 {3}	—
CITY {3}	—
CITY-86 {2}	near Hebron {2}
CITY-102 {1}	in Kiryat Arba {1}
SPECIFIC-LOCATION {2}	IDF post {1}, building {1}
TERROR-ATTACK {1}	attempted attack {1}
KNIFE-ATTACK {3}	attempted stabbing terror attack {2}, stab {1}

Этап 6. Представление результатов интеллектуального контент-анализа

Результаты контент-анализа представлены графически средствами MS Excel. На рис. 3.8–3.9 приведены картограммы распределения терактов по странам и территориям в первом квартале 2019 и 2020 гг. соответственно, на рис. 3.10 представлены способы осуществления (типы) терактов (языковые репрезентации на русском языке для представления концептов на графиках взяты из слотов DEFAULT-NAME-RU соответствующих концептов). Полученные результаты контент-анализа демонстрируют количественные и качественные изменения террористической деятельности и ее освещения в СМИ в выбранный период времени.

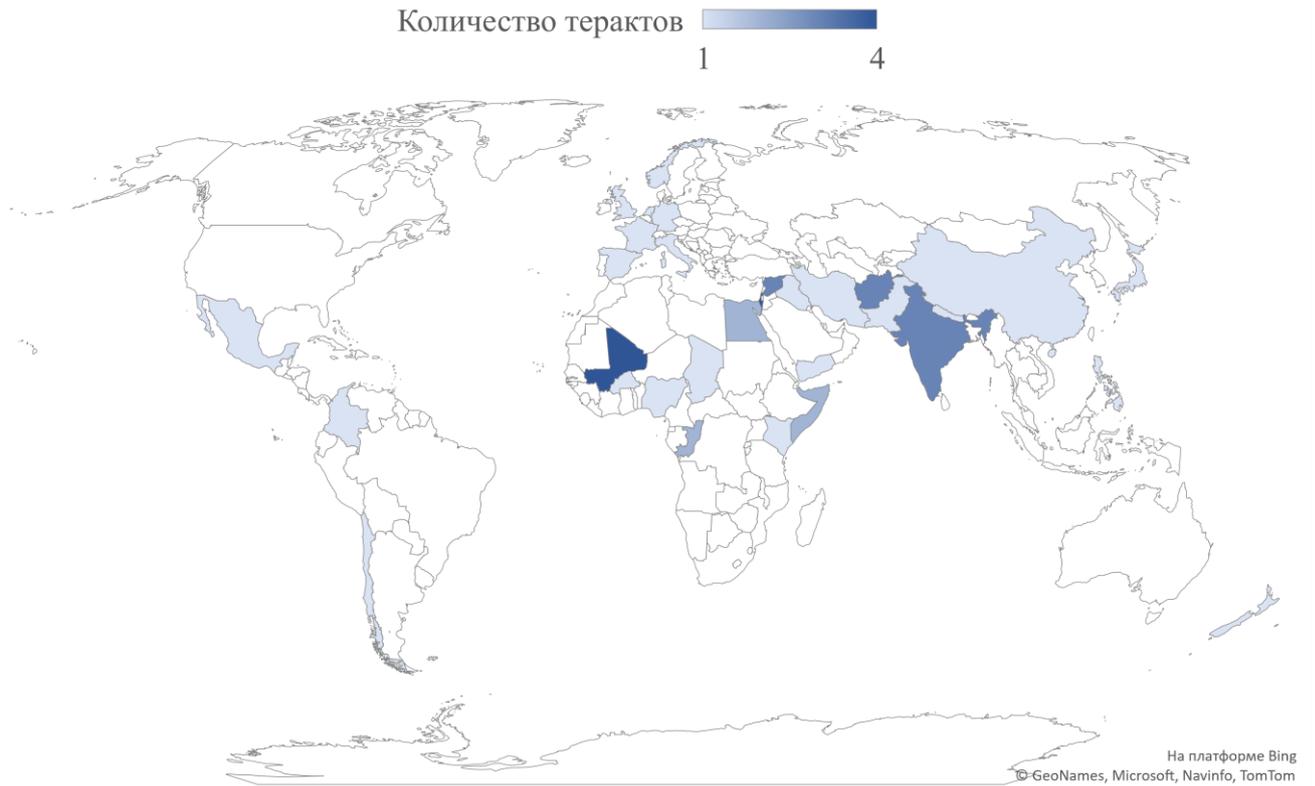


Рисунок 3.8. Количество терактов в мире в I квартале 2019 г.

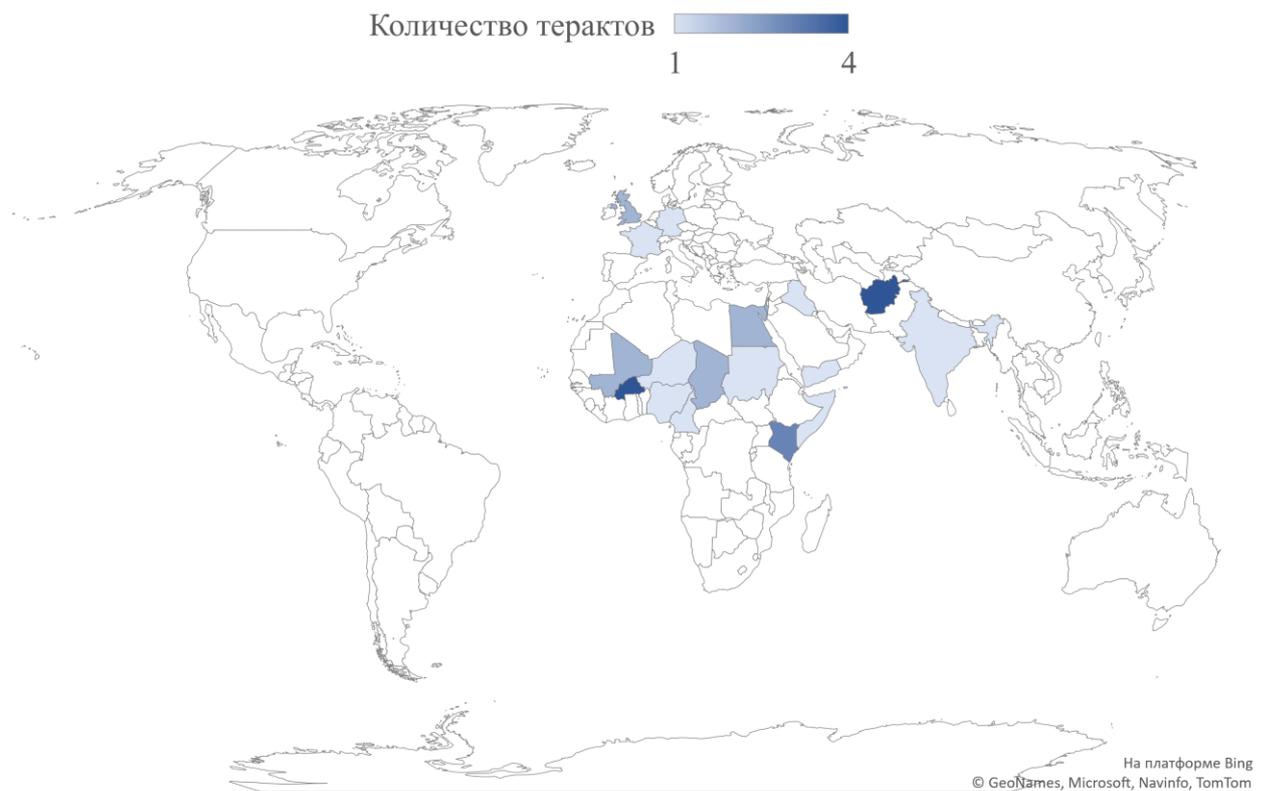


Рисунок 3.9. Количество терактов в мире в I квартале 2020 г.

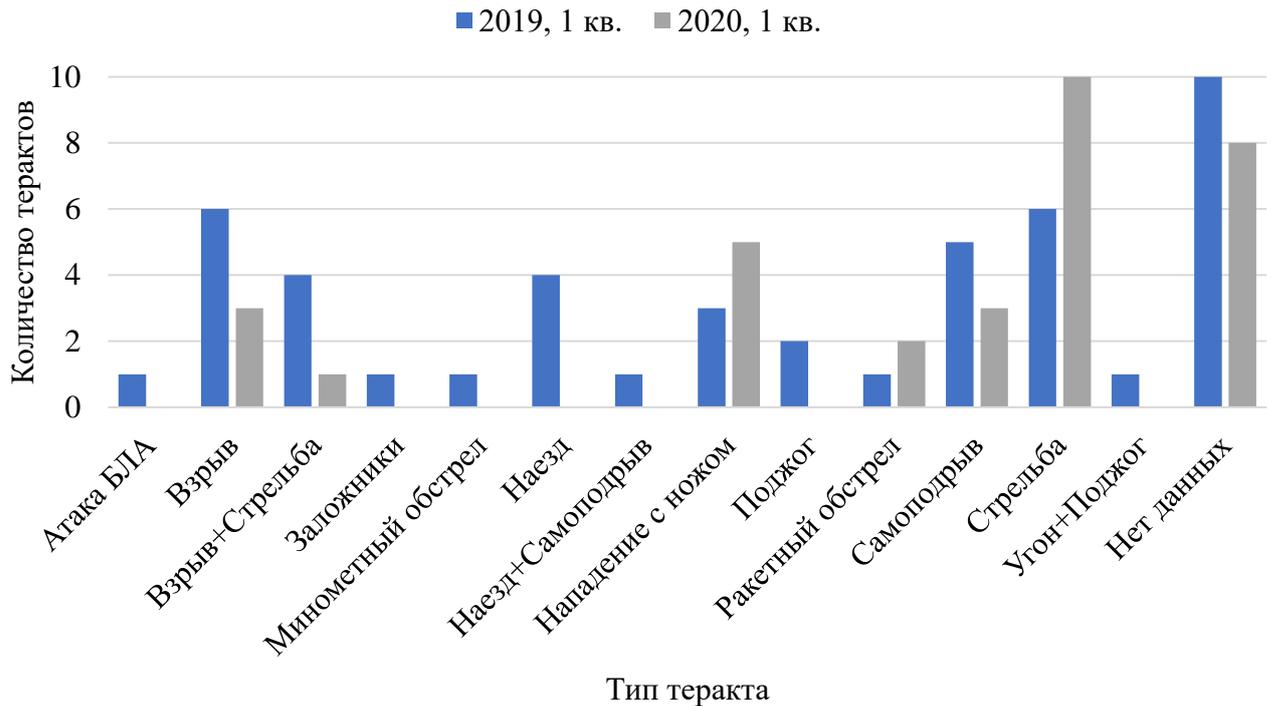


Рисунок 3.10. Сравнение частоты типов терактов в I квартале 2019 и 2020 гг.

3.5.2 Пример второй: анализ репрезентации террористов на материале русскоязычного корпуса

Этап 1. Постановка задачи интеллектуального контент-анализа

Изучить репрезентацию террористов в текстах русскоязычных новостных сообщений (одиночка или участник группы лиц, принадлежность к террористической организации, происхождение, идеология, пол и возраст).

Этап 2. Отбор материала

В качестве материала с учетом критериев отбора было взято 25 сообщений русскоязычных СМИ о терактах за январь-февраль 2020 г.

Этап 3. Онтологический анализ

Ниже приведен один из текстов корпуса на выходе процедуры онтологического анализа, включающего концептуальную разметку текста и разрешение концептуальной неоднозначности.

{Четыре бомбы}~Cd {были взорваны}~Td {в Индии}~Lc-78-N-78
{на праздновании Дня Республики}~Vc

{Взрывы}~Td {четырёх бомб}~Cd произошли {в Индии}~Lc-78-N-78
{во время праздничных мероприятий}~Vc, посвященных {Дню Респуб-
лики}~Vc, {сообщает}~D {26 января}~Vb {телеканал NDTV}~S.

{Решительно осуждаю}~Y {взрывы}~Td {в нескольких местах}~Le
{в штате Ассам}~Ld-35. Это {трусливая попытка}~K {создать хаос}~X
{в священный день}~Vc лишь демонстрирует разочарование {террористиче-
ских групп}~U после того, как народ полностью отверг их. {Наше}~DEF
{правительство}~Ra {примет}~Rb {самые суровые меры}~Ra, чтобы {при-
влечь}~Rb {виновных}~Ab к {ответственности}~CR, {написал}~D {главный
министр штата Ассам Сарбананда Соновал}~Ld-35~S в {Twitter}~S.

На {места взрывов}~Td~Le прибыли {сотрудники правоохранительных
органов}~Ra, по факту {произошедшего}~T начато {расследование}~Rb.

Как {сообщили}~D {полицейские}~S, {все}~DEF {четыре взрыва}~Td
произошли {в течение 10 минут}~B между {8}~Num :{15}~Num и {8}~Num
:{25}~Num ({5}~Num :{45}~Num и {5}~Num :{55}~Num {мск}~B). {Камеры
видеонаблюдения}~Rc зафиксировали, как {двое молодых людей}~Ab~age
~gen на мотоцикле бросают {гранату}~Cd и {скрываются}~Pbe. {Источ-
ник}~S в {администрации}~S Соновала {сообщил}~D, что {министр}~S об-
ратился с просьбой к {главе полиции штата}~Ra Бхаскару Джиоти Маханту
как можно скорее {привлечь}~Rb {виновных}~Ab к {ответственности}~CR.

{Правоохранительные органы}~S {полагают}~I, что {нападение}~T со-
вершено {представителями террористической организации}~Ab~U {Объеди-
ненный фронт освобождения Ассама}~U-15~I ({ULFA}~U-15~I), {запрещен-
ной}~Rb {в Индии}~Lc-78-N-78.

Этап 4. Определение релевантных концептуальных тегов

В соответствии с поставленной задачей релевантными считаются теги ветвей
TERRORISM-AGENT (A), TERRORIST-ORGANIZATION (U), NATION (N), IDEOLOGY (W),
а также теги атрибутов AGE (age) и GENDER (gen).

В приведенном выше тексте отсутствуют лексические единицы, эксплицитно
выражающие концепты ветви IDEOLOGY. Однако на основе одного из правил логи-
ческого вывода идеология террориста может быть определена через идеологию

террористической организации, членом которой он является. В тексте явным образом не указывается, какая именно организация стоит за терактом, однако правоохранительные органы высказывают предположение:

{Правоохранительные органы}~S **{полагают}**~I, что {нападение}~T совершено {представителями террористической организации}~Ab~U **{Объединенный фронт освобождения Ассама}**~U-15~I (**{ULFA}**~U-15~I), {запрещенной}~Rb {в Индии}~Lc-78-N-78.

Таким образом, мы получаем предположение (лексема отмечена тегом I атрибута ASSUMPTION), что искомой организацией является «Объединенный фронт освобождения Ассама», который в базе экземпляров фигурирует под условным наименованием TERRORIST-ORGANIZATION-15. Фрейм экземпляра TERRORIST-ORGANIZATION-15 приведен в таблице 3.18.

Таблица 3.18. Фрейм экземпляра TERRORIST-ORGANIZATION-15

LABEL	TERRORIST-ORGANIZATION-15	VAL
DEFINITION	«Объединенный фронт освобождения Ассама» — сепаратистская организация, запрещенная в Индии, требующая независимости штата Ассам	VAL
INSTANCE-OF	TERRORIST-ORGANIZATION	VAL
HAS-IDEOLOGY	SEPARATISM	VAL
DEFAULT-NAME-RU	ULFA	VAL
DEFAULT-NAME-EN	United Liberation Front of Assam	VAL
DEFAULT-NAME-FR	L'United Liberation Front of Asom	VAL

Поскольку слот HAS-IDEOLOGY экземпляра TERRORIST-ORGANIZATION-15 заполнен концептом SEPARATISM, соответствующий слот концепта TERRORIST-GROUP также может быть заполнен этим концептом на основе приведенного в п. 3.2.1 правила: «Если исполнитель является членом террористической организации, и такая террористическая организация проповедует идеологию, то этот исполнитель также проповедуют эту идеологию». При этом, учитывая, что данные об участии в теракте организации «Объединенный фронт освобождения Ассама» предположительны, данные об идеологии также должны быть предположительными.

В тексте также не обнаружено лексических единиц, репрезентующих концепт NATION в связи с концептами ветви TERRORISM-AGENT. В данном случае их невозможно получить посредством логического вывода, так как в базе знаний нет правил, подходящих для этой ситуации. Кроме того, при попытке проанализировать текст вручную становится понятно, что информация о национальности нападавших в нем в принципе отсутствует.

Этап 5. Заполнение фреймов

Полученная на предыдущем этапе релевантная информация была извлечена из анализируемых текстов, после чего на ее основе были сформированы концептуально-лексические фреймы. Ниже показан фрейм, сформированный на основании приведенного выше текста; символом «*» выделена информация, полученная на основе логического вывода, вопросительный знак перед концептом означает, что эта информация является предположительной.

Идентификатор текста:	Ассам_26.01.txt
TERRORISM-AGENT {4}	—
INDIVIDUAL-TERRORIST {0}	—
TERRORIST-GROUP {4}	двое молодых людей {1}, виновных {2}, представителями террористической организации {1}
TERRORIST-ORGANIZATION {3}	представителями террористической организации {1}
?TERRORIST-ORGANIZATION-15 {2}	Объединенный фронт освобождения Ассама {1}, ULFA {1}
NATION {0}	—
*IDEOLOGY {2}	—
?*SEPARATISM {2}	—
AGE {1}	двое молодых людей {1}
GENDER {1}	двое молодых людей {1}

Этап 6. Представление результатов интеллектуального контент-анализа

Результаты контент-анализа представлены графически средствами MS Excel. На рис. 3.11–3.13 приведена репрезентация состава, происхождения, идеологии террористов, на рис. 3.14 — репрезентация террористических организаций в тестовом русском корпусе за январь-февраль 2020 г.

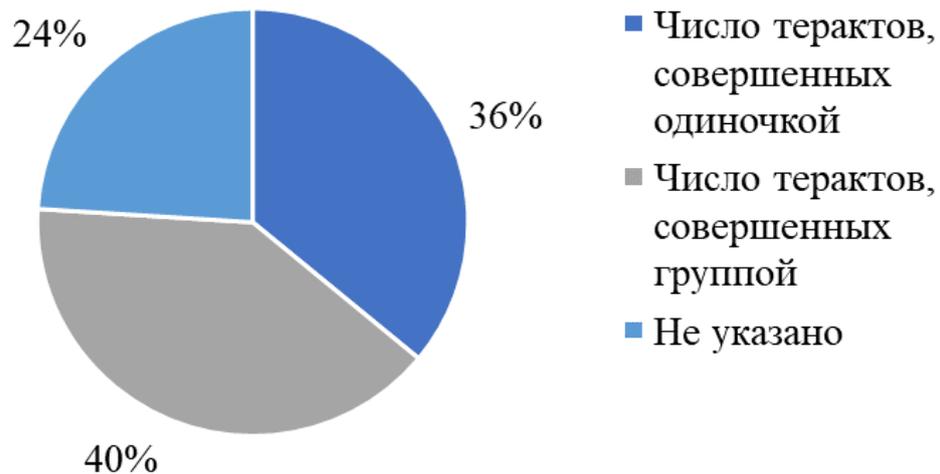


Рисунок 3.11. Репрезентация состава террористов в тестовом русском корпусе за январь-февраль 2020 г.

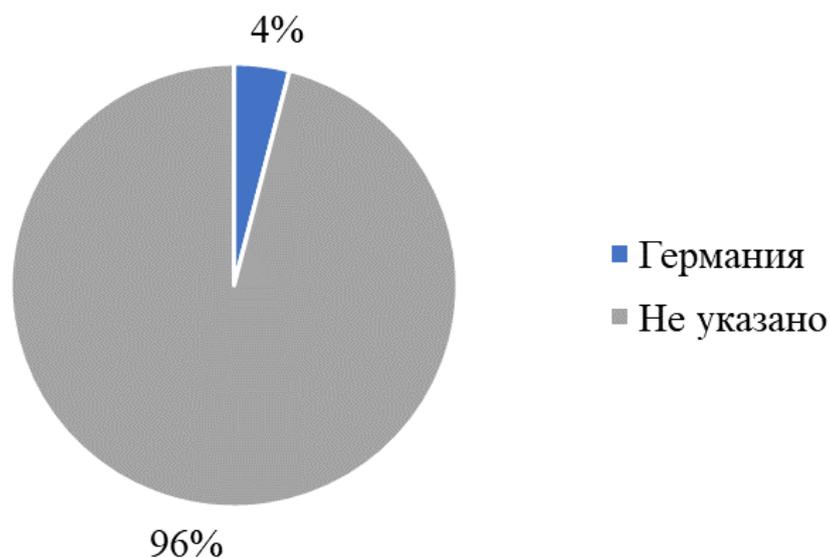


Рисунок 3.12. Репрезентация происхождения террористов в тестовом русском корпусе за январь-февраль 2020 г.

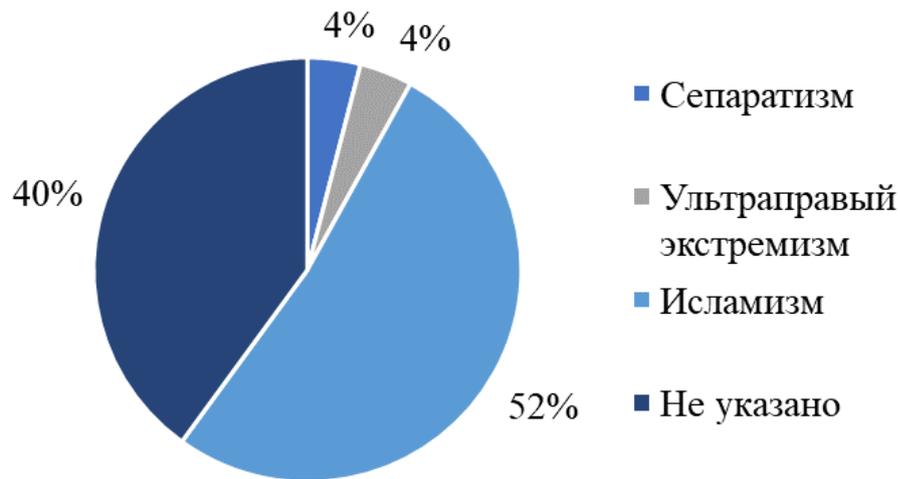


Рисунок 3.13. Репрезентация идеологии террористов в тестовом русском корпусе за январь-февраль 2020 г.

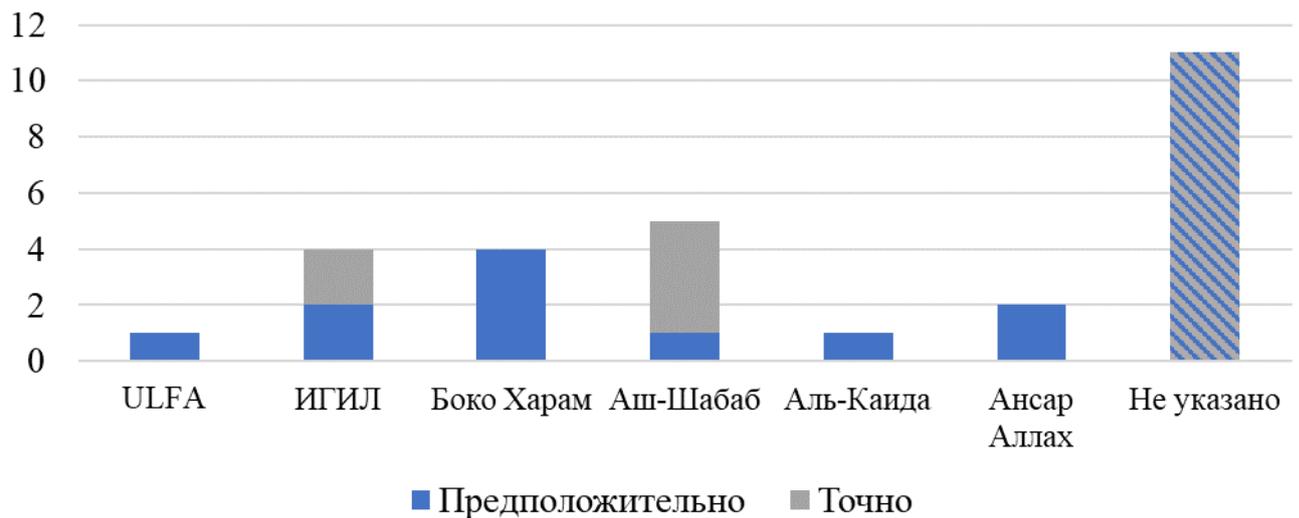


Рисунок 3.14 Репрезентация террористических организаций в тестовом русском корпусе за январь-февраль 2020 г.

3.5.3 Пример третий: репрезентация отдельного теракта в русско-, англо- и франкоязычных СМИ

Этап 1. Постановка задачи интеллектуального контент-анализа

Изучение концептуальной репрезентации в русско-, англо- и франкоязычных СМИ одного террористического акта с целью выявления различий межъязыковой и (или) межкультурной репрезентации; определение наиболее релевантных для анализируемых текстов концептов онтологии.

Этап 2. Отбор материала

В качестве материала интеллектуального контент-анализа выбраны тексты о террористическом акте в Крайстчерче, Новая Зеландия, 15 марта 2019 г. Отобрано по одному новостному сообщению, опубликованному в день теракта, на русском (980 слов), английском (1577 слов) и французском (477 слов) языках.

Этап 3. Онтологический анализ

Ниже приведены фрагменты отобранных текстов на выходе процедуры онтологического анализа, включающего концептуальную разметку текста и разрешение концептуальной неоднозначности.

Почему стала возможна {атака}~T {в Новой Зеландии}~Lc-124~N-124

Это отражение общего тренда: в мире растет популярность {идеологии белого превосходства}~Wd

{Крупнейший в истории}~M {Новой Зеландии}~Lc-124~N-124 {массовый расстрел}~Taa~M {в двух мечетях}~Zbg~Le {унес жизни}~Paa {49}~Num {человек}~Ha. {Эксперты}~S {отмечают}~D {рост правого экстремизма}~Wb {в стране}~N и {возможность}~I {ответных атак}~T со стороны {ислаμισлов}~A~Wa

Что произошло

{В пятницу}~B, {15 марта}~Bb, {злоумышленники}~Ab {совершили нападение}~T на {две мечети}~Zbg~Le {в Крайстчёрче}~La-157 {третьем по величине городе Новой Зеландии}~La~Lc-124~N-124 ({расположен}~L {на востоке Южного острова}~Ld, {наибольшем из двух островов страны}~Lc~Ld).

{49}~Num {killed}~Paa in {mass shooting}~Taa~M {at two mosques}~Zbg~Le {in Christchurch}~La-157, {New Zealand}~Lc-124~N-124

At least {49}~Num {people}~Ha {were killed}~Paa and {20}~Num {seriously injured}~Pab in {mass shootings}~Taa~M {at two mosques}~Zbg~Le {in the New Zealand city of Christchurch}~Lc-124~N-124-La-157 {Friday}~B, in a {carefully planned}~K and {unprecedented atrocity}~T that {shocked}~Y {the}~DEF {usually peaceful nation}~N.

{The}~DEF {attack}~T was unleashed {at lunchtime}~B {local time}~B {Friday}~B, when {mosques}~Zbg~Le were full of {worshippers}~Zae. {Footage}~S of {the}~DEF {massacre}~T {was streamed live online}~E, and a {rambling manifesto}~E laced with {white supremacist references}~Wd {was published}~D {just before}~B {the}~DEF {shootings}~Taa {unfolded}~T.

{Nouvelle-Zélande}~Lc-124~N-124: au moins {49}~Num {morts}~Paa dans une {attaque terroriste}~T de deux {mosquées}~Zbg~Le de {Christchurch}~La-157

Un {bilan provisoire}~S {fait également état}~D d'une vingtaine de {blessés}~Ha~Paa. Quatre {personnes}~A~Hb {ont été interpellées}~Pbd~Rb et l'une d'entre elles {inculpée de meurtres}~Pbd~Rb. {La}~DEF {piste du terrorisme d'extrême droite}~Wb~I est avancée. {La}~DEF {Nouvelle-Zélande}~Lc-124~N-124 vient de connaître un {acte de violence extraordinaire et sans précédent}~T, {selon}~D {la}~DEF {première ministre}~S, {Jacinda Arden}~S.

Этап 4. Определение релевантных концептуальных тегов

С учетом поставленной задачи контент-анализа релевантными признаны теги ветвей концептов, доля которых составляет не менее 1 % от общего количества тегов в тексте. В данном случае нет необходимости выявления имплицитно выраженных концептов, поэтому процедура логического вывода не задействована.

Этап 5. Заполнение фреймов

Выявленная на предыдущем этапе релевантная информация была извлечена из анализируемых текстов и была представлена в виде концептуально-лексических фреймов, фрагменты которых приводятся далее.

Идентификатор текста:	Christchurch_ru.txt
SOURCE {56}	Таппант {4}, AP {2}, Sky News {1} ...
DECLARE {46}	пишет {3}, заявил {2}, отмечают {2} ...
LOCATION {38}	Новой Зеландии {4}, в Крайстчёрче {4} ...

NATION {27}	австралийские власти {1} ...
TERROR-ATTACK {26}	нападение {4}, теракт {3}, массовый расстрел {1}, открыли огонь {1} ...
TERRORISM-AGENT {19}	Таррант {2}, подозреваемого {1}, жестокого праворадикального экстремиста {1} ...
TERRORIST-THREAT {14}	манифест {2}, фотографии оружия {1} ...
PROPERTY-TARGET {13}	Аль-Нур {2}, в двух мечетях {1} ...
IDEOLOGY {13}	идеологии белого превосходства {1}, радикальных исламистов {1} ...
CONSEQUENCES {12}	был убит {1}, были расстреляны {1} ...
TIME {10}	15 мар 2019 {1}, 15 марта {1} ...
Идентификатор текста:	Christchurch_en.txt
SOURCE {71}	statement {4}, Twitter {3}, witness {1} ...
DECLARE {62}	said {30}, tweeted {2}, broadcast live {2} ...
TERROR-ATTACK {56}	attack {12}, horrible massacre {2} ...
CONSEQUENCES {48}	seriously injured {3}, were arrested {2}, straight in the chest {1} ...
LOCATION {36}	in Christchurch {3}, New Zealand {2} ...
NATION {30}	people of New Zealand {3}, country {2}...
COMMUNITY-REACTION {30}	thoughts {2}, warmest sympathy {2} ...
TIME {26}	local time {3}, just before {2} ...
COUNTERTERRORISM-EVENT {25}	investigation {1}, was placed on lockdown {1}, were deployed {1}, were found {1} ...
TERRORISM-AGENT {23}	gunman {7}, suspects {2}, shooter {1} ...
PROPERTY-TARGET {19}	mosque {3}, in the mosque {3} ...
Идентификатор текста:	Christchurch_fr.txt
CONSEQUENCES {25}	blessés {2}, morts {2}, ont été interpellées {2}, corps ensanglantés {1} ...

DECLARE {22}	a annonce {2}, a déclaré {1}, selon {1} ...
SOURCE {20}	AFP {1}, première ministre {1} ...
TERRORISM-AGENT {18}	homme {3}, personnes {2}, tireur {2}, homme blanc {1}, l'auteur {1} ...
LOCATION {14}	Nouvelle-Zélande {4}, dans le centre de Christchurch {1}, Christchurch {1} ...
NATION {13}	pays {4}, premier ministre français {1} ...
TERROR-ATTACK {11}	attaque terroriste {2}, en train de tirer {1} ...
COUNTERTERRORISM-EVENT {6}	ont été interpellées {2}, l'arrestation {1} ...
PROPERTY-TARGET {5}	mosquées {2}, dans une mosquée {2}, à l'intérieur de la mosquée Al Noor {1}

Этап 6. Представление результатов интеллектуального контент-анализа

Результаты интеллектуального контент-анализа представлены графически средствами MS Excel. На рис. 3.15 приведено распределение наиболее частотных концептов верхнего уровня онтологии терроризма в ранних текстах новостей о теракте в Крайстчерче. Доля каждого из включенных в диаграмму концептов по крайней мере в одном из проанализированных текстов составляет больше 1 %.

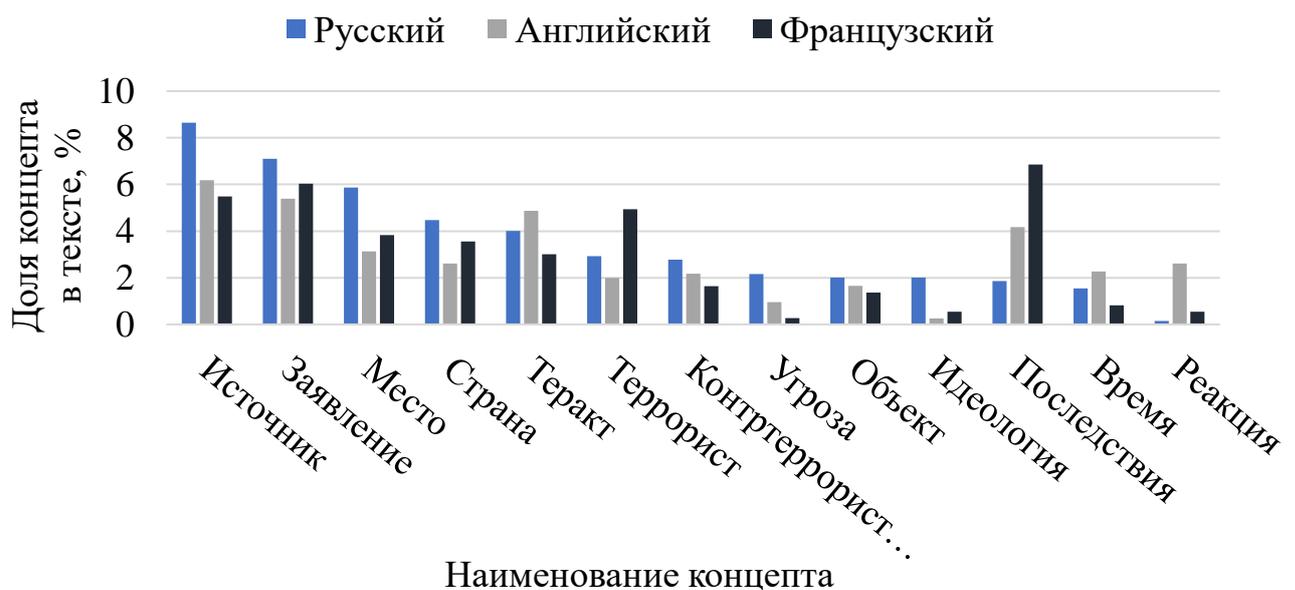


Рисунок 3.15. Концептуальная репрезентация теракта в Крайстчерче в русскоязычных, англоязычных и франкоязычных СМИ

На рис. 3.15 прослеживаются существенные различия между концептуальной репрезентацией теракта в новостных сообщениях на русском, английском и французском языках. Например, в тексте на русском языке превалируют концепты «Источник» (SOURCE) и «Заявление» (DECLARE). Это свидетельствует о том, что в тексте приводятся разнообразные мнения властей и СМИ о произошедшем, при этом почти не упоминается реакция общественности. Кроме того, в отличие от других текстов, информация о последствиях теракта в русском тексте минимальна. В тексте на английском языке более высокую частоту, нежели в других текстах, имеют концепты «Теракт» (TERROR-ATTACK) и «Реакция общественности» (COMMUNITY-REACTION), однако почти не упоминается идеология. В тексте на французском языке больше всего внимания отводится описанию последствий теракта и сведениям о террористе, меньше всего — сведениям о террористической угрозе и реакции общественности. Выявленные различия, вероятно, вызваны культурными особенностями носителей данных языков; кроме того, влияние может оказывать направленность и позиция изданий, в которых были опубликованы эти новости.

ВЫВОДЫ ПО ГЛАВЕ 3

В настоящей главе изложены методологические основы разработки модели многоязычного интеллектуального контент-анализа неструктурированной информации на примере новостных сообщений предметной области «Терроризм», описан ход разработки модели и ее результаты. Кроме того, описан вспомогательный инструмент автоматизации онтологического анализа, настройка которого на обработку французского языка была выполнена автором диссертационного исследования. Работа модели многоязычного интеллектуального контент-анализа продемонстрирована на примере контент-анализа новостных сообщений о террористической деятельности на русском, английском и французском языках.

Созданная модель многоязычного интеллектуального контент-анализа новостных сообщений предметной области «Терроризм» состоит из двух основных компонентов: базы знаний и алгоритма.

База знаний содержит следующие компоненты:

- 1) трехуровневую многоязычную лингвистическую онтологию терроризма, представленную в формализме онтологии верхнего уровня μK и содержащую 113 концептов, обозначающих объекты и события, 28 отношений, 5 атрибутов и 32 аксиомы, а также связанную с онтологией базу экземпляров отдельных концептов;
- 2) перекрестные онтолексиконы русского, английского и французского языков, каждый из которых содержит более 20 000 лексических единиц длиной от 1 до 10 компонентов, а также ономастиконы для каждого из языков, содержащие имена собственные и связанные с экземплярами онтологии;
- 3) правила онтологического анализа, включающие правила концептуальной разметки и правила снятия концептуальной неоднозначности;
- 4) правила логического вывода, соответствующие аксиомам онтологии;
- 5) правила создания динамических концептуально-лексических фреймов.

Алгоритм модели многоязычного интеллектуального контент-анализа включает в себя этап постановки задачи интеллектуального контент-анализа и процедуры решения этой задачи, выполняемые либо автоматически с использованием авторских программ, либо вручную: отбор текстового материала, онтологический анализ, определение релевантных концептуальных тегов, заполнение концептуально-лексических фреймов, представление результатов.

В последнем разделе поэтапно описаны три примера использования модели многоязычного интеллектуального контент-анализа для анализа новостных сообщений о террористической деятельности на русском, английском и французском языках. Примеры демонстрируют возможность успешного извлечения из текстов релевантного для информационного запроса пользователя контента, в том числе выраженного имплицитно, и представления его в графической форме.

Модель интеллектуального контент-анализа, представленная в настоящей главе, основана на независимой от конкретного языка онтологии, что обеспечивает ее универсальный характер: для применения модели к обработке текстов на других языках, не являвшихся языками данного исследования, в рамках предметной области «Терроризм» требуется внесение изменений только в лингвистический

компонент базы знаний, в то время как концептуальный и эпизодический компоненты, а также правила онтологического анализа, логического вывода и формирования фреймов для представления результатов могут быть использованы повторно. Кроме того, описанные принципы разработки базы знаний могут быть применены для построения базы знаний любой другой предметной области. Следовательно, в рамках настоящего исследования нам удалось достигнуть поставленной цели.

ЗАКЛЮЧЕНИЕ

Извлечение из неструктурированных текстовых источников релевантной запросу пользователя информации является одним из широко разрабатываемых направлений автоматической обработки текста. Выполненное исследование, с одной стороны, опирается на существующие разработки по данному направлению, и, с другой стороны, призвано дополнить их новыми теоретическими и практическими положениями в области многоязычного интеллектуального контент-анализа.

В ходе исследования решены поставленные ранее задачи теоретического и прикладного характера. В частности, определены теоретико-методологические основы моделирования многоязычного интеллектуального контент-анализа.

Выявлено, что под многоязычным интеллектуальным контент-анализом понимается качественно-количественный компьютеризированный метод исследования текстов, предполагающий извлечение из неструктурированных разноязычных текстов ограниченной предметной области контента, соответствующего информационному запросу пользователя, его числовая обработка, интерпретация и представление в удобной для пользователя форме. В свою очередь контент, являясь главным понятием контент-анализа, представляет собой то, что содержится в анализируемом тексте в явном виде или может быть логически выведено на основе эксплицитной текстовой информации и фоновых знаний о предметной области.

Привлечение к контент-анализу фоновых знаний о предметной области и мире в целом возможно посредством использования онтологий, которые позволяют осуществлять логический вывод на основе содержащихся в них аксиом. Установлено, что наиболее подходящим типом онтологии для модели многоязычного интеллектуального контент-анализа является многоязычная, но при этом независимая от конкретного естественного языка лингвистическая онтология предметной области, разработанная на основе корпусного анализа текстов предметной области.

Предложена итеративная методика построения модели многоязычного интеллектуального контент-анализа в ограниченной предметной области, которая предусматривает последовательное осуществление трех стадий разработки

с применением ручных и автоматизированных методов: анализ лингвистического материала предметной области, разработка исходной модели многоязычного интеллектуального контент-анализа и совершенствование разработанной модели.

С целью анализа лингвистического материала новостных сообщений предметной области «Терроризм» собраны три псевдопараллельных корпуса текстов на русском, английском и французском языках за 2014–2020 гг., объем каждого из которых составил около двухсот тысяч словоупотреблений.

Проведен анализ подъязыка новостных сообщений предметной области «Терроризм» на материале русского, английского и французского языков, благодаря чему доказано, что данные новостные сообщения обладают характерной структурой релевантности и суперструктурой, демонстрируют некоторые закономерности на морфосинтаксическом уровне и содержат определенный набор концептуальных классов лексики, репрезентирующих релевантный контент предметной области. На основе полученных данных разработана база знаний модели многоязычного интеллектуального контент-анализа.

База знаний модели представлена следующими компонентами:

- трехуровневой многоязычной онтологией предметной области «Терроризм», описанной в формализме онтологии верхнего уровня MikroKosmos, причем данная онтология терроризма состоит из 113 концептов, обозначающих объекты и события, 28 отношений, пяти атрибутов и 32 аксиом;
- базой релевантных для предметной области экземпляров отдельных концептов; перекрестными предметно-релевантными связанными с концептами онтологии лексиконами русского, английского и французского языков объемом более 20 тыс. лексических единиц каждый;
- правилами онтологического анализа, включающими правила концептуальной разметки и правила снятия концептуальной неоднозначности;
- правилами логического вывода, основанными на аксиомах;
- правилами формирования динамических концептуально-лексических фреймов для числовой обработки и промежуточного представления результатов контент-анализа.

Разработан алгоритм интеллектуального контент-анализа, который включает в себя этап постановки задачи контент-анализа и этапы решения данной задачи, выполняемые вручную или автоматически с использованием авторских программ: отбор текстового материала, онтологический анализ, определение релевантных концептуальных тегов, в том числе с использованием правил логического вывода, формирование концептуально-лексических фреймов, представление результатов.

Совместно с коллективом НОЦ «Лингво-инновационные технологии» Южно-Уральского государственного университета разработана платформа концептуального аннотирования новостных сообщений предметной области «Терроризм»; в частности, автором настоящей работы данная платформа была настроена на концептуальное аннотирование новостных сообщений на французском языке.

Работа модели многоязычного интеллектуального контент-анализа продемонстрирована на примере трех задач различного характера, связанных с обработкой русского, английского и смешанного русско-англо-французского корпуса новостных сообщений предметной области «Терроризм». Показана возможность успешного извлечения из данных текстов релевантного для пользователя контента, в том числе не выраженного эксплицитно на текстовом уровне, и представления результатов в графической форме.

На основе разработанной модели возможно создание системы интеллектуального контент-анализа новостных сообщений предметной области «Терроризм», использование которой позволит в значительной степени автоматизировать и повысить оперативность аналитической деятельности в сфере политического прогнозирования и контртерроризма. Конкретные результаты анализа подязыка новостных сообщений предметной области «Терроризм» могут быть использованы для разработки других типов моделей и систем автоматической обработки текста, например систем автоматического реферирования и аннотирования, информационно-поисковых систем, систем машинного перевода. Отдельные положения исследования могут быть включены в курсы по прикладной лингвистике, функциональной стилистике, общественно-политическому переводу.

Перспективы исследования заключаются, во-первых, в возможности дальнейшего расширения и углубления базы знаний предметной области «Терроризм» и увеличения связей между концептами, что позволит проводить более глубокий анализ новостных сообщений предметной области и, таким образом, сможет положительно повлиять на качество аналитики и прогнозирования в области контртеррористической деятельности. Кроме того, в разработанную базу знаний могут быть добавлены дополнительные языки для расширения области применения модели интеллектуального контент-анализа.

Во-вторых, более детального изучения требует проблема концептуальной неоднозначности, возникающая на этапе онтологического анализа. В будущих исследованиях планируется провести апробацию методов снятия концептуальной неоднозначности на более объемных корпусах текстов с целью получения более достоверных данных, а также рассмотреть другие количественные и качественные методы устранения неоднозначности.

Наконец, принципы, лежащие в основе предложенной методики разработки модели многоязычного интеллектуального контент-анализа, могут быть экстраполированы на другие языки и предметные области, что позволит выполнять разработку аналогичных моделей интеллектуального контент-анализа с экономией времени и исследовательских усилий.

СПИСОК ЛИТЕРАТУРЫ

1. Андреев, Н. Д. Статистико-комбинаторные методы в теоретическом и прикладном языкознании / Н. Д. Андреев. — Л.: Наука, 1967. — 403 с.
2. Апресян, Ю. Д. Идеи и методы современной структурной лингвистики (краткий очерк) / Ю. Д. Апресян. — М.: Просвещение, 1966. — 304 с.
3. Баранов, А. В. Опыт текстового анализа газеты / А. В. Баранов // Количественные методы в социальных исследованиях. Информ. бюл. ИКСИ АН СССР. — М., 1968. — № 9.
4. Баранов, А. Н. Введение в прикладную лингвистику: учебное пособие / А. Н. Баранов. — 2-е изд., испр. — М.: Едиториал УРСС, 2003. — 360 с.
5. Богуславский, И. М. Онтология для поддержки задач извлечения смысла из текста на естественном языке / И. М. Богуславский, В. Г. Диконов, С. П. Тимошенко // Информационные технологии и системы. — 2012. — С. 152–161.
6. Большой энциклопедический словарь / гл. ред. А. М. Прохоров. — 2-е изд., перераб. и доп. — СПб.: Норинт, 2004. — 1456 с.
7. Бызов, А. А. Интеллектуальный анализ текстов в социальных науках / А. А. Бызов // Социология: 4М. — 2019. — № 49. — С. 131–160.
8. Власов, Д. Ю. Автоматизация извлечения отношений между понятиями из текстов естественного языка / Д. Ю. Власов, Д. Е. Пальчунов, П. А. Степанов // Вестник НГУ. Сер.: Информ. технологии. — № 8 (3). — 2010. — С. 23–33.
9. Власова, М. Л. Социологические методы в маркетинговых исследованиях / М. Л. Власова. — М.: Издательский дом ГУ ВШЭ, 2005. — 712 с.
10. Воронов, Ю. П. Чтение между строк (контент-анализ в конкурентной разведке) [Электронный ресурс] / Ю. П. Воронов // ЭКО. — № 11. — 2005. — Режим доступа: <http://www.corpus-consulting.ru/Categorys/94>, свободный. — Загл. с экрана (дата обращения: 11.11.2019).
11. Данченков, С. И. Классификация текстов в системе узлов лексической онтологии / С. И. Данченков, В. Н. Поляков // Учен. зап. Казан. ун-та. Сер. Физ.-матем. науки. — 2010. — Т. 152 (1). — С. 255–267.

12. Добров, А. В. Автоматическая рубрикация новостных сообщений средствами синтаксического семантики : дис. ... канд. филол. наук : 10.02.21 / Добров Алексей Владимирович. — Санкт-Петербург, 2014. — 418 с.

13. Добров, А. В. [и др.] Семантический анализ новостных сообщений по теме «Электронные услуги»: опыт применения методов онтологической семантики / А. В. Добров, А. Е. Доброва, Н. Л. Сомс, А. В. Чугунов // Труды XVIII объединенной конференции «Интернет и современное общество» (IMS-2015). — СПб: Библ. РАН, 2015. — С. 120–125.

14. Доброва, В. В. Моделирование как метод познания ненаблюдаемых объектов / В. В. Доброва. — Вестник Вятского государственного университета. — 2015. — № 1. — С. 146–152.

15. Добросклонская, Т. Г. Вопросы изучения медиатекстов: Опыт исследования современной английской медиаречи / Т. Г. Добросклонская. — М.: УРСС Эдиториал, 2008. — 288 с.

16. Ефремова Т. Ф. Новый словарь русского языка. Толково-словообразовательный. — М.: Русский язык, 2001. — 2354 с.

17. Жданова, С. Ю. [и др.] Концепт терроризм в дискурсе международных новостей интернет-ресурсов / С. Ю. Жданова, С. Л. Мишланова, В. Б. Поляков, А. А. Строканов // Вектор науки Тольяттинского государственного университета. Серия: Педагогика, Психология. — 2012. — № 4 (11). — С. 100–103.

18. Жеребило, Т. В. Словарь лингвистических терминов / Т. В. Жеребило. — Назрань: Пилигрим, 2010. — Изд. 5-е, испр. и доп. — 486 с.

19. Зализняк, А. А. Феномен многозначности и способы его описания / А. А. Зализняк // Вопросы языкознания. — М.: Наука, 2004. — № 2. — С. 20–45.

20. Захаров, В. П. Прологомены к корпусной лингвистике / В. П. Захаров // Вопросы психолингвистики. — 2016. — № 2 (28). — С. 150–161.

21. Звегинцев, В. А. Теоретическая и прикладная лингвистика / В. А. Звегинцев. — М.: Просвещение, 1968. — 338 с.

22. Зиновьева, А. Ю. Онтологические модели и перевод (на примере предметной области «Терроризм»): выпускная квалификационная работа / А. Ю. Зиновьева. — Южно-Уральский гос. ун-т, Челябинск, 2018. — 165 с.

23. Иорданская, Л. Н. Автоматический синтаксический анализ / Л. Н. Иорданская. — Новосибирск : Наука Сиб. отд-ние, 1967. — Т. 1. — 231 с.

24. Коваль, С. А. Безэкземплярные и экземплярные онтологии [Электронный ресурс] / С. А. Коваль // Материалы XXXVI Международной филологической конференции 12 марта 2007 г. в Санкт-Петербургском университете. — 2007. — Режим доступа: <http://skowal.ru/old/research/ontology2007.htm>, свободный. — Загл. с экрана (дата обращения: 3.06.2017).

25. Козеренко, Е. Б. Концептуально-лингвистическое моделирование в интеллектуальных системах на основе расширенных семантических сетей : автореф. дис. ... 05.13.17 / Елена Борисовна Козеренко. — Москва, 1995. — 25 с.

26. Колесниченко, А. В. Практическая журналистика : учеб. пособие / А. В. Колесниченко. — М. : Изд-во Моск. ун-та, 2008. — 192 с.

27. Комарова, З. И. Методология, метод, методика и технология научных исследований в лингвистике : учебное пособие / З. И. Комарова. — М.: ФЛИНТА, 2014. — 820 с.

28. Кордонский, С. Прикладная герменевтика информационного пространства: картины мира, теоретические онтологии и верные матрицы / С. Кордонский, В. Бардин // ЛОГОС. — 2015. — Т.25, №3. — С. 19–45.

29. Кубрякова, Е. С. Краткий словарь когнитивных терминов / Е. С. Кубрякова, В. З. Демьянков, Ю. Г. Панкрац, Л. Г. Лузина. — М., 1996. — 245 с.

30. Кудрина, Л. В. Структура новостных текстов и их лексические особенности (на материале социальных сетей и интернет-СМИ): автореф. дис. ... канд-та филол. н.: 10.02.01 / Лада Владимировна Кудрина. — Санкт-Петербург, 2020. — 20 с.

31. Лазарев, С. А. Некоторые аспекты повышения эффективности механизмов контентной фильтрации трафика / С. А. Лазарев, А. В. Демидов, Р. В. Шатеев // Информационные ресурсы, системы и технологии. — 2013. — С. 1–8.

32. Лидерс, Е. В. Доходные дома: мировой опыт, практика и позиционирование в России / Е. В. Лидерс // Современная наука. — 2015. № 1. — С. 27–32.
33. Лингвистический энциклопедический словарь / гл. ред. В. Н. Ярцева. — М.: Советская энциклопедия, 1990. — 685 с.
34. Лосев, А. Ф. Введение в общую теорию языковых моделей / А. Ф. Лосев. — М.: Едиториал УРСС, 2004. — 296 с.
35. Лукашевич, Н. В. [и др.] Онтологические ресурсы и информационно-аналитическая система в предметной области «Безопасность» / Н. В. Лукашевич, Б. В. Добров, А. М. Павлов, С. В. Штернов // Онтология проектирования. — 2018. — Т. 8, № 1 (27). — С. 74–95.
36. Лукашевич, Н. В. Тезаурусы в задачах информационного поиска / Н. В. Лукашевич. — М.: МГУ, 2011. — 495 с.
37. Марков, А. А. Пример статистического исследования над текстом Евгения Онегина, иллюстрирующий связь испытаний в цепи / А. А. Марков // Известия Императорской Академии Наук. — 1913. — Т. 7 (3). — С. 153–162.
38. Мельчук, И. А. Опыт теории лингвистических моделей «Смысл–Текст» / И. А. Мельчук. — М.: Школа «Языки русской культуры», 1999. — 346 с.
39. Митрофанова, О. А. Онтологии как системы хранения знаний [Электронный ресурс] / О. А. Митрофанова, Н. С. Константинова. — 2015. — Режим доступа: <http://ict.edu.ru/ft/005706/68352e2-st08.pdf>, свободный. — Загл. с экрана (дата обращения: 31.10.2016).
40. Мишланова, С. Л. Особенности репрезентации концепта терроризм в англоязычном дискурсе новостей Интернет-ресурсов / С. Л. Мишланова, Е. А. Куприянычева // Филологические заметки. — 2012. — С. 265–276.
41. Нариньяни, А. С. Кентавр по имени ТЕОН: Тезаурус + Онтология / А. С. Нариньяни // Труды международного семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. — Т. 1. — Аксаково, 2001. — С. 184–188.
42. Национальный корпус русского языка [Электронный ресурс]. — 2003–2022. — Режим доступа: ruscorpora.ru, свободный. — Загл. с экрана (дата обращения: 07.01.2022).

43. Неуймин, Я. Г. Модели в науке и технике: история, теория и практика / Я. Г. Неуймин. — Л.: Наука, 1984. — 187 с.
44. Олейник, А. Н. Контент-анализ больших качественных данных / А. Н. Олейник // *International Journal of Open Information Technologies*. — 2019. — Т. 7, № 10. — С. 36–49.
45. Пастухов, А. Г. Содержательный анализ медиатекста: новое в определении метода / А. Г. Пастухов // *Вестник ВГУ. Серия: Лингвистика и межкультурная коммуникация*. — 2014. — № 3. — С. 130–133.
46. Пашимян, И. А. Контент-анализ как метод исследования: достоинства и ограничения / И. А. Пашимян // *Научная периодика: проблемы и решения*. — 2012. — № 3. — С. 13–18.
47. Поляков, В. Н. Использование технологий, ориентированных на лексическое значение, в задачах поиска и классификации / В. Н. Поляков // *Проблемы прикладной лингвистики. Сборник статей*. — 2004. — Вып. 2. — С. 101–117.
48. Рахилина, Е. В. [и др.] Многозначность как прикладная проблема: семантическая разметка в национальном корпусе русского языка / Е. В. Рахилина, Б. П. Кобрицов, Г. И. Кустова, О. Н. Ляшевская, О. Ю. Шеманаева // *Труды международной конференции «Диалог 2006»*. — 2006. — С. 445–450.
49. Ревзин, И. И. Модели языка / И. И. Ревзин. — М.: Изд-во Академии наук СССР, 1962. — 194 с.
50. Ревзин, И. И. Метод моделирования и типология славянских языков / И. И. Ревзин. — М.: Наука, 1967. — 298 с.
51. Ревзин, И. И. Современная структурная лингвистика / И. И. Ревзин. — М.: Наука, 1977. — 263 с.
52. Рубашкин, В. Ш. Онтологии — проблемы и решения. Точка зрения разработчика / В. Ш. Рубашкин // *Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2007»*. — М.: Изд-во РГГУ, 2007. — С. 481–485.

53. Рубашкин, В. Ш. Онтологическая семантика. Знания. Онтологии. Онтологически ориентированные методы информационного анализа текста / В. Ш. Рубашкин. — М.: ФИЗМАТЛИТ, 2013. — 348 с.

54. Сидорова, Е. А. Методы и программные средства для анализа документов на основе модели предметной области : автореф. дис. ... канд-та физ.-мат. наук : 05.13.11 / Елена Анатольевна Сидорова. — Новосибирск, 2006. — 20 с.

55. Смирнова, Л. Н. Профессионально дифференцированное обучение устной научной монологической речи научных работников на продвинутом этапе (на материале английского языка) : дис. ... канд-та пед. наук : 13.00.02 / Людмила Николаевна Смирнова. — Ленинград, 1984. — 252 с.

56. Соколова, Е. Г. Проблемы описания компьютерной лингвистики в виде онтологии для портала знаний / Е. Г. Соколова, И. С. Кононенко, Ю. А. Загорулько // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4–8 июня 2008 г.). — Вып. 7 (14). — М.: РГГУ, 2008. — С. 482–487.

57. Таршис, Е. Я. Перспективы развития метода контент-анализа / Е. Я. Таршис // Социология: 4М. — 2002. — № 15. — С. 71–92.

58. Толковый словарь русского языка в 4-х томах: Описание ЭНИ [Электронный ресурс] / под ред. Д. Н. Ушакова // Фундаментальная электронная библиотека «Русская литература и фольклор» (ФЭБ). — Режим доступа: <http://feb-web.ru/feb/ushakov/ush-abc/0ush.htm>, свободный. — Загл. с экрана (дата обращения: 26.03.2018).

59. Федеральная служба безопасности РФ. Единый федеральный список организаций, в том числе иностранных и международных организаций, признанных в соответствии с законодательством Российской Федерации террористическими (на 5 июля 2019 г.) [Электронный ресурс]. — 2019. — Режим доступа: <http://www.fsb.ru/fsb/npd/terror.htm>, свободный. — Загл. с экрана (дата обращения: 13.07.2020).

60. Химинец, Е. М. Особенности номинаций субъекта террористических действий во французской прессе (на материале газет «Le Figaro» и «Liberation») / Е. М. Химинец // ИСОМ. — 2016. — № 4–2. — С. 172–176.

61. Чернобровкина, Е. П. Контент-анализ в лингвистических исследованиях / Е. П. Чернобровкина // Вестник Бурятского государственного университета. Язык. Литература. Культура. — 2011. — № 11. С. 125–129.

62. Шалак, В. И. Математические методы компьютерного контент-анализа текстов / В. И. Шалак // Труды научно-исследовательского семинара логического центра ИФ РАН. Вып. XVI. / под ред. Карпенко А. С. — М.: 2002. — С. 117–127.

63. Шереметьева, С. О. Извлечение и формализация лингвистических знаний с помощью ЭВМ / С. О. Шереметьева, В. К. Сарьян // НТИ. Сер. 2. — 1994. — № 2. — С. 20–23.

64. Шереметьева, С. О. К вопросу о разработке онтологических ресурсов предметной области «Терроризм» / С. О. Шереметьева, А. Ю. Зиновьева // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. — 2017. — Т. 14, № 4. — С. 48–54.

65. Шереметьева, С. О. Платформа для концептуального аннотирования многоязычных текстов / С. О. Шереметьева, О. И. Бабина // Вестник ЮУрГУ. Серия «Лингвистика». — 2020. — Т. 17, № 4. — С. 53–60.

66. Ядов, В. А. Методология и процедуры социологических исследований / В. А. Ядов. — Тарту: Изд-во Тартуского госуниверситета, 1968. — 281 с.

67. Agić, Ž. Cross-lingual tagger evaluation without test data / Ž. Agić, B. Plank, A. Søgaard // Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. — Valencia, Spain, 2017. — Vol. 2. — P. 248–253.

68. Aguado de Cea, G. [и др.] Using linguistic patterns to enhance ontology development / G. Aguado de Cea, E. Montiel-Ponsoda, A. Gómez-Pérez, M. del Carmen Suárez-Figueroa // Proceedings of the International Conference on Knowledge Engineering and Ontology Development. — 2009. — P. 206–213.

69. Alatrish, E. S. Building ontologies for different natural languages / E. S. Alatrish, D. Tošić, N. Milenkovic // *Computer Science and Information Systems*. — 2014. — Vol. 11 (2). — P. 623–644.

70. Altheide, D. L. Ethnographic content analysis / D. L. Altheide // *Qualitative Sociology*. — 1987. — Vol. 10. — P. 65–77.

71. Anthony, L. AntConc (Version 3.5.9) [Электронный ресурс] / L. Anthony. — 2020. — Режим доступа: <https://www.laurenceanthony.net/software/antconc/releases/AntConc359/>, свободный. — Загл. с экрана (дата обращения: 05.01.2022).

72. Arp, R. Building ontologies with Basic Formal Ontology / R. Arp, B. Smith, A. D. Spear. — Cambridge, MA: MIT Press, 2015. — 248 p.

73. Astrakhantsev, N. A. Automatic Enrichment of Informal Ontology by Analyzing a Domain-Specific Text Collection / N. A. Astrakhantsev, D. G. Fedorenko, D. Y. Turdakov. — *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue"* (2014). — 2014. — Vol. 13. — P. 29–42.

74. Barhamgi, M. [и др.] Social networks data analysis with semantics: Application to the radicalization problem / M. Barhamgi, A. Masmoudi, R. Lara-Cabrera, D. Camacho // *Journal of Ambient Intelligence and Humanized Computing*. — 2018. — Режим доступа: <https://link.springer.com/article/10.1007/s12652-018-0968-z>, платный. — Загл. с экрана (дата обращения: 15.08.2021).

75. Benahmed, K. An ontology design for terrorism activities / K. Benahmed, N. Assouli // *International Journal of Management and Applied Science*. — 2017. — Vol. 3 (9). — P. 90–95.

76. Benamara, F. Cooperative question answering in restricted domains: the WEB-COOP experiment / F. Benamara // *Proceedings of the Conference on Question Answering in Restricted Domains*. — Association for Computational Linguistics, 2004. — P. 31–38.

77. Berelson, B. The analysis of communication content / B. Berelson, P. F. Lazarsfeld. — Chicago: The University of Chicago Press. — 1948. — 254 p.

78. Berelson, B. Content analysis in communications research. — NY: Free Press, 1952. — 220 p.

79. Berners-Lee, T. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities [Электронный ресурс] / T. Berners-Lee, J. Hendler, O. Lassila // Scientific American. — 2001. — Режим доступа: <https://www.scientificamerican.com/article/the-semantic-web/>, по подписке. — Загл. с экрана (дата обращения: 07.01.2022).

80. Bhatt, B. Expectation maximization algorithm for domain specific ontology extraction / B. Bhatt, P. Bhattacharyya // Research in Computing Science. — 2015. — Vol. 90. — P. 11–21.

81. Brewster, Ch. [и др.] The Ontology: Chimaera or Pegasus / Ch. Brewster, J. Iria, F. Ciravegna, Y. Wilks // Proceedings of the Dagstuhl Seminar Machine Learning for the Semantic Web. — Dagstuhl, Germany, 2005. — P. 13–18.

82. Brooks, B.S. [и др.] News reporting and writing / B.S. Brooks, G. Kennedy, D.R. Moen, D. Ranly. — Boston, MA: Bedford/St. Martin's, 2008. — 9th ed. — 214 p.

83. Cambridge Dictionary [Электронный ресурс]. — Режим доступа: <https://dictionary.cambridge.org>, свободный. — Загл. с экрана (дата обращения: 30.03.2018).

84. Chandioux, J. METEO : un système opérationnel pour la traduction automatique des bulletins météorologiques destinés au grand public / J. Chandioux // Meta. — Vol. 21 (2). — 1976. — P. 127–133.

85. Chaves, M. S. Towards a multilingual ontology for ontology-driven content mining in social web sites / M. S. Chaves, C. Trojahn // Proceedings of the 1st Workshop on Cross-Cultural and Cross-Lingual Aspects of the Semantic Web 2010. — 2010. — P. 1–10.

86. Chmielewski, M. [и др.] Semantic knowledge representation in terrorist threat analysis for crisis management systems / M. Chmielewski, A. Gałka, P. Jarema, K. Krawowski, A. Kosiński // Lecture Notes in Computer Science. — 2009. — Vol. 5796 (1). — P. 460–471.

87. Chrizman, N. How to construct a multi-lingual domain ontology / N. Chrizman, A. Itai // Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). — Reykjavik, Iceland, 2014. — P. 4345–4350.

88. Cioffi-Revilla, C. Introduction to computational social science: Principles and applications. — London: Springer, 2014. — 320 p.

89. Dictionnaires français Larousse [Электронный ресурс]. — Режим доступа: <https://www.larousse.fr/dictionnaires/francais>, свободный. — Загл. с экрана (дата обращения: 30.03.2018).

90. Dovring, K. Quantitative semantics in 18th century Sweden / K. Dovring // Public Opinion Quarterly. — 1954. — Vol. 18 (4). — P. 389–394.

91. Edmundson, H. P. New methods in automatic extracting / H. P. Edmundson // Journal of the Association for Computing Machinery. — 1969. — Vol. 16 (2). — P. 264–285.

92. Efimenko, I. V. OntosMiner family: Multilingual IE systems / I. V. Efimenko, V. F. Khoroshevsky, V. P. Klintsov // SPECOM'2004: 9th Conference. Speech and Computer. — Saint-Petersburg, Russia, 2004. — P. 716–720.

93. Embley, D. W. [и др.] Multilingual ontologies for cross-language information extraction and semantic search / D. W. Embley, S. W. Liddle, D. W. Lonsdale, Y. Tijerino // ER 2011: Conceptual Modeling. — 2011. — P. 147–160.

94. Espinoza, M. Enriching an ontology with multilingual information / M. Espinoza, A. Gómez-Pérez, E. Mena // The Semantic Web: Research and Applications. ESWC 2008. Lecture Notes in Computer Science. — 2008. — Vol. 5021. — P. 333–347.

95. European Council. Council Decision (CFSP) 2019/1341 of 8 August 2019 [Электронный ресурс]. — 2019. — Режим доступа: <https://eur-lex.europa.eu/legal-content/en/TXT/HTML/?uri=CELEX:32019D1341&from=en>, свободный. — Загл. с экрана (дата обращения: 07.01.2022).

96. Façanha, R. L. A systematic approach to review legacy schemas based on ontological analysis / R. L. Façanha, M. C. Cavalcanti, M. Campos // Metadata and Semantic Research. — 2019. — P. 63–75.

97. Farrar, S. A linguistic ontology for the Semantic Web / S. Farrar, T. Langendoen // *Glott International*. — 2003. — Vol. 7 (3). — P. 97–100.
98. Fitzpatrick, E. The status of telegraphic sublanguages / E. Fitzpatrick, J. Bachenko, D. Hindle // *Analyzing Language in Restricted Domains: Sublanguage Description and Processing* / ed. R. Grishman and R. Kittredge. — 1986. — P. 39–51.
99. Flaounas, I. [и др.] The structure of the EU mediasphere / I. Flaounas, M. Turchi, O. Ali, T. De Bie, N. Mosdell, J. Lewis, N. Cristianini // *PLoS ONE*. — 2010. — Vol. 5 (12). — P. 1–6.
100. Fong, A. C. M. *Multimedia engineering: A practical guide for Internet implementation* / A. C. M. Fong, S. C. Hui. — Chichester: Wiley, 2006. — 286 p.
101. Foxvog, D. Сус / D. Foxvog // *Theory and Applications of Ontology: Computer Applications*. — Springer Netherlands, 2010. — P. 259–278.
102. Francesconi, E. [и др.] Integrating a bottom-up and top-down methodology for building semantic resources for the multilingual legal domain semantic processing of legal texts / E. Francesconi, M. Montemagni, W. Peters, D. Tiscornia // *LNAE*. — 2010. — Vol. 6036. — P. 95–121.
103. Galjano, Ph. Theoretical investigation of terrorism. Ontology development / Ph. Galjano, V. Popovich // *Information Fusion and Geographic Information Systems*. — 2009. — P. 277–239.
104. Gauch, H. G. *Scientific method in practice* / H. G. Gauch. — Cambridge University Press, 2002. — 456 p.
105. Gómez-Pérez, A. *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the Semantic Web* / A. Gómez-Pérez, M. Fernandez-Lopez, O. Corcho. — L.: Springer, 2010. — 404 p.
106. Gruber, T. R. A translation approach to portable ontology specifications / T. R. Gruber // *Knowledge Acquisition*. — 1993. — Vol. 5 (2). — P. 199–220.
107. Guarino, N. Understanding, building and using ontologies / N. Guarino // *International Journal of Human-Computer Studies*. — 1997. — Vol. 46 (2–3). — P. 293–310.

108. Harris, Z. *Mathematical structures of language* / Z. Harris. — New York: Wiley & Sons, 1968. — 230 p.
109. Hartmann, J. [и др.] *Ontology metadata vocabulary and applications* / J. Hartmann, R. Palma, Y. Sure, M. C. Suárez-Figueroa, P. Haase, A. Gómez-Pérez, R. Studer // *OTM 2005: On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops*. — 2005. — P. 906–915.
110. Hearst, M. A. *Automatic acquisition of hyponyms from large text corpora* / M. A. Hearst // *Proceedings of the 14th International Conference on Computational Linguistics*. — Stroudsburg, PA: Association for Computational Linguistic, 1992. — P. 539–545.
111. Hois, J. *Modular ontologies for spatial information* / J. Hois. — Logos Verlag, 2013. — 293 p.
112. Holsti, O. R. *Content analysis for the social sciences and humanities* / O. R. Holsti. — Addison-Wesley Pub. Co., 1969. — 235 p.
113. Hsieh, H.-F. *Three approaches to qualitative content analysis* / H.-F. Hsieh, S. E. Shannon // *Qualitative Health Research*. — 2005. — Vol. 15 (9). — P. 1277–1288.
114. *Intelligent Content Analysis* [Электронный ресурс]. — Режим доступа: <https://melingo.com/en/text-analysis/ica/>, свободный. — Загл. с экрана (дата обращения: 07.01.2021).
115. Inyaem, U. [и др.] *Construction of fuzzy ontology-based terrorism event extraction* / U. Inyaem, Ch. Haruechaiyasak, Ph. Meesad, D. Tran // *Proceedings of the 2010 Third International Conference on Knowledge Discovery and Data Mining*. — Washington, DC: IEEE Computer Society, 2010. — P. 391–394.
116. Iosif, E. *Ontology-based information extraction under a bootstrapping approach* / E. Iosif, G. Petasis, V. Karkaletsis // *Semi-Automatic Ontology Development: Processes and Resources*. — Hershey, PA: IGI Global, 2012. — P. 1–21.
117. Janis, I. L. *The problem of validating content analysis* / I. L. Janis // *Language of politics: Studies in quantitative semantics* / H. D. Lasswell, N. Leites (eds.). — Cambridge: MIT Press, 1965. — P. 55–82.

118. Khoo, Ch. Identifying semantic relations in text for information retrieval and information extraction / Ch. Khoo, S.-H. Myaeng // *The Semantics of Relationships: An Interdisciplinary Perspective*. — Dordrecht: Kluwer, 2002. — P. 161–180.

119. Konys, A. An approach for ontology-based information extraction system selection and evaluation / A. Konys // *Przegląd Elektrotechniczny*. — 2015. — Vol. 1 (11). — P. 207–211.

120. Kort-Butler, L. A. Content analysis in the study of crime, media and popular culture / L. A. Kort-Butler // *Oxford Research Encyclopedia, Criminology and Criminal Justice*. — 2016. — P. 1–13.

121. Koufakis, A. [и др.] Invited keynote on IOT4SAFE 2020: Semantic Web technologies in fighting crime and terrorism: The CONNEXIONs approach / A. Koufakis, D. Chatzakou, G. Meditskos, Th. Tsikrika, S. Vrochidis, I. Kompatsiaris // *IOT4SAFE@ESWC 2020*. — 2020. — Режим доступа: <http://ceur-ws.org/Vol-2686/invited.pdf>, свободный. — Загл. с экрана (дата обращения: 15.08.2021).

122. Krippendorff, K. An examination of content analysis: a proposal for a general framework and an information calculus for message analytic situations : doctoral dissertation / Klaus Krippendorff. — University of Illinois, 1967. — 413 p.

123. Krippendorff, K. Content analysis: An introduction to its methodology / K. Krippendorff. — Beverly Hills, CA : SAGE Publications, 1980. — 188 p.

124. Krippendorff, K. Content analysis: An introduction to its methodology / K. Krippendorff. — Newbury Park, CA : SAGE Publications, 2018. — 4th ed. — 472 p.

125. Labov, W. Narrative analysis / W. Labov, J. Waletzky // *Essays on the Verbal and Visual Arts* / ed. J.Helm. — Seattle: University of Washington Press, 1967. — P. 12–44.

126. Lassila, O. The role of frame-based representation on the Semantic Web [Электронный ресурс] / O. Lassila, D. McGuinness // *Knowledge Systems Laboratory Report KSL-01-02, Stanford University*. — 2001. — Режим доступа: <https://ida.liu.se/ext/epa/ej/etai/2001/018/01018-etaibody.pdf>, свободный. — Загл. с экрана (дата обращения: 1.04.2017).

127. Lasswell, H. D. Why be quantitative / H. D. Lasswell // *Language of Politics*. — New York: Stewart, 1949. — P. 40–52.
128. Lehrberger, J. Automatic translation and the concept of sublanguage / J. Lehrberger // *Sublanguage: Studies of Language in Restricted Domains* // ed. R. Kittredge and J. Lehrberger. — Berlin: Walter de Gruyter, 1982. — 240 p.
129. Li, Y. [и др.] Assessing collaborative process in cscl with an intelligent content analysis toolkit / Y. Li, J. Wang, J. Liao, D. Zhao, R. Huang // *Proceedings of the Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, 2007. — P. 257–261.
130. Madsen, B. N. [и др.] Automatic ontology construction for a national term bank / B. N. Madsen, H. E. Thomsen, J. Halskov, T. Lassen // *Presenting Terminology and Knowledge Engineering Resources Online*. — 2010. — P. 502–533
131. Mahesh, K. Toward full-text ontology-based word-sense disambiguation / K. Mahesh, S. Nirenburg, S. Beale // *Recent Advances in Natural Language Processing*. — 2000. — Vol. 2. — P. 131–142.
132. Mahesh, K. Ontology development for machine translation: Ideology and methodology [Электронный ресурс] / K. Mahesh // *Memoranda in Computer and Cognitive Science, MCCS-96-292*. — Las Cruces, N.M.: New Mexico State University. — 1996. — Режим доступа: <https://cs.uwaterloo.ca/~cdimarco/pdf/mahesh.pdf>, свободный. — Загл. с экрана (дата обращения: 17.03.2017).
133. Manicheva, E. [и др.] The Compreno Semantic Model as an Integral Framework for a Multilingual Lexical Database / E. Manicheva, M. Petrova, E. Kozlova, T. G. Popova // *24th Int. Conf. on Computational Linguistics. Proc. of the 3rd Workshop on Cognitive Aspects of the Lexicon*. — 2012. — P. 215–230.
134. Mannes, A. Building a terrorism ontology [Электронный ресурс] / A. Mannes, J. Golbeck // *ISWC Workshop on Ontology Patterns for the Semantic Web*. — 2005. — Vol. 36. — Режим доступа: https://www.researchgate.net/profile/AaronMannes/publication/228878519_Building_a_terrorism_ontology/links/0deec51584d54a8400000000/Building-a-terrorism-ontology.pdf, свободный. — Загл. с экрана (дата обращения: 07.01.2022).

135. Mannes, A. Ontology building: A terrorism specialist's perspective [Электронный ресурс] / A. Mannes, J. Golbeck // 2007 IEEE Aerospace Conference. — 2007. — Режим доступа: <https://ieeexplore.ieee.org/document/4161607>, для зарег. пользователей. — Загл. с экрана (дата обращения: 07.01.2022).
136. Martínez, P. [и др.] TrendMiner: Large-scale cross-lingual trend mining summarization of real-time media streams / P. Martínez, I. Segura, Th. Declerck, J. L. Martínez // *Procesamiento del lenguaje natural*. — 2014. — Vol. 53. — P. 163–166.
137. Mayring, Ph. Qualitative content analysis: Theoretical foundation, basic procedures and software solution / Ph. Mayring. — Klagenfurt, 2014. — 144 p.
138. Merriam—Webster Dictionary [Электронный ресурс]. — Режим доступа: <https://www.merriam-webster.com/>, свободный. — Загл. с экрана (дата обращения: 07.01.2022).
139. Miller, G. A. [и др.] Introduction to WordNet: An on-line lexical database / G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, K. J. Miller // *International Journal of Lexicography*. — 1990. — Vol. 3 (4). — P. 235–244.
140. Miller, G. A. [и др.] Five papers on WordNet (CSL Report 43) / G. A. Miller, R. Beckwith, Ch. Fellbaum, D. Gross, K. J. Miller, R. I. Teng. — Princeton, NJ: Princeton University, Cognitive Science Laboratory, 1993. — 86 p.
141. Mindich, D. T. Z. Just the facts: How “objectivity” came to define American journalism / D. T. Z. Mindich. — New York: New York University Press, 1998. — 200 p.
142. Mizoguchi, R. Task ontology for reuse of problem-solving knowledge / R. Mizoguchi, J. Vanwelkenhuysen, M. Ikeda // *Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing*. — IOS Press, 1995. — P. 46–59.
143. Montiel-Ponsoda, E. [и др.] Modelling multilinguality in ontologies / E. Montiel-Ponsoda, G. Aguado de Cea, A. Gómez-Pérez, W. Peters // *Proceedings of COLING 2008, Companion volume — Posters and Demonstrations*. — Manchester: Coling 2008 Organizing Committee, 2008. — P. 67–70.
144. Montiel-Ponsoda, E. [и др.] Enriching ontologies with multilingual information / E. Montiel-Ponsoda, G. Aguado de Cea, A. Gómez-Pérez, W. Peters. — *Natural Language Engineering*. — 2011. — Vol. 17 (3). — P. 283–309.

145. Moreno, A. From text to ontology: Extraction and representation of conceptual information / A. Moreno, Ch. Pérez // Actes de quatrièmes rencontres « Terminologie et Intelligence Artificielle ». — Nancy, 2011. — P. 233–242.
146. Multilingual Computing and Technology. — 1999. — Vol. 10.
147. Nickles, M. [и др.] Ontologies across disciplines / M. Nickles, A. Pease, A. C. Schalley, D. Zaefferer // Ontolinguistics: How ontological status shapes the linguistic coding of concepts. — Berlin: Walter de Gruyter, 2007. — P. 23–67.
148. Niles, I. Linking lexicons and ontologies: Mapping WordNet to the Suggested Upper Merged Ontology / I. Niles, A. Pease // Proceedings of the 2003 International conference on Information and Knowledge Engineering (IKE 03). — Las Vegas, NV, 2003. — P. 412–416.
149. Nirenburg, S. [и др.] Lexicons in the MikroKosmos project [Электронный ресурс] / S. Nirenburg, S. Beale, K. Mahesh, B. Onyshkevych, V. Raskin, E. Viegas, Y. Wilks, R. Zajac // Proceedings of the AISB'96 Workshop on Multilinguality in the Lexicon. — 1996. — Режим доступа: http://www.cogsci.rpi.edu/~nirens/SergeiPapers/Nirenburg_Lexicons_1996.pdf, свободный. — Загл. с экрана (дата обращения: 02.12.2016).
150. Nirenburg, S. Ontological semantics / S. Nirenburg, V. Raskin. — Cambridge: MIT Press, 2004. — 440 p.
151. OntoText [Электронный ресурс]. — 2022. — Режим доступа: <https://www.ontotext.com/>, свободный. — Загл. с экрана (дата обращения: 07.01.2021).
152. Onyshkevych, B. A. An ontological-semantic framework for text analysis : Ph.D. thesis / Boyan A. Onyshkevych. — Carnegie Mellon University of Pittsburgh. — 1997. — 194 p.
153. Ozguven, K. An approach to machine translation / K. Ozguven, J. Tsujii // Proceedings of Turkish Artificial Intelligence and Neural Networks Symposium. — 1992. — P. 103–110.

154. Pöttker, H. News and its communicative quality: The inverted pyramid—When and why did it appear? / H. Pöttker // *Journalism Studies*. — 2003. — Vol. 4. — P. 501–511.

155. Rich, R. C. [и др.] *Empirical political analysis: Quantitative and qualitative research methods* / R. C. Rich, C. L. Brians, J. B. Manheim, L. Willnat. — NY: Routledge, 2018. — 9th ed. — 384 p.

156. Sheremetyeva, S. Acquisition of a language computational model for NLP / S. Sheremetyeva, S. Nirenburg // *The 18th International Conference on Computational Linguistics*. — 2000. — Vol. 2. — P. 1111–1115.

157. Sheremetyeva, S. Automatic extraction of linguistic resources in multiple languages / S. Sheremetyeva // *Proceedings of NLPCS 2012, 9th International Workshop on Natural Language Processing and Cognitive Science in conjunction with ICEIS 2012, Wroclaw, Poland*. — 2012. — P. 44–52.

158. Sheremetyeva, S. O. Linguistic models and tools for processing patent claims : monograph / S. O. Sheremetyeva. — Chelyabinsk: SUSU Publishing Centre, 2017. — 157 p.

159. Sheremetyeva, S. Universal computational formalisms and developer environment for rule-based NLP / S. Sheremetyeva // *International Conference on Computational Linguistics and Intelligent Text Processing. CICLing 2017. Lecture Notes in Computer Science*. — 2018. — Vol. 10761. — P. 67–78.

160. Sheremetyeva, S. Towards creating interoperable resources for conceptual annotation of multilingual domain corpora / S. Sheremetyeva // *Proceedings of the 16th Joint ACL-ISO Workshop Interoperable Semantic Annotation (ISA-16). Language Resources and Evaluation Conference (LREC 2020)*. — 2020. — P. 102–109.

161. Stine, R. A. Sentiment analysis / R. A. Stine // *Annual Review of Statistics and Its Application*. — 2019. — Vol. 6 (1). — P. 17.1–17.22.

162. Stone, Ph. J. [и др.] *The General Inquirer: A computer approach to content analysis* / Ph. J. Stone, D. Dunphy, M. S. Smith, D. M. Ogilvie. — Cambridge: MIT Press, 1966. — 651 p.

163. Streibel, O. Trend ontology for knowledge-based trend mining in textual information / O. Streibel, M. Mochol // 2010 Seventh International Conference on Information Technology: New Generations. — 2010. — P. 1285–1288.
164. Swartout, B. [и др.] Toward distributed use of large-scale ontologies / B. Swartout, R. Patil, K. Knight, T. Russ // Ontological Engineering, AAAI-97 Spring Symposium Series. — 1997. — P. 138–148.
165. Turner, M. D. A simple ontology for the analysis of terrorist attacks [Электронный ресурс] / M. D. Turner, D. M. Weinberg, J. A. Turner. — 2011. — Режим доступа: https://digitalrepository.unm.edu/ece_rpts/41, свободный. — Загл. с экрана (дата обращения: 05.01.2022).
166. Uschold, M. Ontologies: principles, methods and applications / M. Uschold, M. Gruninger // Knowledge Engineering Review. — 1996. — Vol. 11 (2). — P. 93–136.
167. Van Dijk, T. A. News as discourse / T. A. van Dijk. — Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., 1988. — 210 p.
168. Van Heist, G. Using explicit ontologies in KBS / G. Van Heist, A. T. Schreiber, B. J. Wielinga // International Journal of Human-Computer Studies. — 1997. — Vol. 46 (2/3). — P. 183–292.
169. Viegas, E. [и др.] Semantics in action / E. Viegas, K. Mahesh, S. Nirenburg, S. Beale // Predicative Forms in Natural Language and in Lexical Knowledge Bases / ed. by P. Saint-Dizier. — Kluwer Academic Publishers, 1999. — P. 171–204.
170. Viju, J. S. Concept interpretation by semantic knowledge harvesting / J. S. Viju // International Journal for Research in Applied Science and Engineering Technology (IJRASET). — 2018. — Vol. 6 (5). P. 477–484.
171. Weber, L. Ontology-based content analysis of US patent applications from 2001–2010 / L. Weber, T. Böhme, M. Irmer // Pharmaceutical patent analyst. — 2013. — Vol. 2 (1). — P. 39–54.
172. Wendelberg, L. An ontological framework to facilitate early detection of ‘radicalization’ (OFEDR)—A three world perspective / L. Wendelberg // Journal of Imaging. — 2021. — Vol. 7 (3). — P. 1–27.

173. Wielinga, B. J. Reusable and shareable knowledge bases: A European perspective / B. J. Wielinga, A. T. Schreiber // Proceedings of First International Conference on Building and Sharing. — Tokyo and Amsterdam: Ohmsha Ltd and IOS Press. — 1994. — P. 110–120.

174. Wise, K. [и др.] When words collide online: How writing style and video intensity affect cognitive processing of online news / K. Wise, P. Bolls, J. Myers, M. Sternadori // Journal of Broadcasting & Electronic Media. — 2009. — Vol. 53. — P. 632–546.

175. Witt, A. [и др.] Multilingual language resources and interoperability / A. Witt, U. Heid, F. Sasaki, G. Sérasset. — Language Resources and Evaluation. — 2009. — Vol. 43 (1). — P. 1–14.

ПРИЛОЖЕНИЕ А

Морфологическая репрезентация глаголов в подъязыке новостных сообщений предметной области «Терроризм»

Таблица 1. Морфологическая репрезентация глаголов в русском корпусе

Частота (F)	Личные формы						Прошедшее время	Инфинитив	Причастие					Деепричастие		Повелительное наклонение
	Настоящее время			Будущее время					Действительное		Страдательное			Настоящее	Прошедшее	
	1-е лицо	2-е лицо	3-е лицо	1-е лицо	2-е лицо	3-е лицо			Настоящее время	Прошедшее время	Настоящее	Прошедшее				
												Полное	Полное	Краткое		
F _{АБС}	132	9	4132	252	2	197	17579	1921	551	793	180	1776	3224	167	198	27
F _%	0,42	0,03	13,27	0,81	0,01	0,63	56,45	6,17	1,77	2,55	0,58	5,7	10,35	0,54	0,64	0,09

Таблица 2. Глаголы русского корпуса с самой высокой частотой встречаемости

Глагол	Личные формы						Прошедшее время	Инфинитив	Причастие					Деепричастие		Повелительное наклонение
	Настоящее время			Будущее время					Действительное		Страдательное			Настоящее	Прошедшее	
	1-е лицо	2-е лицо	3-е лицо	1-е лицо	2-е лицо	3-е лицо			Настоящее время	Прошедшее время	Настоящее	Прошедшее				
												Полное	Полное	Краткое		
быть	—	—	177	—	6	76	2119	89	—	—	—	—	—	—	29	1
погибнуть ³⁷	—	—	—	—	—	—	1601	2	—	—	—	—	—	—	—	—
произойти	—	—	—	—	—	2	1166	7	—	35	—	—	—	—	—	—
получить	—	—	—	—	—	2	820	8	—	13	—	—	35	5	—	3
сообщать	—	—	653	—	—	—	90	5	—	—	—	—	—	—	—	—
сообщить	—	—	—	—	—	—	644	7	—	—	—	—	—	—	—	1
взять	—	—	—	—	—	—	569	6	—	2	—	—	1	9	—	3
пострадать ³⁷	—	—	—	—	—	—	460	5	—	6	—	—	—	—	—	—
прогреть	—	—	—	—	—	—	433	—	—	25	—	—	—	—	—	—
стать	—	—	—	1	—	7	435	11	—	3	—	—	—	—	—	—

³⁷ В расчет не включена частота встречаемости словоформ *погибший*, *пострадавший*, поскольку они функционируют в рассматриваемом корпусе как субстантивированные причастия и, как следствие, отнесены к именам существительным.

Таблица 3. Морфологическая репрезентация глаголов в английском корпусе

Частота (F)	Present				Past				Participle			Infinitive		Imperative	Future	
	Simple	Perfect	Continuous	Perfect Cont.	Simple	Perfect	Continuous	Perfect Cont.	Present		Past	Simple	Continuous		Simple	Continuous
									Simple	Perfect						
F _{ABC}	4856	1100	318	44	13705	468	357	18	3449	18	2972	2478	28	28	257	1
F%	16,13	3,65	1,06	0,15	45,54	1,55	1,19	0,06	11,46	0,06	9,87	8,23	0,09	0,09	0,85	0

Таблица 4. Глаголы английского корпуса с самой высокой частотой встречаемости

Глагол	Present				Past				Participle			Infinitive		Imperative	Future	
	Simple	Perfect	Continuous	Perfect Cont.	Simple	Perfect	Continuous	Perfect Cont.	Present		Past	Simple	Continuous		Simple	Continuous
									Simple	Perfect						
be	1291	—	—	—	1500	—	—	—	213	—	230	211	—	—	—	—
say	223	18	2	—	2183	5	1	—	100	—	—	—	—	—	—	—
kill	66	130	—	—	548	24	—	—	184	2	972	29	—	—	—	—
have	338	—	—	—	236	—	—	—	29	—	4	158	—	—	—	—
injure	7	7	—	—	174	4	—	—	32	—	279	2	—	—	—	—
take	60	24	8	—	242	11	5	—	21	—	37	12	—	—	10	—
wound	5	—	—	—	160	4	—	—	26	2	161	—	—	—	—	—
tell	5	4	—	—	336	4	—	—	6	—	—	—	—	—	—	—
claim	63	55	2	—	186	4	—	—	16	—	16	6	—	—	—	—
shoot	7	4	—	—	118	3	—	—	144	—	58	—	—	—	2	—

Таблица 5. Морфологическая репрезентация глаголов во французском корпусе

Частота (F)	Indicatif												Conditionnel			Subjonctif			Infinitif		Participe			Gérondif	
	Présent			Passé composé			Imparfait		Plus-que-parfait		Passé antérieur	Futur simple		Présent		Passé	Présent		Passé	Présent	Passé	Présent	Passé composé		Passé
	1-е лицо	2-е лицо	3-е лицо	1-е лицо	2-е лицо	3-е лицо	1-е лицо	3-е лицо	1-е лицо	3-е лицо		3-е лицо	1-е лицо	3-е лицо	1-е лицо		3-е лицо	3-е лицо							
F _{ABC}	266	81	4879	149	1	6485	63	2234	1	814	6	36	151	1	151	106	4	95	18	2533	201	1095	68	6284	125
F _%	1,03	0,31	18,88	0,58	0	25,09	0,24	8,64	0	3,15	0,02	0,14	0,58	0	0,58	0,41	0,02	0,37	0,07	9,8	0,78	4,24	0,26	24,31	0,48

Таблица 6. Глаголы французского корпуса с самой высокой частотой встречаемости

Глагол	Indicatif												Conditionnel			Subjonctif			Infinitif		Participe			Gérondif	
	Présent			Passé composé			Imparfait		Plus-que-parfait		Passé antérieur	Futur simple		Présent		Passé	Présent		Passé	Présent	Passé	Présent	Passé composé		Passé
	1-е лицо	2-е лицо	3-е лицо	1-е лицо	2-е лицо	3-е лицо	1-е лицо	3-е лицо	1-е лицо	3-е лицо		3-е лицо	1-е лицо	3-е лицо	1-е лицо		3-е лицо	3-е лицо							
être	14	23	1593	2	—	1528	18	517	—	250	—	4	39	—	58	35	1	61	6	213	20	16	9	256	—
avoir	43	2	1126	2	—	167	3	254	—	19	—	1	17	1	—	—	—	9	—	54	5	6	—	16	1
faire	4	—	29	7	—	341	—	935	—	111	—	—	1	—	5	4	—	—	—	98	5	130	9	—	3
blessar	—	1	5	—	—	17	—	—	—	5	—	—	—	—	—	2	—	—	—	2	2	17	—	1462	18
mourir	1	—	—	—	—	79	—	—	—	3	—	—	—	—	—	4	—	—	—	13	—	—	—	1255	—
tuer	—	—	9	—	—	54	—	—	—	9	4	—	—	—	—	1	—	—	—	21	9	42	5	720	6
revendiquer	—	—	29	—	—	131	—	1	—	16	—	—	—	—	—	—	—	—	—	3	—	—	2	373	1
déclarer	—	—	1	—	—	366	—	—	—	10	—	—	—	—	—	3	—	—	1	1	—	1	1	20	—
piéger	—	—	7	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	379	—
indiquer	—	—	32	—	—	281	—	3	—	10	—	—	—	—	—	—	—	—	—	1	—	10	—	17	—

ПРИЛОЖЕНИЕ Б

Концептуальные классы лексических единиц предметной области «Терроризм»

1. «Взятие ответственности»

В этом концептуальном классе представлены лексические единицы (как правило, глаголы и глагольные группы) которые сигнализируют о взятии террористической организацией или террористом ответственности за теракт.

Примеры на русском языке: брать на себя ответственность, взять на себя ответственность, взять на себя вину, заявление об ответственности, ответственность за взрыв, ответственный, признать вину.

Примеры на английском языке: accept responsibility, claim attack, claim credit, claim culpability for, claim responsibility for the explosions, claim to have carried out, claim, responsibility for a series of terrorist attacks, responsibility, responsible.

Примеры на французском языке: absence de revendication, affirmer avoir mené, communiqué de revendication, démentir toute responsabilité, être responsable de, prendre la responsabilité, responsabilité, revendiquer la responsabilité, revendiquer.

2. «Время теракта»

В этом концептуальном классе представлены лексические единицы, отражающие дату, время и прочие временные параметры теракта. Концептуальный класс в основном содержит предложные группы и наречия, реже — имена существительные и прилагательные.

Примеры на русском языке: ближе к вечеру, в 11 часов дня по местному времени, в 2018 году, в день празднования католической Пасхи, в разгар предвыборной кампании, вечер понедельника, вечерний, вечером, ежедневно, на протяжении нескольких суток.

Примеры на английском языке: a few days ago, about three o'clock this morning, after the blast, after the evening incidents, April, around 7 a.m. local time, evening rush-hour, February, last Friday, middle of the day, national election day, New Year's Eve.

Примеры на французском языке: 11 décembre, à ce jour, à la fin d'un concert de la chanteuse américaine, depuis cet après-midi, dimanche soir, en 2018, en milieu de journée, fête de l'Achoura, heure locale, mardi, nuit de la Saint-Sylvestre, quelques heures plus tard, réveillon de la nouvelle année, un peu plus tôt.

3. «Заявление»

В этом концептуальном классе представлены лексические единицы, которые сигнализируют о цитате, высказывании. Концептуальный класс содержит в основном глаголы и глагольные группы.

Примеры на русском языке: выразить, высказаться, говорить, говорить на условиях анонимности, давать обновленную информацию, добавить, заявить, заявление, отказаться официально подтвердить, подробно излагать, подчеркнуть, по данным, по словам.

Примеры на английском языке: add, blame, broadcast, characterize, complain, declare, deny any involvement, express condolences, issue a statement, make a statement, post a message on his twitter account, refuse to officially confirm, take the floor, tweet.

Примеры на французском языке: accuser, affirmer, ajouter, citer, condamner, confirmer, d'après, évoquer, indiquer, interview, juger, message, pour sa part évoquer, principale information, proposer, se référer, témoignage, témoigner, tweeter.

4. «Идеология»

В этом концептуальном классе представлены лексические единицы, репрезентирующие идеологию, которой придерживаются террористы и террористические организации. Концептуальный класс представлен именами существительными, именными группами и прилагательными.

Примеры на русском языке: ваххабит, исламизм, национализм, радикализм, сепаратизм, ультраправый экстремизм, ультраправый.

Примеры на английском языке: anti-imperialism, extremism, jihadi feminism, militant extremism, radicalism, ultra-hardline islamist, wahhabi, white supremacism.

Примеры на французском языке: cercle islamiste local, idéologie radicale, organisation indépendantiste kurde, rébellion islamiste, suprémacisme blanc, terrorisme endogène, violence djihadiste.

5. «Источник»

Этот концептуальный класс содержит лексические единицы, представляющие источник сообщения о теракте или любого другого высказывания, в частности источник реакции на теракт (газеты, телевидение, информационные агентства, представители власти, свидетели, пострадавшие и т. д.). Концептуальный класс содержит именные группы, в том числе имена собственные.

Примеры на русском языке: 74-страничный документ, аккаунт полиции, американская администрация, британские СМИ, бывший министр обороны, военная прокуратура, Аль-Джазира, газета «Коммерсантъ», Евросоюз, НТВ.

Примеры на английском языке: ABC News, Afghan news portal, agence, authorities, Binali Yildirim, Bloomberg, breaking news, broadcaster, cable news network, cell phone video, chairman of the provincial council, Chinese embassy, communique.

Примеры на французском языке: AFP, agence Anadolu, ambulanciers, autorités israéliennes, chauffeur de bus, département d'État américain, gouverneur, journaliste, ministère de l'Intérieur, Numan Kurtulmus, police berlinoise, porte-parole du ministère de la Défense, quotidien Hurriyet, sécurité, source des services de sécurité, témoin.

6. «Контртерроризм»

В этом концептуальном классе представлены лексические единицы, используемые для описания представителей контртеррористической деятельности, контртеррористических мер и средств. Ввиду широкого охвата лексики концептуальный класс содержит лексемы, принадлежащие к разным частям речи.

Примеры на русском языке: активная разведка боевых порядков, антитеррористические органы, арест лидера ИГ, армия Сирии, беспилотник, блестящая операция, боец спецназа, в рамках расследования теракта, внимание правоохранительных органов, контролируемый подрыв подозрительного предмета.

Примеры на английском языке: administrative surveillance, Afghan forces, air force night raid, airstrike of US coalition, American special forces, anti-IS forces, anti-terrorism operation, army corps, arrest, counterterrorism, cybersecurity, defense, demining operation, fight against terrorism, government troops, investigation.

Примеры на французском языке: abattre, action résolue et courageuse, agent de la police, ambulance, antijihadiste, antiterroriste, appréhender, armée israélienne, arrestation, arrêter, autorités turques, base de lancement aux opérations des forces progouvernementales, bombarder, boucler, cellule de crise, chasse à l'homme, coalition antijihadiste, incriminer, interdit, juge.

7. «Направление теракта»

Концептуальный класс содержит лексические единицы, сигнализирующие о направленности теракта на кого-л. или что-л. В данном концептуальном классе представлено небольшое количество уникальных лексем, однако частота их встречаемости и значимость для предметной области достаточно высока для образования отдельного концептуального класса.

Примеры на русском языке: затронуть, направить, нацелить, объект многочисленных нападений, подвергаться многочисленным нападениям, против, цель.

Примеры на английском языке: against, aimed, directed to, mainly target, now touched, often targeted, target.

Примеры на французском языке: cibler, être la cible, être particulièrement touché, prendre pour cible, viser.

8. «Масштаб теракта»

Концептуальный класс содержит лексические единицы, сигнализирующие о крупном масштабе теракта, в основном прилагательные и именные группы.

Примеры на русском языке: волна кровавых терактов, кровавый, крупнейшая в мире серия терактов, крупномасштабное нападение, массовая казнь, массовые жертвы, масштабный теракт.

Примеры на английском языке: bloody, cruelest, deadliest, deadliest massacre, deadly attack, ISIS-claimed carnage, large-scale, massive, powerful, terrible.

Примеры на французском языке: attaque d'ampleur, attaque d'envergure, attaque la plus meurtrière, carnage, d'ampleur, de grande ampleur, de masse, d'envergure, épopée meurtrière, mares de sang, sanglant, tuerie.

9. «Место теракта»

В этот концептуальный класс включены лексические единицы, которые обозначают страну, регион, город, район, конкретное место или географический объект, где произошел теракт.

Примеры на русском языке: база в Афганистане, Баку, бедный квартал, близ госпиталя, в кадетской школе полиции, в Лондоне, в мечети, в окрестностях Пальмиры, в центре Будапешта, грузинская территория, значимый квартал Лондона.

Примеры на английском языке: 120 kilometers southeast of Baghdad, along the entire route, Amsterdam, apartment, army-controlled region, at a military camp, at the entrance, country's capital, district, in different areas of Turkey.

Примеры на французском языке: Ain Al Tamer, Etat du Minnesota, près de la province occidentale d'Al Anbar, région autonome du Kurdistan irakien, capitale allemande, dans le nord-est de la capitale, grande ville, ouest de la capitale afghane, cœur historique, haut lieu touristique, promenade populaire, Irak, à environ 5 km du bâtiment de la police, devant le stade, école, zone de marché.

10. «Обладание средством теракта»

Концептуальный класс содержит лексические единицы (в основном глаголы), указывающие на нахождение средства теракта где-либо (имеется при террористе, надето на него или помещено куда-либо, например в автомобиль).

Примеры на русском языке: в руке, владеющий, вооруженный ножом, вооруженный, закрепленный, заминированный, набитый взрывчаткой, начиненный.

Примеры на английском языке: armed with a knife, armed, bomb-laden, filled with, fixed, in his hand, laden, mined, mounted, wielding.

Примеры на французском языке: armé, bourré, chargé, engin piégé, homme armé, muni, piégé, porter, rempli.

11. «Объект теракта»

Концептуальный класс содержит лексические единицы, обозначающие одушевленные и неодушевленные объекты, на которые может быть направлен теракт.

Примеры на русском языке: 9-этажный дом, авиабаза, автобус КСИР, американский солдат, верующий, выпускник полицейской академии, высотное здание, главная городская мечеть, головной вагон поезда, гостинично-офисный комплекс, госучреждение, демонстрант-шиит, депо, депутат.

Примеры на английском языке: administration officer, Afghan base, airplane, airport, Al-Noor Mosque, American university, ancient temple, asylum seeker, busy fish market, cadet school, Canadian priest, checkpoint, Christchurch Mosque, civilian, club-goer, diplomatic mission, displaced resident.

Примеры на французском языке: chiite, groupe de personnes, jeune, piéton, représentants de l'occident honni, acheteur, gens ordinaires, convoi militaire, groupe de soldats israéliens, ambassadeur de Russie en Turquie, boite de nuit, discothèque, église, cathédrale syriaque catholique, lieux de culte, poste de police, car de policiers.

12. «Планы террористов»

В этот концептуальный класс включены лексические единицы, которые отображают намерения террориста или террористической организации, а также неудавшиеся попытки терактов.

Примеры на русском языке: готовить теракты, готовиться к атаке, заговор, задуманный разбой, замысел, новая попытка теракта, организация взрыва, очередное покушение, планировать, попытаться поджечь, хотеть устроить, быть в планах.

Примеры на английском языке: attempt to blow up, attempt an attack, choose a target, intent, part of our plan, plan to liberate the occupied territory, plan a terror attack, plan to undermine, plot, plot to harm, prepare a terrorist attack, try to drive into.

Примеры на французском языке: avoir l'intention de commettre, être en train d'organiser, planification, projeter de commettre, projeter un attentat, tentative d'attaque au couteau, tentative de meurtre, tentative d'homicide, vouloir viser.

13. «Последствия»

В этот концептуальный класс отнесены лексические единицы, репрезентующие любые последствия (как положительные, так и отрицательные) теракта для пострадавших, террористов, местности, зданий, окружающей среды и т. д. Следует отметить, что данный концептуальный класс близок по смыслу к концептуальному классу «Объект теракта», и многие лексические единицы были занесены в оба класса. Между этими классами проведено два основных различия: во-первых, концептуальный класс «Объект теракта» содержит только именные группы, в то время как «Последствия» описываются при помощи как именной (кто пострадал в результате теракта и какие повреждения получены), так и глагольной (как получены эти повреждения) лексики; во-вторых, под «Объектом теракта» понимается именно тот человек или объект, в отношении которого теракт был совершен целенаправленно, а в класс «Последствия» входят все люди и объекты, пострадавшие в результате теракта, в том числе случайно.

Примеры на русском языке: арестовать, без жертв, без сознания, безнаказанный, большое число жертв, в голову, в живот, в реанимации, в розыске, в тяжелом состоянии, взорваться, взять в плен, вспыхнуть, выжить, гибель мирных жителей, мертвый, обширный ожог, огонь, окровавленный.

Примеры на английском языке: aftermath, alive, among the dead, apprehend, arrest, attack victim, barotrauma, bleed, blow themselves up, broken glass shards, bullet wound, burnt stretcher, fatal wound, gas poisoning symptom, get injured, get seriously damaged, killed terrorist, knife.

Примеры на французском языке: activer sa charge explosive, arrestation, assassiné, atteint à l'abdomen, blesser légèrement, blessé au visage et à l'abdomen, blessé légèrement à la main, blessure légère, brûlé, cadavre, carcasse calcinée d'une voiture, carrosserie endommagée, causer la mort, colonne de fumée noire, corps martyrisé.

14. «Предположение»

Концептуальный класс содержит лексические единицы, указывающие на предположения властей, полицейских в отношении террористов, террористических организаций.

Примеры на русском языке: вероятная версия, вероятно, возложить вину, возможный теракт, может быть, обвинение, обвинить, подозреваемый, по всей вероятности, по-видимому.

Примеры на английском языке: accusation, accuse, alleged gunman, apparently, assume, assumption, blame, first version, investigation version, male suspect, may opt to kill, opinion, potential.

Примеры на французском языке: accuser, apparement, attentat terroriste présumé, attribuer, auteur présumé, hypothèse, indication, piste terroriste, probablement, soupçonner, voiture suspecte, vraisemblablement.

15. «Принадлежность к террористической организации»

Концептуальный класс содержит лексические единицы, указывающие на связь террориста или террористической группы (организации) с другой террористической организацией.

Примеры на русском языке: активный участник террористической организации, бандит международной террористической организации, боевик группировки, даишист, игиловец, связанный, сторонник, талиб, ячейка.

Примеры на английском языке: association, adherent, affiliated, Al-Qaeda-linked Al-Shabaab, Al-Qaeda affiliated group al-Shabaab, al Qaeda-linked insurgents, belong to, connection, heroic soldier of the caliphate, loyal to, member, pledge allegiance.

Примеры на французском языке: affilié, appartenance, appartenir, branche, émaner, émanation, faire partie, lié, membre, proche, représentant taliban, soldat du califat, soutenir, sympathisant.

16. «Прочая террористическая деятельность»

В этом концептуальном классе содержатся лексические единицы, описывающие деятельность, ведущуюся террористами и террористическими организациями, которая не является совершением и планированием терактов.

Примеры на русском языке: битва, боестолкновение, вербовка боевиков, вербовка в ряды террористов, возвращение различной криминальной деятельности, военная операция, воздвигать баррикады, вооруженное сопротивление, действия террористов, деятельность деструктивного псевдорелигиозного объединения, джихадистская пропаганда, изготовление бомбы.

Примеры на английском языке: active armed resistance, armed foray, arms smuggling, barbaric action of terrorists, bomb making, bomb-making factory, breeding ground for terror, financing of terrorism, formation of ISIS, recruitment of students.

Примеры на французском языке: apporter un support logistique, combat en Syrie voisine, contrôler, crime de guerre, dangereuse rébellion, longtemps opérer en toute impunité, négociation de paix, prendre la ville, radicalisé.

17. «Реакция общественности»

В этом концептуальном классе содержатся лексические единицы, описывающие реакцию (как физическую, так и психологическую) общественности на произошедший теракт.

Примеры на русском языке: возмущенный, глубокие соболезнования, день траура, компенсация семьям жертв трагедии, однодневный национальный траур, опечаленный, оплакивать, скорбеть, создать горячую линию.

Примеры на английском языке: condemn, drop flowers, grieve, incite more fury, makeshift memorials, mourn, national mourning, offer its condolences, react, saddened, show solidarity, stand with you.

Примеры на французском языке: à se réunir afin de faire face à la barbarie, adresser son soutien aux familles des victimes, cérémonie en l'honneur des victimes, condamner, décréter un deuil national de trois jours, dénoncer, en hommage aux victimes,

exprimer ses condoléances aux familles des victimes, ordonner la mise en berne des drapeaux, solidarité de la nation tout entière.

18. «Религия»

Концептуальный класс содержит лексические единицы, указывающие на религиозную принадлежность террористов и жертв терактов.

Примеры на русском языке: буддистское население, ислам, католик, коптская церковь, мусульманин, суннит, шиит, христианин-копт, христианство.

Примеры на английском языке: Baluch Sunnite, Christian, Christian priest, Coptic Christian, imprisoned Islamic militant.

Примеры на французском языке: cathédrale syriaque catholique, chiite, islam chiite, minorité chiite, musulman chiite, pachtoune sunnite.

19. «Средство теракта»

Концептуальный класс охватывает лексические единицы, обозначающие оружие или подобные ему объекты, используемый для совершения теракта, а также функциональные элементы такого оружия.

Примеры на русском языке: 120-мм миномет, автомат, автомобиль, автомобильный таран, автоцистерна, белый грузовик, беспилотный летательный аппарат, бесхозный предмет, боеприпас, бутылка с зажигательной смесью, взрыватель, взрывпакет, взрывчатка, горчичный газ, джихадмобиль, дистанционное взрывное устройство, чемодан с проводами, шуруп.

Примеры на английском языке: 82-mm mortar, air bomb, ammonium nitrate, anti-tank mine, armed drone, assault rifle, bomb, bomb-laden vehicle, car ram, cartridge clip, cellphone-detonated IED, chlorine, climatic weapon, crude bomb, detonation device, dynamite, explosive belt, firearms, handgun, hexogen, homemade radio-controlled bomb.

Примеры на французском языке: arme blanche, bombe, charge explosive, explosif, fusil d'assaut, fusil mitrailleur, grenade, gilets d'explosifs, poids lourd, véhicule bélier, voiture piégée, véhicule rempli d'explosifs.

20. «Страна»

Концептуальный класс содержит имена собственные, обозначающие названия стран; именные группы, включающие такие имена собственные; а также другие лексические единицы, указывающие на место происхождения или национальность жертвы теракта, террориста или принадлежность террористической организации.

Примеры на русском языке: Австрия, албанец, албанский экстремист, Алжир, алжирец, американец, американские СМИ, американский, арабка, Буркина Фасо, власти Египта, власти Нидерландов, Грузия, египетский, египтянин, житель России, Израиль, Иран, иранец.

Примеры на английском языке: Algerian, Armenian, Belgian national, Burkina Faso, citizen of France, Colombian, Egypt's North Sinai, ethnic Turk, Filipino citizen, Iranian pilgrim, Iraqi, Kurdish gunman, Morocco, Russian woman, Syrian city, US, US-led coalition.

Примеры на французском языке: Afghanistan, Algérie, américain, assaillant palestinien blessé, au Mali, Brésil, burkinabè, capitale allemande, chinois ouïghour, djihadiste ouzbek, égyptien, en Turquie, innocents civils afghans et étrangers, kirghiz.

21. «Террорист»

Концептуальный класс содержит лексические единицы, описывающие лиц, занимающихся террористической деятельностью в составе террористической группы (организации) или самостоятельно. Концептуальный класс состоит в основном из имен существительных и именных групп.

Примеры на русском языке: 12-летний мальчик, 18-летний стрелок, 8-летняя девочка-смертница, автоматчик, автор нового террористического эпизода, авторитетный член ИГИЛ, алжирский террорист, алматинский стрелок, афганский талиб, боевик-экстремист, Брентон Таррант, вооруженный радикал, даишист, джихадист, женщина-смертница, игиловец.

Примеры на английском языке: 22-year-old college student, 8-year-old girl, accomplice, al Qaeda terrorist, Albanian extremist, alleged attacker, alleged perpetrator, armed gang, arsonist, attacker, Daesh supporter, Dozo hunter, female suicide bomber,

female terrorist, former Taliban insurgent, heavily-armed militant, hi-jacker, jihadi, killer, Kurdish separatist, male suspect.

Примеры на французском языке: adolescent, agresseur, artificier, assaillant palestinien blessé, assaillant terroriste, auteur matériel, bombe humaine, Brenton Tarrant, cerveau, chauffeur du camion, combattant iranien, combattant, complice, conducteur, délinquant radicalisé, extrémiste musulman local, homme armé déguisé en Père Noël, homme en armes, insurgé Taliban, jeune homme, seul assaillant, soldat, tireur, type.

22. «Террористическая организация»

Этот концептуальный класс включает лексические единицы, представляющие организации, которые занимаются планированием и осуществлением терактов для достижения определенных политических целей. В концептуальном классе содержатся в основном именные группы, в том числе имена собственные.

Примеры на русском языке: Алеф, Аль-Каида, Аль-Шабаб, Ан-Нусра, Аум Синрикё, Боко Харам, Виляят Хорасан, вооруженное формирование Исламского государства, Джабхат Фатх аш-Шам, радикальная группировка, террористическая группировка.

Примеры на английском языке: Ahrar al-Sham, al Nusra, al Qaeda, al-Shabaab, Ansar Allah, Aum Shinrikyo, Basque separatist group ETA, cell of the Islamic group, faction, Hezbollah, Islamic State of Iraq and al-Sham, Jamaat-ul-Ahrar, Khorasan, PKK.

Примеры на французском языке: Ahrar al-Cham, al Chabaab, al-Nosra, al-Qaïda au Maghreb Islamique, Boko Haram, Etat islamique en Irak et au Levant, ex-branche syrienne d'al-Qaïda, groupe islamiste, mouvement islamiste local, mouvements d'extrême gauche ou anarchistes grecs, PKK, Taliban.

23. «Тип теракта»

В этом концептуальном классе содержатся лексические единицы (именные и глагольные группы), указывающие на теракт и на способ его совершения (тип).

Примеры на русском языке: августовский подрыв террористов-смертников, автомобильный взрыв, автомобильный теракт, акт крайнего насилия, акт террора,

акт терроризм, активировать, артналет, атака беспилотного летательного аппарата, атаковать, бойня, бомбежка, бомбить, варварская атака, взорвать, взрыв автобуса, взрыв велосипеда со взрывчаткой, взять в заложники, воздушный теракт, дистанционный подрыв накладного заряда.

Примеры на английском языке: airport shooting, armed attack, arson, artillery strike, assassination, attack, attempted hijacking, blast, blow himself up, bomb explosion, bombing, capture, carnage, chaotic fire, chemical attack, detonate suicide vests, detonation of a hand grenade, direct hit of a mine, exploded car, fire, gunfight, hostage taking, incident, intimidation act, lone-wolf attack, massacre, plough, truck attack.

Примеры на французском языке: assassinat, attaque, attaque armée, attaque au camion bélier, attaque à la voiture piégée, attaque coordonnée, attentat, attentat à la bombe, attentat suicide, carnage, déflagration, double attentat à la bombe, enlèvement, explosion, explosion meurtrière, foncer, fusillade, percuter, premier coup d'éclat, prise d'otages, raid macabre, s'être fait exploser, s'être fait sauter.

24. «Террористическая угроза»

В концептуальном классе содержатся лексические единицы, которые указывают на террористическую угрозу, запугивание населения террористами, например, посредством демонстрации видео, ложными сообщениями о минировании.

Примеры на русском языке: видеозапись с угрозами, кадры последствий большого взрыва, манифест чёрного халифата, серьезная угроза глобальной безопасности, угрожать, транслировать, фотография тел.

Примеры на английском языке: audiotape, live-stream, multi-page manifesto, online streaming of an attack, propaganda video, threaten, threat to blow up airports.

Примеры на французском языке: manifeste, menacer, menace djihadiste, menace terroriste, vidéo.

25. «Цель теракта»

В этот концептуальный класс объединены лексические единицы, описывающие цель, которую террористы планируют достичь в результате теракта, либо

причину, по которой теракт был совершен. Было принято решение объединить цель и причину в один концептуальный класс, поскольку эти два понятия тесно взаимосвязаны в рассматриваемой предметной области, а также ввиду низкой частоты встречаемости лексических единиц, выражающих данные понятия.

Примеры на русском языке: 15 тысяч долларов, большое число жертв, введение шариата, возмездие за недавние бомбардировки, дестабилизация обстановки в идлибской зоне деэскалации, изменение территориальной целостности Российской Федерации, личная месть, освободить заключенных товарищей, отвлечь внимание, отделение Минданао от Филиппин, разжигать межрелигиозную рознь.

Примеры на английском языке: achieve independence for the Kurds, attempt to impose a radical version of Sharia law, avenge for, bargain for the release, destabilize the political situation, divert attention, do harm, release imprisoned comrades, revenge, seizure of weapons, sow seeds of discord, state independence, terrorist motive, withdrawal

Примеры на французском языке: assassiner des Juifs, causer un grand nombre de victimes, libérer des musulmanes, perturber la vie quotidienne, pour venger le drame de la ville d'Alep, renverser le gouvernement somalien.

ПРИЛОЖЕНИЕ В

Иерархическая структура концептов онтологии терроризма

Онтология терроризма содержит 113 концептов типов ОБЪЕКТ и EVENT, 27 концептов типа RELATION и пять концептов типа ATTRIBUTE. В приведенной ниже иерархической структуре концептов курсивом выделены концепты верхнего уровня онтологии μ К, которые не учитывались при подсчете количества концептов онтологии терроризма. В круглых скобках указаны теги, которыми данные концепты размечаются в ходе онтологического анализа.

ALL

└─>*EVENT*

└─>*MENTAL-EVENT*

└─>ASSUME (I)³⁸

└─>*PHYSICAL-EVENT*

└─>CLAIM-RESPONSIBILITY (CR)

└─>CONSEQUENCES (P)

└─>CONSEQUENCES-FOR-PEOPLE (Pa)

└─>PEOPLE-DYING (Paa)

└─>PEOPLE-INJURING (Pab)

└─>TAKEN-HOSTAGE (Pac)

└─>VICTIM-SURVIVAL (Pad)

└─>CONSEQUENCES-FOR-TERRORIST (Pb)

└─>TERRORIST-ELIMINATION (Pba)

└─>TERRORIST-INJURING (Pbb)

└─>TERRORIST-SUICIDE (Pbc)

└─>TERRORIST-CAPTURING (Pbd)

└─>TERRORIST-FLEEING (Pbe)

³⁸ В скобках указан тег, используемый для концептуального аннотирования.

- └─>DAMAGE (Pc)
- └─>DECLARE (D)
- └─>*SOCIAL-EVENT*
- └─>COUNTERTERRORISM-EVENT (Rb)
- └─>*CRIMINAL-ACTIVITY*
- └─>TERROR-ATTACK (T)
 - └─>ARMED-ASSAULT (Ta)
 - └─>GUN-ATTACK (Taa)
 - └─>KNIFE-ATTACK (Tab)
 - └─>ARSON (Tb)
 - └─>BIOLOGICAL-ATTACK (Tc)
 - └─>BOMB-ATTACK (Td)
 - └─>CAR-BOMB-ATTACK (Tda)
 - └─>SUICIDE-BOMB-ATTACK (Tdb)
 - └─>CHEMICAL-ATTACK (Te)
 - └─>DISTANT-ATTACK (Tf)
 - └─>DRONE-ATTACK (Tfa)
 - └─>MORTAR-ATTACK (Tfb)
 - └─>ROCKET-ATTACK (Tfc)
 - └─>HIJACKING (Tg)
 - └─>HOSTAGE-TAKING (Th)
 - └─>KIDNAPPING (Tha)
 - └─>BARRICADING (Thb)
 - └─>ASSASSINATION (Ti)
 - └─>SUICIDE-ATTACK (Tj)
 - └─>SUICIDE-BOMB-ATTACK (Tdb)
 - └─>UNARMED-ASSAULT (Tk)
 - └─>VEHICLE-RAMMING-ATTACK (Tl)
 - └─>CYBER-ATTACK (Tm)
- └─>TERROR-ATTACK-PLANNING-EVENT (K)

└─>TERRORIST-THREAT (E)

└─>OTHER-TERRORIST-EVENT (O)

└─>TERRORISM-FINANCING (Oa)

└─>TERRORIST-RECRUITING (Ob)

└─>WAR-RELATED-TERRORIST-EVENT (Oc)

└─>COMMUNITY-REACTION (Y)

└─>*OBJECT*

└─>*MENTAL-OBJECT*

└─>GOAL-OF-ATTACK (X)

└─>HARM-PEOPLE (Xa)

└─>MILITARY-GOAL (Xb)

└─>POLITICAL-GOAL (Xc)

└─>PROPERTY-DAMAGE (Xd)

└─>RELIGIOUS-GOAL (Xe)

└─>TIME (B)

└─>EXACT-TIME (Ba)

└─>EXACT-DATE (Bb)

└─>CIRCUMSTANCES (Bc)

└─>*PHYSICAL-OBJECT*

└─>AFFECTED-OBJECT (Hc)

└─>*ARTIFACT*

└─>COUNTERTERRORISM-INSTRUMENT (Rc)

└─>OBJECT-TARGET (Zc)

└─>PROPERTY-TARGET (Zb)

└─>CIVIL-PROPERTY (Zba)

└─>EDUCATIONAL-INSTITUTION (Zbb)

└─>ENTERTAINMENT-FACILITY (Zbc)

└─>MILITARY- INSTITUTION (Zbd)

└─>POLICE-FACILITY (Zbe)

└─>GOVERNMENTAL-INSTITUTION (Zbf)

- └─>RELIGIOUS-INSTITUTION (Zbg)
- └─>MEDICAL-INSTITUTION (Zbh)
- └─>TRANSPORT-FACILITY (Zbi)
- └─>WEAPON (C)
 - └─>BIOLOGICAL-WEAPON (Ca)
 - └─>CHEMICAL-WEAPON (Cb)
 - └─>GUN (Cc)
 - └─>EXPLOSIVE-DEVICE (Cd)
 - └─>SUICIDE-VEST (Cda)
 - └─>CAR-BOMB (Cdb)
 - └─>INCENDIARY-WEAPON (Ce)
 - └─>KNIFE (Cf)
 - └─>RAMMING-VEHICLE (Cg)
 - └─>ROCKET (Ch)
- └─>LOCATION (L)
 - └─>CITY (La)
 - └─>CAPITAL-CITY (Laa)
 - └─>CITY-AREA (Lb)
 - └─>COUNTRY (Lc)
 - └─>COUNTRY-AREA (Ld)
 - └─>SPECIFIC-LOCATION (Le)
 - └─>TERRITORY (Lf)
- └─>*HUMAN*
 - └─>COUNTERTERRORISM-AGENT (Ra)
 - └─>TERRORISM-AGENT (A)
 - └─>INDIVIDUAL-TERRORIST (Aa)
 - └─>TERRORIST-GROUP (Ab)
 - └─>HUMAN-TARGET (Za)
 - └─>CIVIL-TARGET (Zaa)
 - └─>MILITARY-TARGET (Zab)

- └─>POLICE-TARGET (Zac)
- └─>POLITICAL-TARGET (Zad)
- └─>RELIGIOS-TARGET (Zae)
- └─>VICTIM (Ha)
- └─>AFFECTED-TERRORIST (Hb)
- └─>*SOCIAL-OBJECT*
 - └─>*GEOPOLITICAL-ENTITY*
 - └─>NATION (N)
 - └─>IDEOLOGY (W)
 - └─>ISLAMISM (Wa)
 - └─>RIGHT-WING-EXTREMISM (Wb)
 - └─>LEFT-WING-EXTREMISM (Wc)
 - └─>WHITE-SUPREMACISM (Wd)
 - └─>NATIONALISM (We)
 - └─>SEPARATISM (Wf)
 - └─>*ORGANIZATION*
 - └─>TERRORIST-ORGANIZATION (U)
 - └─>SOURCE (S)
 - └─>RELIGION (G)
 - └─>ISLAM (Ga)
 - └─>CHRISTIANITY (Gb)
- └─>*PROPERTY*
 - └─>*RELATION*
 - └─>*DISCOURSE-RELATION*
 - └─>*REASON*
 - └─>PRECONDITION
 - └─>PRECONDITION-OF
 - └─>*CASE-ROLE*
 - └─>AGENT
 - └─>DIRECTION (DA)

- └─>INSTRUMENT
- └─>LOCATED
- └─>*INVERSE-CASE-ROLE*
 - └─>AGENT-OF
 - └─>DIRECTION-OF
 - └─>INSTRUMENT-OF
 - └─>LOCATION-OF
- └─>*PARTITIVE-RELATION*
 - └─>PART-OF
 - └─>HAS-PART
- └─>*OBJECT-RELATION*
 - └─>MEMBER-OF (Q)
 - └─>CAPITAL-OF
 - └─>GOAL-OF
 - └─>WEAPON-OF
- └─>*INVERSE-OBJECT-RELATION*
 - └─>HAS-MEMBER
 - └─>HAS-CAPITAL
 - └─>HAS-GOAL
 - └─>HAS-IDEOLOGY
 - └─>HAS-NATIONALITY
 - └─>HAS-RELIGION
 - └─>HAS-TYPE
 - └─>HAS-WEAPON (HC)
- └─>*OBJECT-EVENT-RELATION*
 - └─>CLAIMED-RESPONSIBILITY-FOR
- └─>*INVERSE-OBJECT-EVENT-RELATION*
 - └─>RESPONSIBILITY-CLAIMED-BY
- └─>*ATTRIBUTE*
 - └─>*LITERAL-ATTRIBUTE*

└─>GENDER (gen)

└─>SCALE-OF-ATTACK (M)

└─>NEGATION (NEG)

└─>ASSUMPTION (I)

└─>*SCALAR-ATTRIBUTE*

└─>AGE (age)

ПРИЛОЖЕНИЕ Г

Фреймы концептов онтологии терроризма

1. Фрейм концепта AFFECTED-OBJECT

LABEL	AFFECTED-OBJECT	VAL
DEFINITION	Неодушевленный объект (место, здание, предмет и т.п.), поврежденный в результате теракта	VAL
IS-A	PHYSICAL-OBJECT	VAL
THEME-OF	DAMAGE	SEM
NEGATION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	пораженный объект	VAL
DEFAULT-NAME-EN	affected object	VAL
DEFAULT-NAME-FR	objet touché	VAL

2. Фрейм концепта AFFECTED-TERRORIST

LABEL	AFFECTED-TERRORIST	VAL
DEFINITION	Террорист, пострадавший в результате теракта либо в результате собственных действий, либо в результате действий представителей контртеррористической деятельности	VAL
IS-A	HUMAN	VAL
GENDER	MALE, FEMALE	VAL
AGE	>0	VAL
HAS-RELIGION	RELIGION	SEM
HAS-NATIONALITY	NATION	SEM
THEME-OF	CONSEQUENCES-FOR-TERRORIST	SEM
ASSUMPTION	TRUE, FALSE	VAL
NEGATION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	затронутый террорист	VAL
DEFAULT-NAME-EN	affected terrorist	VAL
DEFAULT-NAME-FR	terroriste touché	VAL

3. Фрейм концепта ARMED-ASSAULT

LABEL	ARMED-ASSAULT	VAL
DEFINITION	Нападение, совершаемое террористом-одиночкой или группой террористов с использованием оружия	VAL
IS-A	TERROR-ATTACK	VAL
SUBCLASSES	GUN-ATTACK, KNIFE-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, PROPERTY-TARGET, OBJECT-TARGET	SEM
INSTRUMENT	GUN, KNIFE	SEM

LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	вооруженное нападение	VAL
DEFAULT-NAME-EN	armed assault	VAL
DEFAULT-NAME-FR	attaque armée	VAL

4. Фрейм концепта ARSON

LABEL	ARSON	VAL
DEFINITION	Преступление, заключающееся в умышленном поджоге здания или имущества с целью запугивания населения и достижения определенных политических целей.	VAL
IS-A	TERROR-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	OBJECT-TARGET, PROPERTY-TARGET	SEM
INSTRUMENT	INCENDIARY-WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	поджог	VAL
DEFAULT-NAME-EN	arson	VAL
DEFAULT-NAME-FR	incendie criminel	VAL

5. Фрейм концепта ASSASSINATION

LABEL	ASSASSINATION	VAL
DEFINITION	Убийство политического или общественного деятеля, как правило, по идейным мотивам	VAL
IS-A	TERROR-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM

DIRECTION	HUMAN-TARGET	SEM
INSTRUMENT	WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	POLITICAL-GOAL	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	убийство по политическим мотивам	VAL
DEFAULT-NAME-EN	assassination	VAL
DEFAULT-NAME-FR	magnicide	VAL

6. Фрейм концепта ASSUME

LABEL	ASSUME	VAL
DEFINITION	Допускать какую-либо возможности в связи с террористической деятельностью; например, предполагать, что теракт совершен определенной группировкой или что произошедшее является терактом	VAL
IS-A	MENTAL-EVENT	VAL
NEGATION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	предположение	VAL
DEFAULT-NAME-EN	assumption	VAL
DEFAULT-NAME-FR	assumption	VAL

7. Фрейм концепта BARRICADING

LABEL	BARRICADING	VAL
DEFINITION	Захват заложника с баррикадированием здания	VAL
IS-A	HOSTAGE-TAKING	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, PROPERTY-TARGET	SEM
INSTRUMENT	WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL

ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	баррикадирование	VAL
DEFAULT-NAME-EN	barricading	VAL
DEFAULT-NAME-FR	barricades	VAL

8. Фрейм концепта BIOLOGICAL-ATTACK

LABEL	BIOLOGICAL-ATTACK	VAL
DEFINITION	Теракт с использованием биологического оружия	VAL
IS-A	TERROR-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET	SEM
INSTRUMENT	BIOLOGICAL-WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	биологическая атака	VAL
DEFAULT-NAME-EN	biological attack	VAL
DEFAULT-NAME-FR	attaque biologique	VAL

9. Фрейм концепта BIOLOGICAL-WEAPON

LABEL	BIOLOGICAL-WEAPON	VAL
DEFINITION	Вид оружия массового поражения, действие которого основано на использовании болезнетворных свойств микроорганизмов и продуктов их жизнедеятельности	VAL
IS-A	WEAPON	VAL
INSTRUMENT-OF	BIOLOGICAL-ATTACK	SEM
WEAPON-OF	TERRORISM-AGENT, ARTIFACT	SEM
DEFAULT-NAME-RU	биологическое оружие	VAL
DEFAULT-NAME-EN	biological weapon	VAL
DEFAULT-NAME-FR	arme biologique	VAL

10. Фрейм концепта BOMB-ATTACK

LABEL	BOMB-ATTACK	VAL
DEFINITION	Теракт с использованием взрывного устройства	VAL
IS-A	TERROR-ATTACK	VAL

SUBCLASSES	CAR-BOMB-ATTACK, SUICIDE-BOMB-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, PROPERTY-TARGET, OBJECT-TARGET	SEM
INSTRUMENT	EXPLOSIVE-DEVICE	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	взрыв	VAL
DEFAULT-NAME-EN	bomb attack	VAL
DEFAULT-NAME-FR	attentat à la bombe	VAL

11. Фрейм концепта CAPITAL-CITY

LABEL	CAPITAL-CITY	VAL
DEFINITION	Столица государства	VAL
IS-A	CITY	VAL
INSTANCES	CAPITAL-CITY-1, CAPITAL-CITY-2, CAPITAL-CITY-3 ...	VAL
CAPITAL-OF	COUNTRY, COUNTRY-AREA, TERRITORY, NATION	SEM
PART-OF	COUNTRY-AREA, COUNTRY, TERRITORY, NATION	SEM
DEFAULT-NAME-RU	столица	VAL
DEFAULT-NAME-EN	capital	VAL
DEFAULT-NAME-FR	capitale	VAL

12. Фрейм концепта CAR-BOMB

LABEL	CAR-BOMB	VAL
DEFINITION	Автомобиль, начиненный взрывчаткой, используемый в качестве средства теракта	VAL
IS-A	EXPLOSIVE-DEVICE	VAL
INSTRUMENT-OF	CAR-BOMB-ATTACK	SEM
WEAPON-OF	ARTIFACT	SEM
DEFAULT-NAME-RU	начиненный взрывчаткой автомобиль	VAL
DEFAULT-NAME-EN	car bomb	VAL
DEFAULT-NAME-FR	voiture piégée	VAL

13. Фрейм концепта CAR-BOMB-ATTACK

LABEL	CAR-BOMB-ATTACK	VAL
-------	-----------------	-----

DEFINITION	Теракт с использованием взрывного устройства, помещенного в автомобиль или прикрепленного к нему	VAL
IS-A	BOMB-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, PROPERTY-TARGET, OBJECT-TARGET	SEM
INSTRUMENT	CAR-BOMB	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	подрыв начиненного взрывчаткой автомобиля	VAL
DEFAULT-NAME-EN	car bomb attack	VAL
DEFAULT-NAME-FR	attentat à la voiture piégée	VAL

14. Фрейм концепта CHEMICAL-ATTACK

LABEL	CHEMICAL-ATTACK	VAL
DEFINITION	Теракт с использованием химического оружия	VAL
IS-A	TERROR-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET	SEM
INSTRUMENT	CHEMICAL-WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	химическая атака	VAL
DEFAULT-NAME-EN	chemical attack	VAL
DEFAULT-NAME-FR	attaque chimique	VAL

15. Фрейм концепта CHEMICAL-WEAPON

LABEL	CHEMICAL-WEAPON	VAL
-------	-----------------	-----

DEFINITION	Оружие массового поражения, действие которого основано на токсических свойствах отравляющих веществ	VAL
IS-A	WEAPON	VAL
INSTRUMENT-OF	CHEMICAL-ATTACK	SEM
WEAPON-OF	TERRORISM-AGENT, ARTIFACT	SEM
DEFAULT-NAME-RU	химическое оружие	VAL
DEFAULT-NAME-EN	chemical weapon	VAL
DEFAULT-NAME-FR	arme chimique	VAL

16. Фрейм концепта CHRISTIANITY

LABEL	CHRISTIANITY	VAL
DEFINITION	Авраамическая мировая религия, основанная на жизни и учении Иисуса Христа, описанных в Новом Завете	VAL
IS-A	RELIGION	VAL
DEFAULT-NAME-RU	христианство	VAL
DEFAULT-NAME-EN	christianity	VAL
DEFAULT-NAME-FR	chrétienté	VAL

17. Фрейм концепта CIRCUMSTANCES

LABEL	CIRCUMSTANCES	VAL
DEFINITION	Прочие временные обстоятельства теракта, отличные от даты и времени (например, какое-либо празднование)	VAL
IS-A	TIME	VAL
DEFAULT-NAME-RU	обстоятельства	VAL
DEFAULT-NAME-EN	circumstances	VAL
DEFAULT-NAME-FR	circonstances	VAL

18. Фрейм концепта CITY

LABEL	CITY	VAL
DEFINITION	Крупный центр обитания людей, в котором произошел теракт	VAL
IS-A	LOCATION	VAL
SUBCLASSES	CAPITAL-CITY	VAL
PART-OF	COUNTRY-AREA, COUNTRY, TERRITORY, NATION	SEM
HAS-PART	CITY-AREA, SPECIFIC-LOCATION	SEM
LOCATION-OF	TERROR-ATTACK	SEM
INSTANCES	CITY-1, CITY-2, CITY-3, CITY-4, CITY-5, CITY-6 ...	VAL
DEFAULT-NAME-RU	город	VAL
DEFAULT-NAME-EN	city	VAL
DEFAULT-NAME-FR	ville	VAL

19. Фрейм концепта CITY-AREA

LABEL	CITY-AREA	VAL
-------	-----------	-----

DEFINITION	Часть города, в которой произошел теракт	VAL
IS-A	LOCATION	VAL
PART-OF	CITY	SEM
LOCATION-OF	TERROR-ATTACK	SEM
INSTANCES	CITY-AREA-1, CITY-AREA-2, CITY-AREA-3 ...	VAL
DEFAULT-NAME-RU	район города	VAL
DEFAULT-NAME-EN	city area	VAL
DEFAULT-NAME-FR	secteur d'une ville	VAL

20. Фрейм концепта CIVIL-PROPERTY

LABEL	CIVIL-PROPERTY	VAL
DEFINITION	Недвижимое имущество, не принадлежащее государству	VAL
IS-A	PROPERTY-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	частная собственность	VAL
DEFAULT-NAME-EN	private property	VAL
DEFAULT-NAME-FR	propriété privée	VAL

21. Фрейм концепта CIVIL-TARGET

LABEL	CIVIL-TARGET	VAL
DEFINITION	Представители мирного населения, гражданские, на которых направлен теракт	VAL
IS-A	HUMAN-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
AGE	>0	VAL
GENDER	MALE, FEMALE	VAL
HAS-RELIGION	RELIGION	SEM
HAS-NATIONALITY	NATION	SEM
DEFAULT-NAME-RU	мирные жители	VAL
DEFAULT-NAME-EN	civilians	VAL
DEFAULT-NAME-FR	civils	VAL

22. Фрейм концепта CLAIM-RESPONSIBILITY

LABEL	CLAIM-RESPONSIBILITY	VAL
DEFINITION	Факт взятия террористом или террористической организацией ответственности за теракт; не всегда означает, что теракт действительно совершен этим человеком или организацией	VAL
IS-A	PHYSICAL-EVENT	VAL
PRECONDITION	TERROR-ATTACK	SEM
AGENT	TERRORISM-AGENT, TERRORIST-ORGANIZATION	SEM
NEGATION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	взять на себя ответственность	VAL

DEFAULT-NAME-EN	claim responsibility	VAL
DEFAULT-NAME-FR	revendiquer	VAL

23. Фрейм концепта CONSEQUENCES

LABEL	CONSEQUENCES	VAL
DEFINITION	Последствия террористического акта	VAL
IS-A	PHYSICAL-EVENT	VAL
SUBCLASSES	CONSEQUENCES-FOR-PEOPLE, CONSEQUENCES-FOR-TERRORIST, DAMAGE	VAL
THEME	AFFECTED-TERRORIST, AFFECTED-OBJECT, VICTIM	SEM
PRECONDITION	TERROR-ATTACK	SEM
NEGATION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	последствия	VAL
DEFAULT-NAME-EN	consequences	VAL
DEFAULT-NAME-FR	conséquences	VAL

24. Фрейм концепта CONSEQUENCES-FOR-PEOPLE

LABEL	CONSEQUENCES-FOR-PEOPLE	VAL
DEFINITION	Положительные и отрицательные последствия террористического акта для его жертв: погибшие, раненые, выжившие и т. п.	VAL
IS-A	CONSEQUENCES	VAL
SUBCLASSES	PEOPLE-DYING, PEOPLE-INJURING, TAKEN-HOSTAGE, VICTIM-SURVIVAL	VAL
THEME	VICTIM	SEM
NEGATION	TRUE, FALSE	VAL
PRECONDITION	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	последствия для людей	VAL
DEFAULT-NAME-EN	consequences for people	VAL
DEFAULT-NAME-FR	conséquences pour les personnes	VAL

25. Фрейм концепта CONSEQUENCES-FOR-TERRORIST

LABEL	CONSEQUENCES-FOR-TERRORIST	VAL
DEFINITION	Последствия террористического акта для террористов	VAL
IS-A	CONSEQUENCES	VAL
SUBCLASSES	TERRORIST-ELIMINATION, TERRORIST-INJURING, TERRORIST-SUICIDE, TERRORIST-CAPTURING, TERRORIST-FLEEING	VAL
THEME	AFFECTED-TERRORIST	SEM
NEGATION	TRUE, FALSE	VAL
PRECONDITION	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	последствия для террористов	VAL
DEFAULT-NAME-EN	consequences for terrorists	VAL

DEFAULT-NAME-FR	conséquences pour les terroristes	VAL
-----------------	-----------------------------------	-----

26. Фрейм концепта COUNTERTERRORISM-AGENT

LABEL	COUNTERTERRORISM-AGENT	VAL
DEFINITION	Лица и организации, занимающиеся контртеррористической деятельностью	VAL
IS-A	HUMAN	VAL
AGE	>0	VAL
GENDER	MALE, FEMALE	VAL
HAS-RELIGION	*nothing*	VAL
HAS-NATIONALITY	NATION	SEM
AGENT-OF	COUNTERTERRORISM-EVENT	SEM
DEFAULT-NAME-RU	представитель контртеррористической деятельности	VAL
DEFAULT-NAME-EN	counterterrorism agent	VAL
DEFAULT-NAME-FR	agent antiterroriste	VAL

27. Фрейм концепта COUNTERTERRORISM-EVENT

LABEL	COUNTERTERRORISM-EVENT	VAL
DEFINITION	Меры по профилактике терроризма, борьбе с терроризмом и контрмеры в ответ на террористические действия	VAL
IS-A	SOCIAL-EVENT	VAL
AGENT	COUNTERTERRORISM-AGENT	SEM
DEFAULT-NAME-RU	контртеррористические мероприятия	VAL
DEFAULT-NAME-EN	counterterrorism events	VAL
DEFAULT-NAME-FR	événements antiterroristes	VAL

28. Фрейм концепта COUNTERTERRORISM-INSTRUMENT

LABEL	COUNTERTERRORISM-INSTRUMENT	VAL
DEFINITION	Материальные средства, используемые представителями контртеррористической деятельности для борьбы с терроризмом (например, оружие)	VAL
IS-A	ARTIFACT	VAL
DEFAULT-NAME-RU	средство контртеррористической деятельности	VAL
DEFAULT-NAME-EN	means of counterterrorism	VAL
DEFAULT-NAME-FR	moyens de lutte contre le terrorisme	VAL

29. Фрейм концепта COUNTRY

LABEL	COUNTRY	VAL
DEFINITION	Страна, в которой произошел теракт	VAL
IS-A	LOCATION	VAL
HAS-PART	COUNTRY-AREA, CITY, CITY-AREA, SPECIFIC-PLACE	SEM
HAS-CAPITAL	CAPITAL-CITY	SEM

LOCATION-OF	TERROR-ATTACK	SEM
INSTANCES	COUNTRY -1, COUNTRY -2 ...	VAL
DEFAULT-NAME-RU	регион страны	VAL
DEFAULT-NAME-EN	country region	VAL
DEFAULT-NAME-FR	région du pays	VAL

30. Фрейм концепта COUNTRY-AREA

LABEL	COUNTRY-AREA	VAL
DEFINITION	Часть страны, в которой произошел теракт	VAL
IS-A	LOCATION	VAL
PART-OF	COUNTRY	SEM
HAS-PART	CITY, CITY-AREA, SPECIFIC-PLACE	SEM
LOCATION-OF	TERROR-ATTACK	SEM
INSTANCES	COUNTRY-AREA-1, COUNTRY-AREA-2 ...	VAL
DEFAULT-NAME-RU	регион страны	VAL
DEFAULT-NAME-EN	country region	VAL
DEFAULT-NAME-FR	région du pays	VAL

31. Фрейм концепта CYBER-ATTACK

LABEL	CYBER-ATTACK	VAL
DEFINITION	Массовый взлом компьютерных сетей, массовое заражение компьютеров вирусами и т. д. в террористических целях	VAL
IS-A	TERROR-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	PROPERTY-TARGET, OBJECT-TARGET	SEM
LOCATED	LOCATION	SEM
INSTRUMENT	*nothing*	VAL
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	кибератака	VAL
DEFAULT-NAME-EN	cyber attack	VAL
DEFAULT-NAME-FR	attaque cybernétique	VAL

32. Фрейм концепта DAMAGE

LABEL	DAMAGE	VAL
-------	--------	-----

DEFINITION	Материальный ущерб, причиненный неодушевленным объектам в результате теракта	VAL
IS-A	CONSEQUENCES	VAL
THEME	AFFECTED-OBJECT	SEM
PRECONDITION	TERROR-ATTACK	SEM
NEGATION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	материальный ущерб	VAL
DEFAULT-NAME-EN	physical damage	VAL
DEFAULT-NAME-FR	dommages physiques	VAL

33. Фрейм концепта DECLARE

LABEL	DECLARE	VAL
DEFINITION	Публично, официально сообщить о чем-либо	VAL
IS-A	PHYSICAL-EVENT	VAL
AGENT	SOURCE	SEM
NEGATION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	заявить	VAL
DEFAULT-NAME-EN	declare	VAL
DEFAULT-NAME-FR	déclarer	VAL

34. Фрейм концепта DISTANT-ATTACK

LABEL	DISTANT-ATTACK	VAL
DEFINITION	Теракт, совершенный с большого расстояния посредством оружия дальнего боя	VAL
IS-A	TERROR-ATTACK	VAL
SUBCLASSES	DRONE-ATTACK, MORTAR-ATTACK, ROCKET-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, PROPERTY-TARGET, OBJECT-TARGET	SEM
INSTRUMENT	WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	дистанционная атака	VAL
DEFAULT-NAME-EN	distant attack	VAL
DEFAULT-NAME-FR	attaque à distance	VAL

35. Фрейм концепта DRONE-ATTACK

LABEL	DRONE-ATTACK	VAL
DEFINITION	Атака с использованием беспилотного летательного аппарата	VAL
IS-A	DISTANT-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, PROPERTY-TARGET, OBJECT-TARGET	SEM
INSTRUMENT	WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	атака БПЛ	VAL
DEFAULT-NAME-EN	drone attack	VAL
DEFAULT-NAME-FR	attaque par drone	VAL

36. Фрейм концепта EDUCATIONAL-INSTITUTION

LABEL	EDUCATIONAL-INSTITUTION	VAL
DEFINITION	Учреждение, в котором происходит обучение, например школа или университет	VAL
IS-A	PROPERTY-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	образовательное учреждение	VAL
DEFAULT-NAME-EN	educational institution	VAL
DEFAULT-NAME-FR	établissement d'enseignement	VAL

37. Фрейм концепта ENTERTAINMENT-FACILITY

LABEL	ENTERTAINMENT-FACILITY	VAL
DEFINITION	Учреждение, в котором проводятся различные развлекательные мероприятия, например ТРЦ или клуб	VAL
IS-A	PROPERTY-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	развлекательное учреждение	VAL
DEFAULT-NAME-EN	entertainment facility	VAL
DEFAULT-NAME-FR	établissement de divertissement	VAL

38. Фрейм концепта EXACT-DATE

LABEL	EXACT-DATE	VAL
DEFINITION	Точная дата совершения теракта	VAL
IS-A	TIME	VAL
DEFAULT-NAME-RU	точная дата	VAL
DEFAULT-NAME-EN	exact date	VAL
DEFAULT-NAME-FR	date exacte	VAL

39. Фрейм концепта EXACT-TIME

LABEL	EXACT-TIME	VAL
DEFINITION	Точное время совершения теракта	VAL
IS-A	TIME	VAL
DEFAULT-NAME-RU	точное время	VAL
DEFAULT-NAME-EN	exact time	VAL
DEFAULT-NAME-FR	heure exacte	VAL

40. Фрейм концепта EXPLOSIVE-DEVICE

LABEL	EXPLOSIVE-DEVICE	VAL
DEFINITION	Предмет, снабженный разрывным зарядом и взрываемый с помощью взрывателя, путем удара или иным образом	VAL
IS-A	WEAPON	VAL
SUBCLASSES	SUICIDE-VEST, CAR-BOMB	VAL
INSTRUMENT-OF	BOMB-ATTACK	SEM
WEAPON-OF	TERRORISM-AGENT, ARTIFACT	SEM
DEFAULT-NAME-RU	бомба	VAL
DEFAULT-NAME-EN	bomb	VAL
DEFAULT-NAME-FR	bombe	VAL

41. Фрейм концепта GOAL-OF-ATTACK

LABEL	GOAL-OF-ATTACK	VAL
DEFINITION	Цель, которую террористы планируют достичь в результате теракта, либо причина, по которой теракт совершен	VAL
IS-A	MENTAL-OBJECT	VAL
SUBCLASSES	HARM-PEOPLE, MILITARY-GOAL, POLITICAL-GOAL, PROPERTY-DAMAGE, RELIGIOUS-GOAL	VAL
GOAL-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	цель теракта	VAL
DEFAULT-NAME-EN	goal of attack	VAL
DEFAULT-NAME-FR	motif de l'attaque	VAL

42. Фрейм концепта GOVERNMENTAL-INSTITUTION

LABEL	GOVERNMENTAL-INSTITUTION	VAL
-------	--------------------------	-----

DEFINITION	Государственное учреждение, на которое направлен теракт	VAL
IS-A	PROPERTY-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	государственное учреждение	VAL
DEFAULT-NAME-EN	governmental institution	VAL
DEFAULT-NAME-FR	institution gouvernementale	VAL

43. Фрейм концепта GUN

LABEL	GUN	VAL
DEFINITION	Стрелковое оружие, из которого выстреливается пороховой снаряд	VAL
IS-A	WEAPON	VAL
INSTRUMENT-OF	GUN-ATTACK	SEM
WEAPON-OF	TERRORISM-AGENT	SEM
DEFAULT-NAME-RU	огнестрельное оружие	VAL
DEFAULT-NAME-EN	gun	VAL
DEFAULT-NAME-FR	arme à feu	VAL

44. Фрейм концепта GUN-ATTACK

LABEL	GUN-ATTACK	VAL
DEFINITION	Теракт с использованием огнестрельного оружия	VAL
IS-A	ARMED-ASSAULT	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET	SEM
INSTRUMENT	GUN	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	стрельба	VAL
DEFAULT-NAME-EN	gun attack	VAL
DEFAULT-NAME-FR	fusillade	VAL

45. Фрейм концепта HARM-PEOPLE

LABEL	HARM-PEOPLE	VAL
DEFINITION	Нанести вред людям	VAL

IS-A	GOAL-OF-ATTACK	VAL
GOAL-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	вред людям	VAL
DEFAULT-NAME-EN	harm to people	VAL
DEFAULT-NAME-FR	dommage aux personnes	VAL

46. Фрейм концепта HIJACKING

LABEL	HIJACKING	VAL
DEFINITION	Взять под контроль самолет или другое транспортное средство во время путешествия, особенно с применением насилия	VAL
IS-A	TERROR-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, OBJECT-TARGET	SEM
INSTRUMENT	WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	угон	VAL
DEFAULT-NAME-EN	hijacking	VAL
DEFAULT-NAME-FR	détournement	VAL

47. Фрейм концепта HOSTAGE-TAKING

LABEL	HOSTAGE-TAKING	VAL
DEFINITION	Теракт, суть которого заключается в захвате людей в качестве заложников в каком-либо здании или транспортном средстве для достижения каких-либо политических или идеологических целей или получения выкупа	VAL
IS-A	TERROR-ATTACK	VAL
SUBCLASSES	BARRICADING, KIDNAPPING	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, PROPERTY-TARGET	SEM
INSTRUMENT	WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-	TERRORIST-ORGANIZATION	DEF

CLAIMED-BY	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	заложники	VAL
DEFAULT-NAME-EN	hostages	VAL
DEFAULT-NAME-FR	otages	VAL

48. Фрейм концепта HUMAN-TARGET

LABEL	HUMAN-TARGET	VAL
DEFINITION	Лица, являющиеся объектом (целью) террористического акта	VAL
IS-A	HUMAN	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
AGE	>0	VAL
GENDER	MALE, FEMALE	VAL
HAS-RELIGION	RELIGION	SEM
HAS-NATIONALITY	NATION	SEM
DEFAULT-NAME-RU	человек-объект	VAL
DEFAULT-NAME-EN	human target	VAL
DEFAULT-NAME-FR	cible humaine	VAL

49. Фрейм концепта IDEOLOGY

LABEL	IDEOLOGY	VAL
DEFINITION	Идеология, которой следует террорист или террористическая организация	VAL
IS-A	SOCIAL-OBJECT	VAL
SUBCLASSES	ISLAMISM, RIGHT-WING-EXTREMISM, LEFT-WING-EXTREMISM, WHITE-SUPREMACISM, NATIONALISM, SEPARATISM	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	идеология	VAL
DEFAULT-NAME-EN	ideology	VAL
DEFAULT-NAME-FR	idéologie	VAL

50. Фрейм концепта INCENDIARY-WEAPON

LABEL	INCENDIARY-WEAPON	VAL
DEFINITION	Оружие, поражающее действие которого основано на использовании боевых зажигательных веществ	VAL
IS-A	WEAPON	VAL
INSTRUMENT-OF	ARSON	SEM
WEAPON-OF	TERRORISM-AGENT, ARTIFACT	SEM

DEFAULT-NAME-RU	зажигательное оружие	VAL
DEFAULT-NAME-EN	incendiary weapon	VAL
DEFAULT-NAME-FR	arme incendiaire	VAL

51. Фрейм концепта INDIVIDUAL-TERRORIST

LABEL	INDIVIDUAL-TERRORIST	VAL
DEFINITION	Террорист, действующий в одиночку, не являющийся членом какой-либо террористической организации	VAL
IS-A	TERRORISM-AGENT	VAL
AGE	>0	VAL
GENDER	MALE, FEMALE	VAL
HAS-IDEOLOGY	IDEOLOGY	SEM
HAS-RELIGION	RELIGION	SEM
HAS-NATIONALITY	NATION	SEM
ASSUMPTION	TRUE, FALSE	VAL
AGENT-OF	TERROR-ATTACK, TERROR-ATTACK-PLANNING-EVENT, TERRORIST-THREAT, OTHER-TERRORIST-EVENT	VAL
CLAIMED-RESPONSIBILITY-FOR	TERROR-ATTACK	VAL
MEMBER-OF	*nothing*	VAL
DEFAULT-NAME-RU	террорист-одиночка	VAL
DEFAULT-NAME-EN	lone wolf	VAL
DEFAULT-NAME-FR	loup solitaire	VAL

52. Фрейм концепта ISLAM

LABEL	ISLAM	VAL
DEFINITION	Одна из наиболее распространенных религий мира, последователями которой являются мусульмане	VAL
IS-A	RELIGION	VAL
DEFAULT-NAME-RU	ислам	VAL
DEFAULT-NAME-EN	Islam	VAL
DEFAULT-NAME-FR	islam	VAL

53. Фрейм концепта ISLAMISM

LABEL	ISLAMISM	VAL
DEFINITION	Политическая и социальная идеология, часто радикальная, направленная на насильственное распространение строгого соблюдения требований ислама	VAL
IS-A	IDEOLOGY	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	исламизм	VAL
DEFAULT-NAME-EN	Islamism	VAL
DEFAULT-NAME-FR	islamisme	VAL

54. Фрейм концепта KIDNAPPING

LABEL	KIDNAPPING	VAL
DEFINITION	Похищение людей с целью получения выкупа	VAL
DEFAULT-NAME-RU	киднеппинг	VAL
DEFAULT-NAME-EN	kidnapping	VAL
DEFAULT-NAME-FR	enlèvement	VAL
IS-A	HOSTAGE-TAKING	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET	SEM
INSTRUMENT	WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL

55. Фрейм концепта KNIFE

LABEL	KNIFE	VAL
DEFINITION	Холодное оружие, т. е. оружие с лезвием, например нож, меч, мачете и т. д.	VAL
IS-A	WEAPON	VAL
INSTRUMENT-OF	KNIFE-ATTACK	SEM
WEAPON-OF	TERRORISM-AGENT	SEM
DEFAULT-NAME-RU	холодное оружие	VAL
DEFAULT-NAME-EN	knife	VAL
DEFAULT-NAME-FR	couteau	VAL

56. Фрейм концепта KNIFE-ATTACK

LABEL	KNIFE-ATTACK	VAL
DEFINITION	Теракт с применением холодного оружия	VAL
IS-A	ARMED-ASSAULT	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET	SEM
INSTRUMENT	KNIFE	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM

RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
NEGATION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	нападение с ножом	VAL
DEFAULT-NAME-EN	knife attack	VAL
DEFAULT-NAME-FR	attaque au couteau	VAL

57. Фрейм концепта LEFT-WING-TERRORISM

LABEL	LEFT-WING-TERRORISM	VAL
DEFINITION	Терроризм, используемый леворадикальными организациями	VAL
IS-A	IDEOLOGY	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	левый терроризм	VAL
DEFAULT-NAME-EN	left-wing terrorism	VAL
DEFAULT-NAME-FR	terrorisme de gauche	VAL

58. Фрейм концепта LOCATION

LABEL	LOCATION	VAL
DEFINITION	Место, где произошел террористический акт	VAL
IS-A	PHYSICAL-OBJECT	VAL
SUBCLASSES	CITY, CITY-AREA, COUNTRY, COUNTRY-AREA, SPECIFIC-LOCATION, TERRITORY	VAL
LOCATION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	место теракта	VAL
DEFAULT-NAME-EN	location	VAL
DEFAULT-NAME-FR	emplacement	VAL

59. Фрейм концепта MEDICAL-INSTITUTION

LABEL	MEDICAL-INSTITUTION	VAL
DEFINITION	Медицинское учреждение, на которое направлен террористический акт, например госпиталь, больница	VAL
IS-A	PROPERTY-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	медицинское учреждение	VAL
DEFAULT-NAME-EN	medical institution	VAL
DEFAULT-NAME-FR	institution médicale	VAL

60. Фрейм концепта MILITARY-GOAL

LABEL	MILITARY-GOAL	VAL
DEFINITION	Цель в рамках военных действий, которую стремятся достичь террористы посредством теракта	VAL
IS-A	GOAL-OF-ATTACK	VAL
GOAL-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	военная цель	VAL
DEFAULT-NAME-EN	military goal	VAL
DEFAULT-NAME-FR	objectif militaire	VAL

61. Фрейм концепта MILITARY-INSTITUTION

LABEL	MILITARY-INSTITUTION	VAL
DEFINITION	Военное учреждение, на которое направлен теракт	VAL
IS-A	PROPERTY-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	военное учреждение	VAL
DEFAULT-NAME-EN	military institution	VAL
DEFAULT-NAME-FR	institution militaire	VAL

62. Фрейм концепта MILITARY-TARGET

LABEL	MILITARY-TARGET	VAL
DEFINITION	Военные, которые являются объектом теракта	VAL
IS-A	HUMAN-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
AGE	>0	VAL
GENDER	MALE, FEMALE	VAL
HAS-RELIGION	RELIGION	SEM
HAS-NATIONALITY	NATION	SEM
DEFAULT-NAME-RU	военные	VAL
DEFAULT-NAME-EN	military target	VAL
DEFAULT-NAME-FR	cible militaire	VAL

63. Фрейм концепта MORTAR-ATTACK

LABEL	MORTAR-ATTACK	VAL
DEFINITION	Обстрел террористами целей с использованием минометов	VAL
IS-A	DISTANT-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	PROPERTY-TARGET	SEM
INSTRUMENT	WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM

PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	минометный обстрел	VAL
DEFAULT-NAME-EN	mortar attack	VAL
DEFAULT-NAME-FR	attaque au mortier	VAL

64. Фрейм концепта NATION

LABEL	NATION	VAL
DEFINITION	Совокупность людей, организованная под единым правительством; отдельная страна, рассматриваемая вместе с ее общественно-политическими структурами	VAL
IS-A	GEOPOLITICAL-ENTITY	VAL
INSTANCES	NATION-1, NATION-2, NATION-3, NATION-4 ...	VAL
HAS-CAPITAL	CAPITAL-CITY	SEM
DEFAULT-NAME-RU	страна	VAL
DEFAULT-NAME-EN	nation	VAL
DEFAULT-NAME-FR	nation	VAL

65. Фрейм концепта NATIONALISM

LABEL	NATIONALISM	VAL
DEFINITION	Идеология и политика, направленная на разжигание национальной вражды и принижение других наций	VAL
IS-A	IDEOLOGY	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	национализм	VAL
DEFAULT-NAME-EN	nationalism	VAL
DEFAULT-NAME-FR	nationalisme	VAL

66. Фрейм концепта OBJECT-TARGET

LABEL	OBJECT-TARGET	VAL
DEFINITION	Предметы, на которые направлен террористический акт и которые могут быть отнесены к движимому имуществу, например автомобили	VAL
IS-A	ARTIFACT	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	объект-цель	VAL
DEFAULT-NAME-EN	object target	VAL
DEFAULT-NAME-FR	objet cible	VAL

67. Фрейм концепта OTHER-TERRORIST-EVENT

LABEL	OTHER-TERRORIST-EVENT	VAL
DEFINITION	Прочая террористическая деятельность, отличная от планирования и совершения терактов, например финансирование терроризма, вербовка новых террористов и т. п.	VAL
IS-A	CRIMINAL-ACTIVITY	VAL
SUBCLASSES	TERRORISM-FINANCING, TERRORIST-RECRUITING, WAR-RELATED-TERRORIST-EVENT	VAL
AGENT	TERRORISM-AGENT, TERRORIST-ORGANIZATION	SEM
DEFAULT-NAME-RU	прочая террористическая деятельность	VAL
DEFAULT-NAME-EN	other terrorist activity	VAL
DEFAULT-NAME-FR	autre activité terroriste	VAL

68. Фрейм концепта PEOPLE-DYING

LABEL	PEOPLE-DYING	VAL
DEFINITION	Гибель людей (кроме террористов) в результате теракта	VAL
IS-A	CONSEQUENCES-FOR-PEOPLE	VAL
THEME	VICTIM	SEM
NEGATION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	погибшие	VAL
DEFAULT-NAME-EN	killed	VAL
DEFAULT-NAME-FR	morts	VAL

69. Фрейм концепта PEOPLE-INJURING

LABEL	PEOPLE-INJURING	VAL
DEFINITION	Травмы людей (кроме террористов) в результате теракта	VAL
IS-A	CONSEQUENCES-FOR-PEOPLE	VAL
THEME	VICTIM	SEM
NEGATION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	раненые	VAL
DEFAULT-NAME-EN	injured	VAL
DEFAULT-NAME-FR	blessés	VAL

70. Фрейм концепта POLICE-FACILITY

LABEL	POLICE-FACILITY	VAL
DEFINITION	Полицейское учреждение, на которое направлен террористический акт, например полицейский участок	VAL
IS-A	PROPERTY-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	полицейское учреждение	VAL
DEFAULT-NAME-EN	police facility	VAL
DEFAULT-NAME-FR	établissement de police	VAL

71. Фрейм концепта POLICE-TARGET

LABEL	POLICE-TARGET	VAL
DEFINITION	Сотрудники полиции, на которых направлен теракт	VAL
IS-A	HUMAN-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
AGE	>0	VAL
GENDER	MALE, FEMALE	VAL
HAS-RELIGION	RELIGION	SEM
HAS-NATIONALITY	NATION	SEM
DEFAULT-NAME-RU	полицейские	VAL
DEFAULT-NAME-EN	police target	VAL
DEFAULT-NAME-FR	cible policière	VAL

72. Фрейм концепта POLITICAL-GOAL

LABEL	POLITICAL-GOAL	VAL
DEFINITION	Политическая цель, которую намерены достичь террористы в результате теракта	VAL
IS-A	GOAL-OF-ATTACK	VAL
GOAL-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	политическая цель	VAL
DEFAULT-NAME-EN	political goal	VAL
DEFAULT-NAME-FR	objectif politique	VAL

73. Фрейм концепта POLITICAL-TARGET

LABEL	POLITICAL-TARGET	VAL
DEFINITION	Политический деятель, на которого направлен теракт	VAL
IS-A	HUMAN-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
AGE	>0	VAL
GENDER	MALE, FEMALE	VAL
HAS-RELIGION	RELIGION	SEM
HAS-NATIONALITY	NATION	SEM
DEFAULT-NAME-RU	политический деятель	VAL
DEFAULT-NAME-EN	political target	VAL
DEFAULT-NAME-FR	cible politique	VAL

74. Фрейм концепта PROPERTY-DAMAGE

LABEL	PROPERTY-DAMAGE	VAL
DEFINITION	Нанесение вреда имуществу	VAL
IS-A	GOAL-OF-ATTACK	VAL
GOAL-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	вред имуществу	VAL

DEFAULT-NAME-EN	property damage	VAL
DEFAULT-NAME-FR	dommages matériels	VAL

75. Фрейм концепта PROPERTY-TARGET

LABEL	PROPERTY-TARGET	VAL
DEFINITION	Объекты инфраструктуры и частная собственность, на которые направлен террористический акт	VAL
IS-A	ARTIFACT	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	недвижимое имущество	VAL
DEFAULT-NAME-EN	property target	VAL
DEFAULT-NAME-FR	immobilier ciblé	VAL

76. Фрейм концепта RAMMING-VEHICLE

LABEL	RAMMING-VEHICLE	VAL
DEFINITION	Транспортное средство, используемое террористами в качестве тарана для атаки на толпу людей	VAL
IS-A	WEAPON	VAL
INSTRUMENT-OF	VEHICLE-RAMMING-ATTACK	SEM
WEAPON-OF	INDIVIDUAL-TERRORIST	
DEFAULT-NAME-RU	таран	VAL
DEFAULT-NAME-EN	ramming vehicle	VAL
DEFAULT-NAME-FR	véhicule bélier	VAL

77. Фрейм концепта RELIGION

LABEL	RELIGION	VAL
DEFINITION	Мировоззрение, основанное на вере в существование бога, сверхъестественных сил	VAL
IS-A	SOCIAL-OBJECT	VAL
SUBCLASSES	ISLAM, CHRISTIANITY	VAL
DEFAULT-NAME-RU	религия	VAL
DEFAULT-NAME-EN	religion	VAL
DEFAULT-NAME-FR	religion	VAL

78. Фрейм концепта RELIGIOUS-GOAL

LABEL	RELIGIOUS-GOAL	VAL
DEFINITION	Религиозная цель, которую намерены достичь террористы в результате теракта	VAL
IS-A	GOAL-OF-ATTACK	VAL
GOAL-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	религиозная цель	VAL
DEFAULT-NAME-EN	religious goal	VAL

DEFAULT-NAME-FR	objectif religieux	VAL
-----------------	--------------------	-----

79. Фрейм концепта RELIGIOUS-INSTITUTION

LABEL	RELIGIOUS-INSTITUTION	VAL
DEFINITION	Религиозное учреждение, на которое направлен теракт, например церковь	VAL
IS-A	PROPERTY-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	религиозное учреждение	VAL
DEFAULT-NAME-EN	religious institution	VAL
DEFAULT-NAME-FR	institution religieuse	VAL

80. Фрейм концепта RELIGIOUS-TARGET

LABEL	RELIGIOUS-TARGET	VAL
DEFINITION	Религиозный деятель, на которого направлен теракт	VAL
IS-A	HUMAN-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
AGE	>0	VAL
GENDER	MALE, FEMALE	VAL
HAS-RELIGION	RELIGION	SEM
HAS-NATIONALITY	NATION	SEM
DEFAULT-NAME-RU	религиозный деятель	VAL
DEFAULT-NAME-EN	religious target	VAL
DEFAULT-NAME-FR	cible religieuse	VAL

81. Фрейм концепта RIGHT-WING-TERRORISM

LABEL	RIGHT-WING-TERRORISM	VAL
DEFINITION	Терроризм, применяемый праворадикальными организациями	VAL
IS-A	IDEOLOGY	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	правый терроризм	VAL
DEFAULT-NAME-EN	right-wing terrorism	VAL
DEFAULT-NAME-FR	terrorisme de droite	VAL

82. Фрейм концепта ROCKET

LABEL	ROCKET	VAL
DEFINITION	Оружие дальнего боя, в котором средства поражения доставляются к цели с помощью ракет	VAL
IS-A	WEAPON	VAL
INSTRUMENT-OF	ROCKET-ATTACK	SEM
WEAPON-OF	TERRORISM-AGENT, ARTIFACT	SEM
DEFAULT-NAME-RU	ракетное оружие	VAL

DEFAULT-NAME-EN	rocket	VAL
DEFAULT-NAME-FR	missile	VAL

83. Фрейм концепта ROCKET-ATTACK

LABEL	ROCKET-ATTACK	VAL
DEFINITION	Теракт с использование ракетного оружия	VAL
IS-A	DISTANT-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	PROPERTY-TARGET	SEM
INSTRUMENT	ROCKET	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	ракетный обстрел	VAL
DEFAULT-NAME-EN	rocket attack	VAL
DEFAULT-NAME-FR	attaque missile	VAL

84. Фрейм концепта SEPARATISM

LABEL	SEPARATISM	VAL
DEFINITION	Идеология, призывающая к отделению части территории от государства	VAL
IS-A	IDEOLOGY	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	сепаратизм	VAL
DEFAULT-NAME-EN	separatism	VAL
DEFAULT-NAME-FR	séparatisme	VAL

85. Фрейм концепта SOURCE

LABEL	SOURCE	VAL
DEFINITION	Источник сообщения о террористическом акте и его последствиях: газеты, телеканалы, власти, свидетели	VAL
IS-A	SOCIAL-OBJECT	VAL
INSTANCES	SOURCE-1, SOURCE-2, SOURCE-3 ...	VAL
HAS-NATIONALITY	NATION	SEM
DEFAULT-NAME-RU	источник	VAL

DEFAULT-NAME-EN	source	VAL
DEFAULT-NAME-FR	source	VAL

86. Фрейм концепта SPECIFIC-LOCATION

LABEL	SPECIFIC-LOCATION	VAL
DEFINITION	Конкретное место совершения теракта	VAL
IS-A	LOCATION	VAL
PART-OF	CITY-AREA, CITY, CAPITAL-CITY, COUNTRY-AREA, COUNTRY, TERRITORY	SEM
LOCATION-OF	TERROR-ATTACK	SEM
INSTANCES	SPECIFIC-LOCATION-1, SPECIFIC-LOCATION-2 ...	VAL
DEFAULT-NAME-RU	конкретное место	VAL
DEFAULT-NAME-EN	specific location	VAL
DEFAULT-NAME-FR	endroit précis	VAL

87. Фрейм концепта SUICIDE-ATTACK

LABEL	SUICIDE-ATTACK	VAL
DEFINITION	Теракт, совершаемый террористом или группой террористов, в результате которого они погибают	VAL
IS-A	TERROR-ATTACK	VAL
SUBCLASSES	SUICIDE-BOMB-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, PROPERTY-TARGET, OBJECT-TARGET	SEM
INSTRUMENT	WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	теракт-самоубийство	VAL
DEFAULT-NAME-EN	suicide attack	VAL
DEFAULT-NAME-FR	attentat suicide	VAL

88. Фрейм концепта SUICIDE-BOMB-ATTACK

LABEL	SUICIDE-BOMB-ATTACK	VAL
-------	---------------------	-----

DEFINITION	Теракт, совершаемый террористом или группой террористов с использованием взрывного устройства, в результате которого они погибают	VAL
IS-A	SUICIDE-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, PROPERTY-TARGET, OBJECT-TARGET	SEM
INSTRUMENT	SUICIDE-VEST	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	самоподрыв	VAL
DEFAULT-NAME-EN	suicide bomb attack	VAL
DEFAULT-NAME-FR	attentat suicide à la bombe	VAL

89. Фрейм концепта SUICIDE-VEST

LABEL	SUICIDE-VEST	VAL
DEFINITION	Самодельное взрывное устройство, которое террорист-смертник надевает на тело под одежду для совершения убийства путем самоподрыва	VAL
IS-A	EXPLOSIVE DEVICE	VAL
INSTRUMENT-OF	SUICIDE-BOMB-ATTACK	SEM
WEAPON-OF	INDIVIDUAL-TERRORIST	SEM
DEFAULT-NAME-RU	пояс смертника	VAL
DEFAULT-NAME-EN	suicide vest	VAL
DEFAULT-NAME-FR	ceinture explosive	VAL

90. Фрейм концепта TAKEN-HOSTAGE

LABEL	TAKEN-HOSTAGE	VAL
DEFINITION	В результате теракта люди взяты в заложники	VAL
IS-A	CONSEQUENCES-FOR-PEOPLE	VAL
THEME	VICTIM	SEM
PRECONDITION	HOSTAGE-TAKING	SEM
DEFAULT-NAME-RU	заложники	VAL
DEFAULT-NAME-EN	hostages	VAL
DEFAULT-NAME-FR	otages	VAL

91. Фрейм концепта TERRITORY

LABEL	TERRITORY	VAL
DEFINITION	Территория, не являющаяся признанным государством, на которой произошел теракт	VAL
IS-A	LOCATION	VAL
HAS-CAPITAL	CAPITAL-CITY	SEM
LOCATION-OF	TERROR-ATTACK	SEM
INSTANCES	TERRITORY-1, TERRITORY-2, TERRITORY-3 ...	VAL
DEFAULT-NAME-RU	территория	VAL
DEFAULT-NAME-EN	territory	VAL
DEFAULT-NAME-FR	territoire	VAL

92. Фрейм концепта TERROR-ATTACK

LABEL	TERROR-ATTACK	VAL
DEFINITION	Нападение, совершаемое террористом или группой террористов для устрашения населения для достижения политических или иных целей	VAL
IS-A	CRIMINAL-ACTIVITY	VAL
SUBCLASSES	ARMED-ASSAULT, ARSON, BIOLOGICAL-ATTACK, BOMB-ATTACK, CHEMICAL-ATTACK, DISTANT-ATTACK, HIJACKING, HOSTAGE-TAKING, ASSASSINATION, SUICIDE-ATTACK, UNARMED-ASSAULT, VEHICLE-RAMMING-ATTACK, CYBER-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, PROPERTY-TARGET, OBJECT-TARGET	SEM
INSTRUMENT	WEAPON	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	ARMED-ASSAULT, ARSON, BIOLOGICAL-ATTACK, BOMB-ATTACK, CHEMICAL-ATTACK, DISTANT-ATTACK, HIJACKING, HOSTAGE-TAKING, ASSASSINATION, SUICIDE-ATTACK, UNARMED-ASSAULT, VEHICLE-RAMMING-ATTACK, CYBER-ATTACK	SEM
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	теракт	VAL
DEFAULT-NAME-EN	terror attack	VAL
DEFAULT-NAME-FR	attentat	VAL

93. Фрейм концепта TERROR-ATTACK-PLANNING-EVENT

LABEL	TERROR-ATTACK-PLANNING-EVENT	VAL
DEFINITION	Действия по планированию террористического акта	VAL
IS-A	CRIMINAL-ACTIVITY	VAL
AGENT	TERRORISM-AGENT, TERRORIST-ORGANIZATION	SEM
PRECONDITION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	планирование теракта	VAL
DEFAULT-NAME-EN	terror attack planning	VAL
DEFAULT-NAME-FR	planification d'un attentat	VAL

94. Фрейм концепта TERRORISM-AGENT

LABEL	TERRORISM-AGENT	VAL
DEFINITION	Лицо или группа лиц, занимающиеся террористической деятельностью либо в составе террористической организации, либо самостоятельно и использующие методы террора для достижения своих целей	VAL
IS-A	HUMAN	VAL
SUBCLASSES	INDIVIDUAL-TERRORIST, TERRORIST-GROUP	VAL
AGENT-OF	TERROR-ATTACK, TERROR-ATTACK-PLANNING-EVENT, OTHER-TERRORIST-EVENT	SEM
AGE	>0	VAL
GENDER	MALE, FEMALE	VAL
HAS-NATIONALITY	NATION	SEM
HAS-RELIGION	RELIGION	SEM
HAS-IDEOLOGY	IDEOLOGY	SEM
DEFAULT-NAME-RU	террорист	VAL
DEFAULT-NAME-EN	terrorist	VAL
DEFAULT-NAME-FR	terroriste	VAL

95. Фрейм концепта TERRORISM-FINANCING

LABEL	TERRORISM-FINANCING	VAL
DEFINITION	Финансирование террористической деятельности	VAL
IS-A	OTHER-TERRORIST-EVENT	VAL
AGENT	TERRORISM-AGENT, TERRORIST-ORGANIZATION	SEM
DEFAULT-NAME-RU	финансирование терроризма	VAL
DEFAULT-NAME-EN	terrorism financing	VAL
DEFAULT-NAME-FR	financement du terrorisme	VAL

96. Фрейм концепта TERRORIST-CAPTURING

LABEL	TERRORIST-CAPTURING	VAL
DEFINITION	Задержание (арест) террориста	VAL
IS-A	CONSEQUENCES-FOR-TERRORIST	VAL

THEME	AFFECTED-TERRORIST	SEM
DEFAULT-NAME-RU	арест террориста	VAL
DEFAULT-NAME-EN	terrorist arrest	VAL
DEFAULT-NAME-FR	arrestation d'un terroriste	VAL

97. Фрейм концепта TERRORIST-FLEEING

LABEL	TERRORIST-FLEEING	VAL
DEFINITION	Покидание террористом места теракта	VAL
IS-A	CONSEQUENCES-FOR-TERRORIST	VAL
THEME	AFFECTED-TERRORIST	SEM
DEFAULT-NAME-RU	бегство террориста	VAL
DEFAULT-NAME-EN	terrorist fleeing	VAL
DEFAULT-NAME-FR	en fuite	VAL

98. Фрейм концепта TERRORIST-GROUP

LABEL	TERRORIST-GROUP	VAL
DEFINITION	Террористы, совершающие теракты небольшой группой, которая может быть или не быть частью более крупной организации	VAL
IS-A	TERRORISM-AGENT	VAL
AGENT-OF	TERROR-ATTACK, TERROR-ATTACK-PLANNING-EVENT, TERRORIST-THREAT, OTHER-TERRORIST-EVENT	SEM
CLAIM-RESPONSIBILITY	TERROR-ATTACK	SEM
AGE	*nothing*	VAL
GENDER	MALE, FEMALE	VAL
HAS-NATIONALITY	NATION	SEM
HAS-RELIGION	RELIGION	SEM
HAS-IDEOLOGY	IDEOLOGY	SEM

99. Фрейм концепта TERRORIST-ELIMINATION

LABEL	TERRORIST-ELIMINATION	VAL
DEFINITION	Убийство (нейтрализация) террориста представителями контртеррористической деятельности	VAL
IS-A	CONSEQUENCES-FOR-TERRORIST	VAL
THEME	AFFECTED-TERRORIST	SEM
DEFAULT-NAME-RU	убийство террориста	VAL
DEFAULT-NAME-EN	terrorist elimination	VAL
DEFAULT-NAME-FR	assassinat d'un terroriste	VAL

100. Фрейм концепта TERRORIST-INJURING

LABEL	TERRORIST-INJURING	VAL
-------	--------------------	-----

DEFINITION	Ранение террориста представителями контртеррористической деятельности	VAL
IS-A	CONSEQUENCES-FOR-TERRORIST	VAL
THEME	AFFECTED-TERRORIST	SEM
DEFAULT-NAME-RU	ранение террориста	VAL
DEFAULT-NAME-EN	terrorist injuring	VAL
DEFAULT-NAME-FR	blessure d'un terroriste	VAL

101. Фрейм концепта TERRORIST-ORGANIZATION

LABEL	TERRORIST-ORGANIZATION	VAL
DEFINITION	Организация, совершающая террористические акты против людей или имущества с целью запугивания или принуждения общества или правительства к чему-либо, часто в идеологических или политических целях	VAL
IS-A	ORGANIZATION	VAL
HAS-MEMBER	TERRORIST-GROUP	SEM
AGENT-OF	TERROR-ATTACK, TERROR-ATTACK-PLANNING-EVENT, TERRORIST-THREAT, OTHER-TERRORIST-EVENT	SEM
CLAIM-RESPONSIBILITY	TERROR-ATTACK	SEM
HAS-IDEOLOGY	IDEOLOGY	SEM
INSTANCES	TERRORIST-ORGANIZATION-1, TERRORIST-ORGANIZATION-2, TERRORIST-ORGANIZATION-3 ...	VAL
DEFAULT-NAME-RU	террористическая организация	VAL
DEFAULT-NAME-EN	terrorist organization	VAL
DEFAULT-NAME-FR	organisation terroriste	VAL

102. Фрейм концепта TERRORIST-RECRUITING

LABEL	TERRORIST-RECRUITING	VAL
DEFINITION	Вербовка в террористы	VAL
IS-A	OTHER-TERRORIST-EVENT	VAL
AGENT	TERRORISM-AGENT	SEM
DEFAULT-NAME-RU	вербовка	VAL
DEFAULT-NAME-EN	recruitment	VAL
DEFAULT-NAME-FR	recrutement	VAL

103. Фрейм концепта TERRORIST-SUICIDE

LABEL	TERRORIST-SUICIDE	VAL
DEFINITION	Самоубийство террориста в результате теракта, часто в результате самоподрыва	VAL
IS-A	CONSEQUENCES-FOR-TERRORIST	VAL
THEME	AFFECTED-TERRORIST	SEM
DEFAULT-NAME-RU	самоубийство террориста	VAL

DEFAULT-NAME-EN	terrorist suicide	VAL
DEFAULT-NAME-FR	suicide d'un terroriste	VAL

104. Фрейм концепта TERRORIST-THREAT

LABEL	TERRORIST-THREAT	VAL
DEFINITION	Террористическая угроза, запугивание населения террористами, например, посредством демонстрации видео, ложными сообщениями о минировании	VAL
IS-A	CRIMINAL-ACTIVITY	VAL
AGENT	TERRORISM-AGENT, TERRORIST-ORGANIZATION	SEM
DEFAULT-NAME-RU	террористическая угроза	VAL
DEFAULT-NAME-EN	terrorist threat	VAL
DEFAULT-NAME-FR	menace terroriste	VAL

105. Фрейм концепта TIME

LABEL	TIME	VAL
DEFINITION	Время, дата, обстоятельства совершения теракта	VAL
IS-A	MENTAL-OBJECT	VAL
SOBCLASSES	CICUMSTANCES, EXACT-DATE, EXACT-TIME	VAL
DEFAULT-NAME-RU	время теракта	VAL
DEFAULT-NAME-EN	time of attack	VAL
DEFAULT-NAME-FR	heure de l'attentat	VAL

106. Фрейм концепта TRANSPORT-FACILITY

LABEL	TRANSPORT-FACILITY	VAL
DEFINITION	Транспортный объект, на которое направлен террористический акт, например метро, вокзал	VAL
IS-A	PROPERTY-TARGET	VAL
DIRECTION-OF	TERROR-ATTACK	SEM
DEFAULT-NAME-RU	транспортное учреждение	VAL
DEFAULT-NAME-EN	transportation facility	VAL
DEFAULT-NAME-FR	installation de transport	VAL

107. Фрейм концепта UNARMED-ASSAULT

LABEL	UNARMED-ASSAULT	VAL
DEFINITION	Нападение террористов без оружия	VAL
IS-A	TERROR-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET, PROPERTY-TARGET, OBJECT-TARGET	SEM
INSTRUMENT	*nothing*	VAL
LOCATED	LOCATION	SEM

PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	невооруженное нападение	VAL
DEFAULT-NAME-EN	unarmed assault	VAL
DEFAULT-NAME-FR	agression non-armée	VAL

108. Фрейм концепта VEHICLE-RAMMING-ATTACK

LABEL	VEHICLE-RAMMING-ATTACK	VAL
DEFINITION	Теракт в виде наезда на толпу людей транспортным средством, часто грузовиком	VAL
IS-A	TERROR-ATTACK	VAL
AGENT	TERRORISM-AGENT	DEF
	TERRORIST-ORGANIZATION	SEM
DIRECTION	HUMAN-TARGET	SEM
INSTRUMENT	RAMMING-VEHICLE	SEM
LOCATED	LOCATION	SEM
PRECONDITION	TERROR-ATTACK-PLANNING-EVENT	SEM
PRECONDITION-OF	CLAIM-RESPONSIBILITY, CONSEQUENCES	SEM
RESPONSIBILITY-CLAIMED-BY	TERRORIST-ORGANIZATION	DEF
	TERRORISM-AGENT	SEM
HAS-GOAL	GOAL-OF-ATTACK	SEM
HAS-TYPE	*nothing*	VAL
SCALE-OF-ATTACK	TRUE, FALSE	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	наезд	VAL
DEFAULT-NAME-EN	ramming attack	VAL
DEFAULT-NAME-FR	attaque au bélier	VAL

109. Фрейм концепта VICTIM

LABEL	VICTIM	VAL
DEFINITION	Человек, пострадавший в результате теракта	VAL
IS-A	HUMAN	VAL
THEME-OF	CONSEQUENCES-FOR-PEOPLE	SEM
AGE	>0	VAL
GENDER	MALE, FEMALE	VAL
HAS-RELIGION	RELIGION	SEM
HAS-NATIONALITY	NATION	SEM
DEFAULT-NAME-RU	пострадавший	VAL

DEFAULT-NAME-EN	victim	VAL
DEFAULT-NAME-FR	victime	VAL

110. Фрейм концепта VICTIM-SURVIVAL

LABEL	VICTIM-SURVIVAL	VAL
DEFINITION	Спасение человека после теракта (например, вызволение из заложников)	VAL
IS-A	CONSEQUENCES-FOR-PEOPLE	VAL
PRECONDITION	TERROR-ATTACK	SEM
THEME	VICTIM	SEM
DEFAULT-NAME-RU	спасение	VAL
DEFAULT-NAME-EN	survival	VAL
DEFAULT-NAME-FR	survie	VAL

111. Фрейм концепта WAR-RELATED-TERRORIST-EVENT

LABEL	WAR-RELATED-TERRORIST-EVENT	VAL
DEFINITION	Военные действия террористов	VAL
IS-A	OTHER-TERRORIST-EVENT	VAL
AGENT	TERRORISM-AGENT, TERRORIST-ORGANIZATION	SEM
DEFAULT-NAME-RU	военная деятельность террористов	VAL
DEFAULT-NAME-EN	war-related terrorist activity	VAL
DEFAULT-NAME-FR	activité terroriste liée à la guerre	VAL

112. Фрейм концепта WEAPON

LABEL	WEAPON	VAL
DEFINITION	Оружие или подобные объекты (например, грузовик), используемые для совершения террористического акта, а также функциональные элементы такого оружия	VAL
IS-A	ARTIFACT	VAL
SUBCLASSES	BIOLOGICAL-WEAPON, CHEMICAL-WEAPON, GUN, EXPLOSIVE-DEVICE, INCENDIARY-WEAPON, KNIFE, RAMMING-VEHICLE, ROCKET	VAL
INSTRUMENT-OF	TERROR-ATTACK	SEM
WEAPON-OF	TERRORISM-AGENT, ARTIFACT	SEM
DEFAULT-NAME-RU	оружие	VAL
DEFAULT-NAME-EN	weapon	VAL
DEFAULT-NAME-FR	arme	VAL

113. Фрейм концепта WHITE-SUPREMACISM

LABEL	WHITE-SUPREMACISM	VAL
DEFINITION	Политическая идеология, утверждающая, что белые люди превосходят других людей по расовым особенностям	VAL

IS-A	IDEOLOGY	VAL
ASSUMPTION	TRUE, FALSE	VAL
DEFAULT-NAME-RU	идеология белого превосходства	VAL
DEFAULT-NAME-EN	white supremacism	VAL
DEFAULT-NAME-FR	suprémacisme blanc	VAL

114. Фрейм концепта MEMBER-OF

LABEL	MEMBER-OF	VAL
DEFINITION	Связывает террориста и организацию	VAL
IS-A	OBJECT-RELATION	VAL
DOMAIN	TERRORISM-AGENT	SEM
RANGE	TERRORIST-ORGANIZATION	SEM
INVERSE	HAS-MEMBER	VAL

115. Фрейм концепта AGENT

LABEL	AGENT	VAL
DEFINITION	Субъект, совершающий или побуждающий действие	VAL
IS-A	CASE-ROLE	VAL
DOMAIN	EVENT	SEM
RANGE	OBJECT	SEM
INVERSE	AGENT-OF	VAL

116. Фрейм концепта AGENT-OF

LABEL	AGENT-OF	VAL
DEFINITION	Субъект, совершающий или побуждающий действие	VAL
IS-A	INVERSE-CASE-ROLE	VAL
DOMAIN	OBJECT	SEM
RANGE	EVENT	SEM
INVERSE	AGENT	VAL

117. Фрейм концепта CAPITAL-OF

LABEL	CAPITAL-OF	VAL
DEFINITION	Связывает столицу и государство	VAL
IS-A	OBJECT-RELATION	VAL
DOMAIN	CAPITAL-CITY	SEM
RANGE	COUNTRY, COUNTRY-AREA, TERRITORY, NATION	SEM
INVERSE	HAS-CAPITAL	VAL

118. Фрейм концепта CLAIMED-RESPONSIBILITY-FOR

LABEL	CLAIMED-RESPONSIBILITY-FOR	VAL
-------	----------------------------	-----

DEFINITION	Связывает теракт с террористической организацией или (реже) террористом, которые заявили, что этот теракт был совершен ими	VAL
IS-A	OBJECT-EVENT-RELATION	VAL
DOMAIN	TERRORISM-AGENT, TERRORIST-ORGANIZATION	SEM
RANGE	TERROR-ATTACK	SEM
INVERSE	RESPONSIBILITY-CLAIMED-BY	VAL

119. Фрейм концепта DIRECTION

LABEL	DIRECTION	VAL
DEFINITION	Связывает теракт и объект теракта	VAL
IS-A	CASE-ROLE	VAL
DOMAIN	TERROR-ATTACK	SEM
RANGE	HUMAN-TARGET, PROPERTY-TARGET, OBJECT-TARGET	SEM
INVERSE	DIRECTION-OF	VAL

120. Фрейм концепта DIRECTION-OF

LABEL	DIRECTION-OF	VAL
DEFINITION	Связывает теракт и объект теракта	VAL
IS-A	INVERSE-CASE-ROLE	VAL
DOMAIN	HUMAN-TARGET, PROPERTY-TARGET, OBJECT-TARGET	SEM
RANGE	TERROR-ATTACK	SEM
INVERSE	DIRECTION	VAL

121. Фрейм концепта GOAL-OF

LABEL	GOAL-OF	VAL
DEFINITION	Связывает теракт и цель (причину) теракта	VAL
IS-A	OBJECT-RELATION	VAL
DOMAIN	GOAL-OF-ATTACK	SEM
RANGE	TERROR-ATTACK	SEM
INVERSE	HAS-GOAL	VAL

122. Фрейм концепта HAS-CAPITAL

LABEL	HAS-CAPITAL	VAL
DEFINITION	Связывает столицу и государство	VAL
IS-A	INVERSE-OBJECT-RELATION	VAL
DOMAIN	COUNTRY, TERRITORY, NATION	SEM
RANGE	CAPTIAL-CITY	SEM
INVERSE	CAPITAL-OF	VAL

123. Фрейм концепта HAS-GOAL

LABEL	HAS-GOAL	VAL
-------	----------	-----

DEFINITION	Связывает теракт и цель (причину) теракта	VAL
IS-A	INVERSE-OBJECT-RELATION	VAL
DOMAIN	TERROR-ATTACK	SEM
RANGE	GOAL-OF-ATTACK	SEM
INVERSE	GOAL-OF	VAL

124. Фрейм концепта HAS-IDEOLOGY

LABEL	HAS-IDEOLOGY	VAL
DEFINITION	Связывает террориста или террористическую организацию с идеологией	VAL
IS-A	INVERSE-OBJECT-RELATION	VAL
DOMAIN	TERRORISM-AGENT, TERRORIST-ORGANIZATION	SEM
RANGE	IDEOLOGY	SEM

125. Фрейм концепта HAS-NATIONALITY

LABEL	HAS-NATIONALITY	VAL
DEFINITION	Указывает на национальную принадлежность террориста, жертвы теракта, объекта теракта, источника сообщения	VAL
IS-A	INVERSE-OBJECT-RELATION	VAL
DOMAIN	HUMAN, SOURCE	SEM
RANGE	NATION	SEM

126. Фрейм концепта HAS-MEMBER

LABEL	HAS-MEMBER	VAL
DEFINITION	Связывает террориста и организацию	VAL
IS-A	INVERSE-OBJECT-RELATION	VAL
DOMAIN	TERRORIST-ORGANIZATION	SEM
RANGE	TERRORISM-AGENT	SEM
INVERSE	MEMBER-OF	VAL

127. Фрейм концепта HAS-PART

LABEL	HAS-PART	VAL
DEFINITION	Партитивные отношения	VAL
IS-A	PARTITIVE-RELATION	VAL
DOMAIN	EVENT, OBJECT	SEM
RANGE	EVENT, OBJECT	SEM
INVERSE	PART-OF	VAL

128. Фрейм концепта HAS-RELIGION

LABEL	HAS-RELIGION	VAL
DEFINITION	Указывает на религиозную принадлежность террориста, жертвы теракта, объекта теракта	VAL

IS-A	INVERSE-OBJECT-RELATION	VAL
DOMAIN	HUMAN	SEM
RANGE	RELIGION	SEM

129. Фрейм концепта HAS-TYPE

LABEL	HAS-TYPE	VAL
DEFINITION	Указывает на тип теракта	VAL
IS-A	INVERSE-OBJECT-RELATION	VAL
DOMAIN	TERROR-ATTACK	SEM
RANGE	ARMED-ASSAULT, ARSON, BIOLOGICAL-ATTACK, BOMB-ATTACK, CHEMICAL-ATTACK, DISTANT-ATTACK, HIJACKING, HOSTAGE-TAKING, ASSASSINATION, SUICIDE-ATTACK, UNARMED-ASSAULT, VEHICLE-RAMMING-ATTACK, CYBER-ATTACK	SEM

130. Фрейм концепта HAS-WEAPON

LABEL	HAS-WEAPON	VAL
DEFINITION	Связывает террориста и оружие, которым террорист обладает, либо предмет и оружие, соединенное с этим предметом	VAL
IS-A	INVERSE-OBJECT-RELATION	VAL
DOMAIN	TERRORISM-AGENT, ARTIFACT	SEM
RANGE	WEAPON	SEM
INVERSE	WEAPON-OF	VAL

131. Фрейм концепта INSTRUMENT

LABEL	INSTRUMENT	VAL
DEFINITION	Указывает на средство совершения теракта	VAL
IS-A	CASE-ROLE	VAL
DOMAIN	TERROR-ATTACK	SEM
RANGE	WEAPON	SEM
INVERSE	INSTRUMENT-OF	VAL

132. Фрейм концепта INSTRUMENT-OF

LABEL	INSTRUMENT-OF	VAL
DEFINITION	Указывает на средство совершения теракта	VAL
IS-A	INVERSE-CASE-ROLE	VAL
DOMAIN	WEAPON	SEM
RANGE	TERROR-ATTACK	SEM
INVERSE	INSTRUMENT	VAL

133. Фрейм концепта LOCATED

LABEL	LOCATED	VAL
-------	---------	-----

DEFINITION	Указывает на место совершения теракта	VAL
IS-A	CASE-ROLE	VAL
DOMAIN	TERROR-ATTACK	SEM
RANGE	LOCATION	SEM
INVERSE	LOCATION-OFF	VAL

134. Фрейм концепта LOCATION-OF

LABEL	LOCATION-OF	VAL
DEFINITION	Указывает на место совершения теракта	VAL
IS-A	INVERSE-CASE-ROLE	VAL
DOMAIN	LOCATION	SEM
RANGE	TERRORISM-EVENT	SEM
INVERSE	LOCATED	VAL

135. Фрейм концепта PART-OF

LABEL	PART-OF	VAL
DEFINITION	Партитивные отношения	VAL
IS-A	PARTITIVE-RELATION	VAL
DOMAIN	EVENT, OBJECT	SEM
RANGE	EVENT, OBJECT	SEM
INVERSE	HAS-PART	VAL

136. Фрейм концепта PRECONDITION

LABEL	PRECONDITION	VAL
DEFINITION	Связывает события с другими событиями, которые являются их обязательными предпосылками	VAL
IS-A	REASON	VAL
DOMAIN	EVENT	SEM
RANGE	EVENT	SEM
INVERSE	PRECONDITION-OF	VAL

137. Фрейм концепта PRECONDITION-OF

LABEL	PRECONDITION-OF	VAL
DEFINITION	Связывает события с другими событиями, которые являются их обязательными предпосылками	VAL
IS-A	REASON	VAL
DOMAIN	EVENT	SEM
RANGE	EVENT	SEM
INVERSE	PRECONDITION	VAL

138. Фрейм концепта RESPONSIBILITY-CLAIMED-BY

LABEL	RESPONSIBILITY-CLAIMED-BY	VAL
-------	---------------------------	-----

DEFINITION	Связывает теракт с террористической организацией или (реже) террористом, которые заявили, что этот теракт был совершен ими	VAL
IS-A	EVENT-OBJECT-RELATION	VAL
DOMAIN	TERROR-ATTACK	SEM
RANGE	TERRORISM-AGENT, TERRORIST-ORGANIZATION	SEM
INVERSE	CLAIMED-RESPONSIBILITY-FOR	VAL

139. Фрейм концепта THEME

LABEL	THEME	VAL
DEFINITION	Связывает действие с сущностью, на которую воздействуют	VAL
IS-A	CASE-ROLE	VAL
DOMAIN	DAMAGE, CONSEQUENCES-FOR-TERRORISTS, CONSEQUENCES-FOR-PEOPLE	SEM
RANGE	AFFECTED-OBJECT, AFFECTED-TERRORIST, VICTIM	SEM
INVERSE	THEME-OF	VAL

140. Фрейм концепта THEME-OF

LABEL	THEME-OF	VAL
DEFINITION	Связывает действие с сущностью, на которую воздействуют	VAL
IS-A	INVERSE-CASE-ROLE	VAL
DOMAIN	AFFECTED-OBJECT, AFFECTED-TERRORIST, VICTIM	SEM
RANGE	DAMAGE, CONSEQUENCES-FOR-TERRORISTS, CONSEQUENCES-FOR-PEOPLE	SEM
INVERSE	THEME	VAL

141. Фрейм концепта WEAPON-OF

LABEL	WEAPON-OF	VAL
DEFINITION	Связывает террориста и оружие, которым террорист обладает, либо предмет и оружие, соединенное с этим предметом	VAL
IS-A	OBJECT-RELATION	VAL
DOMAIN	WEAPON	SEM
RANGE	TERRORISM-AGENT, ARTIFACT	SEM
INVERSE	HAS-WEAPON	VAL

142. Фрейм концепта AGE

LABEL	AGE	VAL
DEFINITION	Возраст человека	VAL
IS-A	SCALAR-ATTRIBUTE	VAL
DOMAIN	HUMAN	SEM
RANGE	> 0	SEM

143. Фрейм концепта ASSUMPTION

LABEL	ASSUMPTION	VAL
DEFINITION	Предположение	VAL
IS-A	LITERAL-ATTRIBUTE	VAL
DOMAIN	AFFECTED-TERRORIST, IDEOLOGY, TERRORISM-AGENT, TERRORIST-ORGANIZATION, TERROR-ATTACK, TERROR-ATTACK-PLANNING-EVENT, OTHER-TERRORIST-EVENT, TERRORIST-THREAT	SEM
RANGE	TRUE, FALSE	SEM

144. Фрейм концепта GENDER

LABEL	GENDER	VAL
DEFINITION	Пол человека	VAL
IS-A	LITERAL-ATTRIBUTE	VAL
DOMAIN	HUMAN	SEM
RANGE	MALE, FEMALE	SEM

145. Фрейм концепта NEGATION

LABEL	NEGATION	VAL
DEFINITION	Отрицание	VAL
IS-A	LITERAL-ATTRIBUTE	VAL
DOMAIN	ASSUME, AFFECTED-TERRORIST, AFFECTED-OBJECT, CONSEQUENCES, CLAIM-RESPONSIBILITY, DECLARE, TERRORIST-ORGANIZATION, VICTIM	SEM
RANGE	TRUE, FALSE	SEM

146. Фрейм концепта SCALE-OF-ATTACK

LABEL	SCALE-OF-ATTACK	VAL
DEFINITION	Масштабность (массовость) теракта	VAL
IS-A	LITERAL-ATTRIBUTE	VAL
DOMAIN	TERROR-ATTACK	SEM
RANGE	TRUE, FALSE	SEM

ПРИЛОЖЕНИЕ Д

Фреймы экземпляров онтологии терроризма

В приложении приведены фреймы экземпляров онтологии терроризма. Ввиду большого количества экземпляров в базе знаний модели интеллектуального контент-анализа, в приложении в качестве примера приведены фреймы только тех экземпляров, которые упомянуты в тексте диссертационной работы.

1. Фрейм экземпляра CITY-86

LABEL	CITY-86	VAL
DEFINITION	Крупнейший город в южной части Западного берега реки Иордан, столица палестинской провинции Хеврон	VAL
INSTANCE-OF	CITY	VAL
PART-OF	TERRITORY-2	VAL
DEFAULT-NAME-RU	Хеврон	VAL
DEFAULT-NAME-EN	Hebron	VAL
DEFAULT-NAME-FR	Hébron	VAL

2. Фрейм экземпляра CITY-102

LABEL	CITY-102	VAL
DEFINITION	Израильский местный совет и городское поселение в Иудее (на Западном берегу реки Иордан), вплотную примыкающее к Хеврону	VAL
INSTANCE-OF	CITY	VAL
PART-OF	TERRITORY-2	VAL
DEFAULT-NAME-RU	Кирьят-Арба	VAL
DEFAULT-NAME-EN	Kiryat Arba	VAL
DEFAULT-NAME-FR	Kiryat Arba	VAL

3. Фрейм экземпляра CITY-157

LABEL	CITY-157	VAL
DEFINITION	Город на Южном острове Новой Зеландии, административный центр региона Кентербери	VAL
INSTANCE-OF	CITY	VAL
PART-OF	COUNTRY-124	VAL
DEFAULT-NAME-RU	Крайстчерч	VAL
DEFAULT-NAME-EN	Christchurch	VAL
DEFAULT-NAME-FR	Christchurch	VAL

4. Фрейм экземпляра CAPITAL-CITY-186

LABEL	CAPITAL-CITY-186	VAL
DEFINITION	Столица и крупнейший город Соединенного Королевства Великобритании и Северной Ирландии	VAL
INSTANCE-OF	CAPITAL-CITY	VAL
PART-OF	COUNTRY-186	VAL
CAPITAL-OF	COUNTRY-186	VAL
DEFAULT-NAME-RU	Лондон	VAL
DEFAULT-NAME-EN	London	VAL
DEFAULT-NAME-FR	Londres	VAL

5. Фрейм экземпляра COUNTRY-78

LABEL	COUNTRY-78	VAL
DEFINITION	Государство в Южной Азии, граничащее с Пакистаном на западе, с Китаем, Непалом и Бутаном на северо-востоке, с Бангладеш и Мьянмой на востоке	VAL
INSTANCE-OF	COUNTRY	VAL
HAS-PART	CAPITAL-CITY-78, COUNTRY-AREA-35	VAL
HAS-CAPITAL	CAPITAL-CITY-78	VAL
DEFAULT-NAME-RU	Индия	VAL
DEFAULT-NAME-EN	India	VAL
DEFAULT-NAME-FR	Inde	VAL

6. Фрейм экземпляра COUNTRY-124

LABEL	COUNTRY-124	VAL
DEFINITION	Государство в Полинезии	VAL
INSTANCE-OF	COUNTRY	VAL
HAS-PART	CAPITAL-CITY-124, CITY-157 ...	VAL
HAS-CAPITAL	CAPITAL-CITY-124	VAL
DEFAULT-NAME-RU	Новая Зеландия	VAL
DEFAULT-NAME-EN	New Zealand	VAL
DEFAULT-NAME-FR	Nouvelle-Zélande	VAL

7. Фрейм экземпляра COUNTRY-AREA-35

LABEL	COUNTRY-AREA-35	VAL
DEFINITION	Штат на востоке Индии, столица — Диспур, крупнейший город — Гувахати	VAL
INSTANCE-OF	COUNTRY-AREA	VAL
PART-OF	COUNTRY-78	VAL
DEFAULT-NAME-RU	Ассам	VAL
DEFAULT-NAME-EN	Assam	VAL
DEFAULT-NAME-FR	Assam	VAL

8. Фрейм экземпляра NATION-9

LABEL	NATION-9	VAL
DEFINITION	Государство в Южном полушарии, занимающее одноименный материк, остров Тасмания и несколько других островов Индийского и Тихого океанов	VAL
INSTANCE-OF	NATION	VAL
HAS-CAPITAL	CAPITAL-CITY-9	VAL
DEFAULT-NAME-RU	Австралия	VAL
DEFAULT-NAME-EN	Australia	VAL
DEFAULT-NAME-FR	Australie	VAL

9. Фрейм экземпляра NATION-78

LABEL	NATION-78	VAL
DEFINITION	Государство в Южной Азии, граничащее с Пакистаном на западе, с Китаем, Непалом и Бутаном на северо-востоке, с Бангладеш и Мьянмой на востоке	VAL
INSTANCE-OF	NATION	VAL
HAS-CAPITAL	CAPITAL-CITY-78	VAL
DEFAULT-NAME-RU	Индия	VAL
DEFAULT-NAME-EN	India	VAL
DEFAULT-NAME-FR	Inde	VAL

10. Фрейм экземпляра NATION-124

LABEL	NATION-124	VAL
DEFINITION	Государство в Полинезии	VAL
INSTANCE-OF	NATION	VAL
HAS-CAPITAL	CAPITAL-CITY-124	VAL
DEFAULT-NAME-RU	Новая Зеландия	VAL
DEFAULT-NAME-EN	New Zealand	VAL
DEFAULT-NAME-FR	Nouvelle-Zélande	VAL

11. Фрейм экземпляра TERRITORY-2

LABEL	TERRITORY-2	VAL
DEFINITION	Регион на Ближнем Востоке, который является одной из двух частей частично признанного арабского государства Палестина	VAL
INSTANCE-OF	TERRITORY	VAL
HAS-PART	CITY-86, CITY-102 ...	VAL
DEFAULT-NAME-RU	Западный берег реки Иордан	VAL
DEFAULT-NAME-EN	West Bank	VAL
DEFAULT-NAME-FR	Cisjordanie	VAL

12. Фрейм экземпляра TERRORIST-ORGANIZATION-3

LABEL	TERRORIST-ORGANIZATION-3	VAL
DEFINITION	Курдская националистическая вооруженная группировка, которая борется за независимое курдское государство в юго-восточной Турции	VAL
INSTANCE-OF	TERRORIST-ORGANIZATION	VAL
HAS-IDEOLOGY	NATIONALISM, SEPARATISM	VAL
DEFAULT-NAME-RU	Ястребы свободы Курдистана	VAL
DEFAULT-NAME-EN	Kurdistan Freedom Hawks	VAL
DEFAULT-NAME-FR	Les Faucons de la liberté du Kurdistan	VAL

13. Фрейм экземпляра TERRORIST-ORGANIZATION-8

LABEL	TERRORIST-ORGANIZATION-8	VAL
DEFINITION	Военизированная ливанская шиитская организация и политическая партия, выступающая за создание в Ливане исламского государства по образцу Ирана	VAL
INSTANCE-OF	TERRORIST-ORGANIZATION	VAL
HAS-IDEOLOGY	ISLAMISM, NATIONALISM	VAL
DEFAULT-NAME-RU	Хезболла	VAL
DEFAULT-NAME-EN	Hezbollah	VAL
DEFAULT-NAME-FR	Hezbollah	VAL

14. Фрейм экземпляра TERRORIST-ORGANIZATION-15

LABEL	TERRORIST-ORGANIZATION-15	VAL
DEFINITION	Сепаратистская организация, запрещенная в Индии, требующая независимости штата Ассам	VAL
INSTANCE-OF	TERRORIST-ORGANIZATION	VAL
HAS-IDEOLOGY	NATIONALISM, SEPARATISM	VAL
DEFAULT-NAME-RU	ULFA	VAL
DEFAULT-NAME-EN	United Liberation Front of Assam	VAL
DEFAULT-NAME-FR	L'United Liberation Front of Asom	VAL

ПРИЛОЖЕНИЕ Е

Аксиомы онтологии предметной области «Терроризм»

Ниже приведен список всех аксиом, введенных в онтологию терроризма.

$$1. \forall T \text{ located}(T, La) \cap \text{part of}(La, Lc) \rightarrow \text{located}(T, Lc),$$

где T — TERROR-ATTACK, La — CITY, Lc — COUNTRY; т. е. для всех терактов верно, что, если теракт произошел в некоем городе и известно, что этот город находится в некоей стране, то, следовательно, теракт произошел в этой стране.

$$2. \forall T \text{ located}(T, Laa) \cap \text{part of}(Laa, Lc) \rightarrow \text{located}(T, Lc),$$

где T — TERROR-ATTACK, Laa — CAPITAL-CITY, Lc — COUNTRY; т. е. для всех терактов верно, что, если теракт произошел в некоей столице и известно, что эта столица находится в некоей стране, то, следовательно, теракт произошел в этой стране.

$$3. \forall T \text{ located}(T, Lb) \cap \text{part of}(Lb, Lc) \rightarrow \text{located}(T, Lc),$$

где T — TERROR-ATTACK, Lb — CITY-AREA, Lc — COUNTRY; т. е. для всех терактов верно, что, если теракт произошел в некоей части города и известно, что эта часть города находится в некоей стране, то, следовательно, теракт произошел в этой стране.

$$4. \forall T \text{ located}(T, Ld) \cap \text{part of}(Ld, Lc) \rightarrow \text{located}(T, Lc),$$

где T — TERROR-ATTACK, Ld — COUNTRY-AREA, Lc — COUNTRY; т. е. для всех терактов верно, что, если теракт произошел в некоем регионе и известно, что этот регион находится в некоей стране, то, следовательно, теракт произошел в этой стране.

$$5. \forall T \text{ located}(T, Le) \cap \text{part of}(Le, Lc) \rightarrow \text{located}(T, Lc),$$

где T — TERROR-ATTACK, Le — SPECIFIC-LOCATION, Lc — COUNTRY; т. е. для всех терактов верно, что, если теракт произошел в некоем конкретном месте и известно, что это конкретное место находится в некоей стране, то, следовательно, теракт произошел в этой стране.

$$6. \forall T \text{ located}(T, La) \cap \text{part of}(La, Lf) \rightarrow \text{located}(T, Lf),$$

где T — TERROR-ATTACK, La — CITY, Lf — TERRITORY; т. е. для всех терактов верно, что, если теракт произошел в некоем городе и известно, что этот город находится на некоей территории, то, следовательно, теракт произошел на этой территории.

$$7. \forall T \text{ located}(T, Laa) \cap \text{part of}(Laa, Lf) \rightarrow \text{located}(T, Lf),$$

где T — TERROR-ATTACK, Laa — CAPITAL-CITY, Lf — TERRITORY; т. е. для всех терактов верно, что, если теракт произошел в некоей столице и известно, что эта столица находится на некоей территории, то, следовательно, теракт произошел на этой территории.

$$8. \forall T \text{ located}(T, Lb) \cap \text{part of}(Lb, Lf) \rightarrow \text{located}(T, Lf),$$

где T — TERROR-ATTACK, Lb — CITY-AREA, Lf — TERRITORY; т. е. для всех терактов верно, что, если теракт произошел в некоей части города и известно, что этот город находится на некоей территории, то, следовательно, теракт произошел на этой территории.

$$9. \forall T \text{ located}(T, Ld) \cap \text{part of}(Ld, Lf) \rightarrow \text{located}(T, Lf),$$

где T — TERROR-ATTACK, Ld — COUNTRY-AREA, Lf — TERRITORY; т. е. для всех терактов верно, что, если теракт произошел в некоем регионе и известно, что этот регион находится на некоей территории, то, следовательно, теракт произошел на этой территории.

$$10. \forall T \text{ located}(T, Le) \cap \text{part of}(Le, Lf) \rightarrow \text{located}(T, Lf),$$

где T — TERROR-ATTACK, Le — SPECIFIC-LOCATION, Lf — TERRITORY; т. е. для всех терактов верно, что, если теракт произошел в некоем конкретном месте и известно, что это место находится на некоей территории, то, следовательно, теракт произошел на этой территории.

$$11. \forall T \text{ instrument}(T, Ca) \rightarrow \text{type}(T, Tc),$$

где T — TERROR-ATTACK, Ca — BIOLOGICAL-WEAPON, Tc — BIOLOGICAL-ATTACK; т. е. для всех терактов верно, что если теракт был совершен с использованием биологического оружия, то типом теракта является биологическая атака.

$$12. \forall T \text{ instrument}(T, Cb) \rightarrow \text{type}(T, Te),$$

где T — TERROR-ATTACK, C_b — CHEMICAL-WEAPON, T_e — CHEMICAL-ATTACK; т. е. для всех терактов верно, что если теракт был совершен с использованием химического оружия, то типом теракта является химическая атака.

$$13. \forall T \text{ instrument}(T, C_c) \rightarrow \text{type}(T, T_{aa}),$$

где T — TERROR-ATTACK, C_c — GUN, T_{aa} — GUN-ATTACK; т. е. для всех терактов верно, что если теракт был совершен с использованием огнестрельного оружия, то типом теракта является стрельба.

$$14. \forall T \text{ instrument}(T, C_d) \rightarrow \text{type}(T, T_d),$$

где T — TERROR-ATTACK, C_d — EXPLOSIVE-DEVICE, T_d — BOMB-ATTACK; т. е. для всех терактов верно, что если теракт был совершен с использованием взрывного устройства, то типом теракта является взрыв.

$$15. \forall T \text{ instrument}(T, C_{da}) \rightarrow \text{type}(T, T_{db}),$$

где T — TERROR-ATTACK, C_{da} — SUICIDE-VEST, T_{db} — SUICIDE-BOMB-ATTACK; т. е. для всех терактов верно, что если теракт был совершен с использованием пояса шахида, то типом теракта является самоподрыв террориста-смертника.

$$16. \forall T \text{ instrument}(T, C_{db}) \rightarrow \text{type}(T, T_{da}),$$

где T — TERROR-ATTACK, C_{db} — CAR-BOMB, T_{da} — CAR-BOMB-ATTACK; т. е. для всех терактов верно, что если теракт был совершен с использованием начиненного взрывчаткой автомобиля, то типом теракта является подрыв автомобиля.

$$17. \forall T \text{ instrument}(T, C_e) \rightarrow \text{type}(T, T_b),$$

где T — TERROR-ATTACK, C_e — INCENDIARY-WEAPON, T_b — ARSON; т. е. для всех терактов верно, что если теракт был совершен с использованием зажигательного оружия, то типом теракта является поджог.

$$18. \forall T \text{ instrument}(T, C_f) \rightarrow \text{type}(T, T_{ab}),$$

где T — TERROR-ATTACK, C_f — KNIFE, T_{ab} — KNIFE-ATTACK; т. е. для всех терактов верно, что если теракт был совершен с использованием холодного оружия, то типом теракта является нападение с ножом.

$$19. \forall T \text{ instrument}(T, C_g) \rightarrow \text{type}(T, T_l),$$

где T — TERROR-ATTACK, Cg — RAMMING-VEHICLE, Tl — VEHICLE-RAMMING-ATTACK; т. е. для всех терактов верно, что если теракт был совершен с использованием автомобиля в качестве тарана, то типом теракта является наезд автомобиля.

$$20. \forall T \text{ instrument}(T, Ch) \rightarrow \text{type}(T, Tfc),$$

где T — TERROR-ATTACK, Ch — ROCKET, Tfc — ROCKET-ATTACK; т. е. для всех терактов верно, что если теракт был совершен с использованием ракеты, то типом теракта является ракетный обстрел.

$$21. \forall A \text{ member of}(A, U) \cap \text{has ideology}(U, Wa) \rightarrow \text{has ideology}(A, Wa),$$

где A — TERRORISM-AGENT, U — TERRORIST-ORGANIZATION, Wa — ISLAMISM; т. е. для всех исполнителей терактов верно, что если исполнитель является членом некой террористической организации и такая террористическая организация проповедует исламизм, то, следовательно, этот исполнитель является исламистом.

$$22. \forall A \text{ member of}(A, U) \cap \text{has ideology}(U, Wb) \rightarrow \text{has ideology}(A, Wb),$$

где A — TERRORISM-AGENT, U — TERRORIST-ORGANIZATION, Wb — RIGHT-WING-EXTREMISM; т. е. для всех исполнителей терактов верно, что если исполнитель является членом некой террористической организации и такая террористическая организация проповедует праворадикальный экстремизм, то, следовательно, этот исполнитель является праворадикальным экстремистом.

$$23. \forall A \text{ member of}(A, U) \cap \text{has ideology}(U, Wc) \rightarrow \text{has ideology}(A, Wc),$$

где A — TERRORISM-AGENT, U — TERRORIST-ORGANIZATION, Wc — LEFT-WING-EXTREMISM; т. е. для всех исполнителей терактов верно, что если исполнитель является членом некой террористической организации и такая террористическая организация проповедует леворадикальный экстремизм, то, следовательно, этот исполнитель является леворадикальным экстремистом.

$$24. \forall A \text{ member of}(A, U) \cap \text{has ideology}(U, Wd) \rightarrow \text{has ideology}(A, Wd),$$

где A — TERRORISM-AGENT, U — TERRORIST-ORGANIZATION, Wd — WHITE-SUPREMACISM; т. е. для всех исполнителей терактов верно, что если исполнитель является членом некой террористической организации и такая террористическая

организация проповедует идеологию белого превосходства, то, следовательно, этот исполнитель является последователем идеологии белого превосходства.

25. $\forall A \text{ member of } (A, U) \cap \text{has ideology}(U, We) \rightarrow \text{has ideology}(A, We)$,

где A — TERRORISM-AGENT, U — TERRORIST-ORGANIZATION, We — NATIONALISM; т. е. для всех исполнителей терактов верно, что если исполнитель является членом некой террористической организации и такая террористическая организация проповедует национализм, то, следовательно, этот исполнитель является националистом.

26. $\forall A \text{ member of } (A, U) \cap \text{has ideology}(U, Wf) \rightarrow \text{has ideology}(A, Wf)$,

где A — TERRORISM-AGENT, U — TERRORIST-ORGANIZATION, Wf — SEPARATISM; т. е. для всех исполнителей терактов верно, что если исполнитель является членом некой террористической организации и такая террористическая организация проповедует сепаратизм, то, следовательно, этот исполнитель является представителем идеологии сепаратизма.

27. $\forall Ab \text{ member of } (Ab, U) \cap \text{has ideology}(U, Wa) \rightarrow \text{has ideology}(Ab, Wa)$,

где Ab — TERRORIST-GROUP, U — TERRORIST-ORGANIZATION, Wa — ISLAMISM; т. е. для всех групп террористов верно, что если группа террористов состоит в некой террористической организации и такая террористическая организация проповедует исламизм, то, следовательно, эта группа террористов является исламистами.

28. $\forall Ab \text{ member of } (Ab, U) \cap \text{has ideology}(U, Wb) \rightarrow \text{has ideology}(Ab, Wb)$,

где Ab — TERRORIST-GROUP, U — TERRORIST-ORGANIZATION, Wb — RIGHT-WING-EXTREMISM; т. е. для всех групп террористов верно, что если группа террористов состоит в некой террористической организации и такая террористическая организация проповедует праворадикальный экстремизм, то, следовательно, эта группа террористов является праворадикальными экстремистами.

29. $\forall Ab \text{ member of } (Ab, U) \cap \text{has ideology}(U, Wc) \rightarrow \text{has ideology}(Ab, Wc)$,

где Ab — TERRORIST-GROUP, U — TERRORIST-ORGANIZATION, Wc — LEFT-WING-EXTREMISM; т. е. для всех групп террористов верно, что если группа террористов состоит в некой террористической организации и такая террористическая

организация проповедует леворадикальный экстремизм, то, следовательно, эта группа террористов является леворадикальными экстремистами.

30. $\forall Ab \text{ member of } (Ab, U) \cap \text{has ideology}(U, Wd) \rightarrow \text{has ideology}(Ab, Wd)$,

где Ab — TERRORIST-GROUP, U — TERRORIST-ORGANIZATION, Wd — WHITE-SUPREMACISM; т. е. для всех групп террористов верно, что если группа террористов состоит в некой террористической организации и такая террористическая организация проповедует идеологию белого превосходства, то, следовательно, эта группа террористов является последователями идеологии белого превосходства.

31. $\forall Ab \text{ member of } (Ab, U) \cap \text{has ideology}(U, We) \rightarrow \text{has ideology}(Ab, We)$,

где Ab — TERRORIST-GROUP, U — TERRORIST-ORGANIZATION, We — NATIONALISM; т. е. для всех групп террористов верно, что если группа террористов состоит в некой террористической организации и такая террористическая организация проповедует национализм, то, следовательно, эта группа террористов является националистами.

32. $\forall Ab \text{ member of } (Ab, U) \cap \text{has ideology}(U, Wf) \rightarrow \text{has ideology}(Ab, Wf)$,

где Ab — TERRORIST-GROUP, U — TERRORIST-ORGANIZATION, Wf — SEPARATISM; т. е. для всех групп террористов верно, что если группа террористов состоит в некой террористической организации и такая террористическая организация проповедует сепаратизм, то, следовательно, эта группа террористов является представителями идеологии сепаратизма.