



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ
ГУМАНИТАРНО-ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ЮУрГГПУ»)

ФАКУЛЬТЕТ ИНОСТРАННЫХ ЯЗЫКОВ
КАФЕДРА АНГЛИЙСКОЙ ФИЛОЛОГИИ

**Построение электронного франко-англо-русского словаря
на основе онтологической модели предметной области
«Терроризм»**

**Выпускная квалификационная работа по направлению
45.03.02 Лингвистика**

**Направленность программы бакалавриата
«Перевод и переводоведение»**

Форма обучения очная

Проверка на объем заимствований:
96 % авторского текста

Выполнила:
Студентка группы ОФ-403-074-4-1

Работа рекоменд. к защите
рекомендована/не рекомендована
«25» июня 2020 г.
зав. кафедрой английской
филологии
Афанасьева Ольга Юрьевна

Еремеева Алена Андреевна

Научный руководитель:
Зиновьева Анастасия Юрьевна

Челябинск
2020

ОГЛАВЛЕНИЕ

Введение	3
ЭЛЕКТРОННЫЙ СЛОВАРЬ КАК ИНСТРУМЕНТ СОВРЕМЕННОЙ ЛЕКСИКОГРАФИИ.....	7
2.1 Словарь как объект лингвистического исследования.....	7
2.1.1 Понятие словаря.....	7
2.1.2 Роль словаря в лексикографии.....	10
2.2 Классификация словарей.....	12
2.3 Электронные словари.....	19
2.3.1 Характеристика электронных словарей.....	19
2.3.2 Преимущества и недостатки электронных словарей	20
2.3.3 Способы разработки электронных словарей.....	22
Выводы по главе 1.....	25
ГЛАВА 2. ДАННЫЕ ДЛЯ РАЗРАБОТКИ ЭЛЕКТРОННОГО ФРАНКО-АНГЛО-РУССКОГО СЛОВАРЯ ПРЕДМЕТНОЙ ОБЛАСТИ «ТЕРРОРИЗМ».....	26
2.1 Сбор и подготовка материала.....	26
2.2 Подготовка частотных списков.....	28
2.3 Прототип электронного словаря.....	35
ВЫВОДЫ ПО ГЛАВЕ 2.....	38
ЗАКЛЮЧЕНИЕ.....	40
БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....	42
Приложение А.....	45
Приложение Б	50

ВВЕДЕНИЕ

Актуальность работы обусловлена прежде всего тем, что развитие информационных технологий во всем мире происходит очень стремительно. Еще несколько лет назад людям, изучающим иностранные языки, приходилось постоянно работать с массивными бумажными словарями, а в некоторых случаях, заказывать необходимые экземпляры по почте из-за границы, что не является бюджетным вариантом. На сегодняшний день в этом нет необходимости, так как любую информацию об определенной лексической единице можно найти в электронном словаре. Электронные словари являются более простыми и удобными в использовании, чем бумажные. Кроме того, их проще обновлять.

Предметной областью исследования был выбран терроризм, так как это очень важная проблема современности, которая часто рассматривается в СМИ. Тексты СМИ нередко переводятся при помощи сервисов автоматического перевода или же переводчиками, не обладающими знаниями об особенностях терминологии предметной области, и, как следствие, результат такого перевода зачастую оставляет желать лучшего. Соответственно, задача разработки электронного словаря предметной области, который описывал бы высокочастотную терминологию на трех распространенных языках, нам достаточно актуальной.

Отсюда вытекает цель настоящего исследования — собрать данные для последующей разработки электронного словаря предметной области для переводчиков на основе онтологии предметной области «Терроризм».

Объектом исследования является лексика предметной области на русском, французском и английском языках, предметом – грамматические и семантические особенности этой лексики.

В рамках исследования были поставлены следующие задачи:

- 1) рассмотреть понятие словаря, его функции и особенности;
- 2) определить свойства и элементы электронного словаря;

- 3) выполнить отбор категорий словаря на основе онтологии предметной области «Терроризм»;
- 4) отобрать лексику по частоте встречаемости;
- 5) сопоставить лексику разных языков друг с другом;
- 6) выделить грамматические и семантические особенности отобранной лексики.

В соответствии с целью и задачами исследования основными методами в данной работе являются:

- 1) количественный (подсчет частоты встречаемости);
- 2) сравнительно-сопоставительный;
- 3) контекстологический.

Теоретической базой исследования являются работы отечественных и зарубежных лингвистов в области лексикографии: Т.Р. Грубера, В.Г. Гака, В.И. Даля, С.И. Ожегова, Л.В. Щербы, Д.Н. Ушакова и многих других.

Практической базой исследования являются псевдопараллельные (сопоставимые) корпуса текстов энциклопедических статей о терактах на трех языках (русском, английском, французском) объемом в 30 000 словоупотреблений каждый. В текстах описываются такие резонансные теракты последних двух десятилетий, как стрельба в мечетях Крайстчерча, стрельба в синагоге в Питтсбурге, захват заложников в Беслане, теракты 11 сентября, теракт на Дубровке. В качестве языков исследования выбраны русский, английский и французский, поскольку они покрывают довольно большое количество текстов СМИ. В частности, английский язык является языком международного общения, поэтому многие новости дублируются на английском даже в тех странах, где он не является официальным. Французский в свою очередь является официальным языком Франции, одним из официальных языков ООН, языком франкоязычного населения Бельгии, Швейцарии, Канады, Африки, Карибского бассейна, Французской Гвианы и т. д.; при этом в Африке совершается значительное количество терактов. Наконец, русский язык является языком, используемым на территории

бывшего СССР, где теракты также нередки. Соответственно, на каждом из этих языков ежедневно генерируется огромное количество материалов для перевода.

Научная новизна исследования состоит в том, что нами впервые был разработан трехязычный словарь предметной области «Терроризм» на материале текстов из энциклопедии на основе онтологии предметной области.

Практическая значимость работы заключается в том, что результаты можно будет использовать при обучении переводу в предметной области «Терроризм», при переводе текстов предметной области, а также для расширения лексикона онтологического ресурса предметной области, следовательно, в перспективе еще и для машинного перевода.

На защиту выносятся следующие положения:

1. На основе энциклопедических текстов можно собрать список лексем, который наиболее полно отражает предметную область.
2. Онтологии предметной области могут использоваться в качестве основы для создания словаря, поскольку помогают выделить категории.

Работа общим объемом 57 страниц печатного текста состоит из введения, двух глав, заключения, библиографического списка и приложения.

Во введении обосновывается актуальность исследования, определяются цели и задачи работы.

В первой главе даётся определение словаря, приводятся различные подходы к классификации словарей, рассматриваются основные характеристики электронных словарей, описываются способы разработки электронных словарей.

Во второй главе производится отбор категорий, наиболее важных для данной предметной области, сбор корпусов текстов на трех языках, выявляется релевантная лексика, составляются частотные списки, проводится сравнительно-сопоставительный анализ лексики на разных языках в рамках каждой отдельной категории.

В заключении подводятся итоги исследования.

В библиографическом списке приводится 23 наименования работ, в том числе 4 на английском языке.

ГЛАВА 1. ЭЛЕКТРОННЫЙ СЛОВАРЬ КАК ИНСТРУМЕНТ СОВРЕМЕННОЙ ЛЕКСИКОГРАФИИ

1.1 Словарь как объект лингвистического исследования

1.1.1 Понятие словаря

Словарь — это книга или иной источник, информация которого организована посредством деления на краткие статьи, расположенные по названию или теме [2; Т. 30].

Определение понятия «словарь» варьируется от источника к источнику.

Согласно «Новому энциклопедическому словарю» [8], словарь — это организованная определенным образом «справочная книга, содержащая собрание слов» или иных единиц языка, обладающая сведениями «об их значениях, употреблении, происхождении, переводе на другой язык», а также имеющая информацию о понятиях из разных областей жизни.

В соответствии с «Толковым словарем живого великорусского языка» В. И. Даля [5] словарь — это «сборник слов, речений какого-либо языка», включающий их трактовку.

«Толковый словарь русского языка» под редакцией С. И. Ожегова и Н. Ю. Шведовой [13] предлагает свою версию: словарь — это «собрание слов, устойчивых выражений», организованное определенным образом и обладающее пояснением значения единиц языка и их переводом на другой язык.

Согласно «Толковому словарю русского языка» под редакцией Д. Н. Ушакова [14] словарь — это «книга, содержащая перечень слов», расположенных в определенном порядке и имеющих то или иное толкование.

Таким образом, на основе представленных дефиниций можно вывести следующее определение словаря: словарь — это сборник слов, состав-

ленный определенным образом, обладающий толкованием значений представленных в нем лексем и вариантами их перевода на другой язык.

Существует три основных периода развития истории лексикографии [11]:

1. Дословарный период (XI–XV вв.), с характерными формами словарей: глоссарии (глоссы) — пояснение неясных слов в отдельных работах; вокабулярии — учебные сборники слов; ономастиконы — собрания и толкование наименований. Преимуществом обладала функция обучения: объяснение неизвестных (заимствованных) слов. Функции древнерусских словарей отчетливо видны в их названиях, например «Толкование о неразумных словесех», «Толкование неудобь познаваемых в писании речем» и др.

2. Ранний словарный период (XVI–XVII вв.), для которого характерно возникновение азбуковников, лексиконов и переводных словарей. Главенствующая функция словарей в тот временной промежуток — постижение литературного языка, а также — межъязыковое взаимодействие. Одними из самых известных словарей того периода являются: «Толкование именам по алфавиту» Максима Грека и «Лексис, сиречь речения вкратце собраны и со словенского языка на просты русский диялект истолкованы» Лаврентия Зизания (1596 г.).

3. Период развитой лексикографии (XVIII–XIX вв.), для которого характерно возникновение толкового словаря национального языка, а после и различных отраслевых словарей. Главное назначение словарей в этот промежуток времени — описание и упорядочение родного языка, а вместе с тем — увеличение количества, подготовка и обработка словарных материалов для различных языковых анализов в сфере лексикологии (и иных областей языкознания), истории языка и т.п. Общеизвестные словари XVIII века: «Лексикон треязычный» Федора Поликарпова (1704 г.) и «Словарь Академии Российской» (1789-1794) (вместе со вторым изданием этого словаря — «по азбучному порядку»: 1806-1822 гг.). Общеизвестные

словари XIX века: «Словарь церковнославянского и русского языка» в 4-х томах (1847 г.), незавершенный академический «Словарь русского языка» Я.К. Грота и А.А. Шахматова (изд. с 1895) и «Толковый словарь живого великорусского языка» В.И. Даля в 4-х томах (1-е изд. 1863-1866 гг.), переиздававшийся много раз и не утративший значимости на сегодняшний день.

К XX веку в лексикографии уже сформировались основные назначения и виды словарей, и наступил период развития теоретической лексикографии. Ниже приведены выделяющиеся сегодня разновидности словарной работы [11]:

- а) двуязычная лексикография, обладающая функцией межъязыкового взаимодействия;
- б) учебная лексикография, обладающая функцией обучения языку (родному и иностранному);
- в) описательная, или академическая лексикография, обладающая функцией научного познания и описания лексических, семантических и грамматических значений языка;
- г) нормализующая (регламентирующая) лексикография, обладающая соответствующей справочной функцией, создающая различные нормативные словари.

Некоторые словари обладают не одной функцией и поэтому относятся к нескольким разновидностям.

Множество словарей, составленных в разных странах мира в разные временные отрезки, предопределили некоторые социальные задачи, связанные с основными функциями лексикографии:

- 1) межъязыкового общения (социальная функция), осуществляемая в основном двуязычными словарями и разговорниками;
- 2) обучения языку (обучающая функция), осуществляемая различными одноязычными словарями, включая специальные учебные словари;

- 3) описания и упорядочения родного языка (справочная функция), осуществляемая толковыми и другими нормативными словарями;
- 4) научного постижения языка (научная функция), осуществляемая словарями специального назначения [11].

1.1.2 Роль словаря в лексикографии

В наше время стремительного сбора информации поток публикаций в различных сферах знания стремительно растет. Все это непременно повышает значимость всевозможной справочной литературы, имеющей свое предназначение в быстром и удобном добывании необходимых сведений различного рода. Сегодня невозможно обойтись без словарей и справочников, так как в них собрано бесчисленное множество человеческих знаний. Именно словари нацелены на отражение успехов развития нации в вещественной и духовной областях. Нормы языка совершенствуются по мере его развития. Невозможно научиться разбираться в нормах литературного языка в одночасье, необходимо непрерывно улучшать свою речевую культуру. Для этого и обращаются за помощью к словарям.

Словари — наиболее эффективное средство улучшения речевой культуры, одно из главных средств непрерывной и научно оправданной языковой политики. Таким образом, кроме того, что словарь литературного языка выражает языковую деятельность какого-либо общества в конкретный период, он еще и является первым постановителем ныне существующего словоупотребления. Впрочем, словари национального языка являются не только могучим орудием культуры речи, но и вместе с тем продуктом этой культуры. Поэтому создание словарей исконного языка — важная государственная задача.

Объемные толковые словари национальных языков в большинстве своем строятся на основе обширной базы литературно-художественных

источников. Так, словари являются информационными источниками в сфере национальной литературы и культуры.

Любой словарь — это не только информационное пособие, но и теоретические знания того или иного раздела языка, обладающее тем же научным статусом, что и теоретическая лингвистика. Именно поэтому словарь — это важнейший предмет обучения. Словарь не только экономит время и силы, предоставляя доступ к знаниям, собранным предшественниками, но и осуществляет это в соответствии с одним из главных требований дидактики — учить тому, в чем есть необходимость, и приучать к знаниям, применимым в жизни. Таким образом, каждый словарь — это творение, предназначенное для обучения. Наиболее четко эта характеристика прослеживается в особых учебных словарях.

Словарь обладает обилием теоретических материалов в сфере языкознания. Именно в словарях многие теоретические вопросы впервые получают свои ответы. В них осуществляются и испытываются последствия лингвистической теории, и часто словари обгоняют науку практическими решениями еще не постигнутых вопросов. Лексикографические описания содержат в себе итоги лингвистического анализа и вследствие этого обладают внушительной научной важностью. Таким образом, лексикография перестает быть «искусством составления словарей», как говорил испанский лексикограф Хуан Кассарес, а все в большей мере становится особой научной дисциплиной, со своими задачами, объектом изучения и назначением.

Следовательно, теорией и практикой составления разнообразных словарей занимается лексикография (от греч.: *lexis, lexicos* — ‘слово’, ‘словарный’ и *grapho* — ‘пишу, описываю’). Как научный термин слово вошло в употребление относительно недавно, чуть более ста лет назад. Энциклопедические словари 1896 г. (Брокгауз и Эфрон) [3] и 1916 г. (Гранат) [19] определяют лексикографию как «словарную технику», или «научные спо-

собы обработки словесного материала языка для составления лексикона».

В настоящее время термин обладает несколькими значениями:

1) процесс составления словарей (словарное дело) — т.е. работа, заключающаяся в составлении словарей (является прикладной дисциплиной);

2) теория и методика составления словарей — т.е. научный поиск, создание норм и способов описания слов и других единиц языка (теоретическая дисциплина);

3) общее число всех словарей данного языка [11].

Таким образом, лексикография является частью языкознания, занимающейся теорией и практикой составления словарей, связанной с другими его разделами и дисциплинами, имеющей определенный предмет изучения (словари), научные и методологические принципы, конкретную теоретическую проблематику. В отечественном языкознании эта точка зрения впервые была высказана известным русским лингвистом Л.В. Щербой в его известной статье «Опыт общей теории лексикографии» (1940 г.) [18].

1.2 Классификация словарей

Существует два главных типа словарей: энциклопедические и лингвистические [11].

Объект описания лингвистических (языковых) словарей — языковые единицы (слова, словоформы, морфемы). Словарь такого рода может характеризовать слова с разных сторон может характеризоваться с разных сторон, в зависимости от назначения, размера и задач словаря: со стороны лексических и грамматических аспектов языка. В зависимости от количества признаков слова словари делятся на одноаспектные и многоаспектные [11].

Синхронические лингвистические словари представляют язык в конкретный временной отрезок (например, язык XVIII века, современный язык).

Диахронические (например, этимологические) демонстрируют процесс развития языка с течением времени.

Энциклопедические словари состоят из экстралингвистических сведений об описываемых единицах языка; такие словари включают информацию о научных представлениях, терминах, исторических явлениях и т. п. Энциклопедический словарь не содержит сведений о слове, а даёт информацию о значении, которым обладает слово.

Терминологический словарь охватывает термины конкретной сферы знания или предмета и их определения.

Типология словарей — это их научная классификация, созданная на основе представления совершенного словаря (т.е. словаря-образца). Обычно в ее основе лежат уже созданные словари. Итак, рассмотрим принципы построения подобной типологии.

Словари можно классифицировать по различным аспектам. Например, по типу, что часто можно обнаружить в их названиях: толковый, частотный, обратный, семантический, грамматический и т.п. Учитывая все аспекты при классификации словарей, возможно дать точную и всестороннюю характеристику каждому словарю согласно системе словарей.

Безусловно, именно лингвистический параметр языковой системы позволяет определить тип словаря. Тем не менее, внутренние семиотические свойства также имеют важность. К ним относятся: жанр, формальные показатели, назначение и функции словаря.

В результате за последние годы сформировались основные параметры классификации словарей: лингвистический, семиотический, формальный и социально-прагматический. Их мы и рассмотрим.

Исследованием параметров классификации занимались многие отечественные (и зарубежные) лингвисты: Л.В. Щерба, В.В. Виноградов, С.И.

Ожегов, В.Г. Гак, А.М. Цывин, Л.П. Ступин, Л.А. Новиков, П.Н. Денисов и др. Тем не менее, именно Л.В. Щерба произвел на свет научную типологию словарей в своей работе «Опыт общей теории лексикографии» 1940 года. Ниже приведены противопоставления словарей, предложенные Л.В. Щербой [18]:

- 1) словарь академического типа — словарь-справочник,
- 2) энциклопедический словарь — общий словарь,
- 3) thesaurus — обычный (толковый или переводный) словарь,
- 4) обычный словарь — идеологический словарь,
- 5) толковый словарь — переводный словарь,
- 6) неисторический словарь — исторический словарь.

Как мы можем заметить, в классификацию Л.В. Щербы заложен тип идеального словаря, который может быть полезным в описании словарного фонда национального языка с различных сторон.

Типологии других исследователей только добавляют сведения и конкретизируют информацию, освещенную в классификации Л.В. Щербы.

Так, классификация словарей по В.Г. Гаку рассматривает следующие дифференциальные признаки [4]:

- 1) содержание;
- 2) отбор лексики;
- 3) способ описания;
- 4) единица описания;
- 5) расположение информации;
- 6) эпоха функционирования слова;
- 7) назначение;
- 8) число языков.

Подобные признаки (с несколько отличающимися названиями) были предложены и другими авторами.

Исчерпывающая классификация словарей представлена в работе [11]. Автор выделяет четыре основных параметра для классификации сло-

варей (семиотический, формальный, прагматический, лингвистический), каждый из которых в свою очередь делится на четыре дифференциальных признака:

1. Семиотический параметр рассматривает специфику описания или вид языкового знака.

1.1. Со стороны единицы описания словари делятся на словари слов, фразеологизмов, частей слова и словосочетаний. Большая часть лексических словарей являются орфографическими, толковыми, синонимическими, словообразовательными, а также грамматическими. Однако существуют и словари, предназначенные для описания других языковых единиц, например, морфем или единиц синтаксиса.

1.2. Со стороны принципа описания словари делятся на семасиологические и ономасиологические. Семасиологические описывают слова от слова к значению — например, толковые словари, ономасиологические — от значения к обозначающему его слову или группе слов — например, синонимические словари.

1.3. Со стороны способа описания словари делятся на толковые и указатели (реестры, конкордансы).

1.4. Со стороны языка описания словари делятся на одноязычные и двуязычные, или переводные.

2. Формальный параметр рассматривает общие показатели словаря.

2.1. Со стороны объема словари делятся на большие и малые (краткие). Иногда выделяют еще и средние словари. С.И. Ожегов первым представил следующую классификацию для толковых словарей [9]: большой — «представляющий современный литературный язык в широкой исторической перспективе»; средний — «с детальной разработкой исторически оправданного стилистического многообразия современного литературного языка» — и краткий — «популярного типа, стремящийся к активной нормализации современной литературной речи».

2.2. Со стороны состава словари делятся на полные (тезаурусы: греч. thesaurus — сокровищница) и дифференциальные. Полные словари описывают всю лексику языка, дифференциальные — лексику, отобранную определенным способом: например, только литературную, или только диалектную и т.п.

2.3. Со стороны формы словари делятся на односторонние и двусторонние. Односторонние словари включают только словники или списки слов, т.е. это словари-реестры.

2.4. Со стороны способа расположения слов словари делятся на алфавитные и неалфавитные.

В алфавитных словарях используется алфавитный порядок расположения вокабул.

Алфавитные словари, могут быть прямыми (по алфавиту первых букв слова) и обратными (по алфавиту последних букв слова). Такой способ расположения вокабул очень удобен при анализе словообразования и словоизменения, т.к. то и другое происходит с использованием служебных морфем, стоящих в конце слов. Остальные словари являются прямыми алфавитными.

Варианты расположения слов в неалфавитном порядке: частотные (слова располагаются в порядке убывания частоты употребления в речи), тематические (слова описываются по тематическим группам, смысловым «полям»).

Также существует промежуточная группа, которая совмещает алфавитный и групповой способ расположения лексем: алфавитно-гнездовые словари (когда одна словарная статья «гнездового» характера объединяет однокоренные слова, но сами «гнезда» располагаются в алфавитном порядке исходных слов), аналогические (когда одна словарная статья описывает близкие по смыслу слова, в то время как словарные статьи располагаются в алфавитном порядке (например, словари синонимов, антонимов, паронимов)).

3. Прагматический параметр рассматривает социально-психологические качества пользователей словарей.

3.1. Со стороны национальности (читателя, пользователя) словари делятся на словари родного языка.

3.2. Со стороны возраста пользователя словари делятся на детские и взрослые. Возраст читателей начал стали учитывать еще в XVIII веке (в образовательных целях), а узкоспециальные детские словари появились недавно.

3.3. Со стороны учета категории читателей словари делятся на общие и специальные.

3.4. Со стороны назначения словари делятся на общие и учебные.

Общие словари в свою очередь делятся на описательно-академические и справочные. Учебные словари схожи со справочными, но они обладают конкретно обучающей функцией; они подобны учебникам, и иногда даже являются их частью.

4. Лингвистический параметр рассматривает языковые аспекты исследования русского языка.

4.1. Со стороны отношения к норме словари делятся на нормативные и ненормативные.

Нормативные словари описывают литературную лексику и используют нормативные пометы. В настоящее время все толковые словари относятся к нормативному типу.

К нормативным словарям относятся орфографические и орфоэпические словари, словари ударений, словари трудностей или правильности русской речи.

К ненормативным словарям относятся словари по типу тезаурусов, которые описывают и литературная и нелитературная лексика, а также диалектные, или областные словари, которые описывают диалекты, просторечия, жаргон.

4.2. Со стороны отношения ко времени словари делятся на современные и исторические.

Современные словари занимаются описанием лексики языка в определенный период времени.

Исторические словари представляют лексику в диахроническом аспекте. Они всегда содержат элементы соотношения с современностью. К ним можно отнести историко-этимологические словари, а также исторические словари, которые описывают лексику и семантику прошлых времен и показывают изменение словарного состава и языковых значений.

4.3. Со стороны отношения к объекту отражения словари делятся на общие и частные.

Общие словари рассматривают словарный и лексический состав языка, как национального, так и литературного. Орфографический или грамматический словарь также считается общим, если он представляет лексику русского языка без объяснения значений.

Частные словари представляют лишь некую часть словарного состава языка, например, конкретный говор, язык определенного писателя, произведения или жанра и т.п.

4.4. Со стороны отношения к языковому аспекту словари делятся на универсальные (многоаспектные, комплексные) и аспектные (или одноаспектные).

Целью универсальных многоаспектных словарей является всесторонняя характеристика слова: семантическая, грамматическая, стилистическая и т.п. Традиционные толковые словари, дающие многоаспектные грамматические и семантические характеристики, стремятся к такому типу, правда, пока безрезультатно.

1.3 Электронные словари

1.3.1 Характеристика электронных словарей

Компьютерная среда является неотъемлемой частью современного мира. Подавляющее большинство населения планеты пользуется всевозможными электронными устройствами ежедневно. В данную информационную модель хорошо вписываются современные электронные словари, отвечающие последним техническим характеристикам и адаптированные к электронной среде.

Несмотря на то, что такой тип словарей широко распространен, он не всегда является удобным и точным. К тому же, современная лексикография пока не выделяет достоверного значения электронного словаря, поэтому существует некий спор вокруг вопроса о том, считается ли таковым печатный словарь, преобразованный в электронный формат или же это изначально созданная особым способом компьютерная программа. Подавляющая часть исследователей склоняются ко второму варианту и считают, что такой тип словаря должен быть максимально оснащенным и оптимальным для любого пользователя.

Стремительное освоение электронного пространства привлекло к некоторым теоретическим и практическим сложностям в лексикографии. Рассмотрим некоторые из них. Как было сказано выше, заметно расхождение в понимании того, что следует считать электронным словарем, так как разница в определениях этого термина довольно внушительна.

Если условно поделить эти определения на две группы, то в первую можно отнести все термины, обозначающие электронные версии печатных лексикографических изданий, а во вторую — специально созданные для дальнейшего использования в электронном формате компьютерные словари. Любой термин из этих групп может являться определением электронного словаря, поскольку под этим словосочетанием подразумевается лю-

бое программное приложение словаря для персонального компьютера. В таком случае, пользователю не важна основа данного словаря.

Отсюда можно сделать вывод, что нет смысла в спорах об особенностях терминологии, так как все эти словари можно назвать и компьютерными, и электронными, и машинными и пр. Стоит отметить, что первые два варианта являются наиболее точными, так как отражают сферу их использования.

1.3.2 Преимущества и недостатки электронных словарей

Электронный словарь имеет множество преимуществ в отношении к печатному изданию [15].

Во-первых, электронный формат словаря не только обладает высокой скоростью поиска необходимой словарной единицы, но и предлагает варианты правильного написания, в случае ошибки. Такие словари оснащены гиперссылками, что значительно упрощает их использование. Кроме того, в них можно найти дополнительные сведения к запросу, например, варианты употребления искомой словарной единицы во фразеологизмах. Все эти возможности называются гибкостью словаря или гипертекстовостью.

Во-вторых, компьютерный словарь предлагает расширенные возможности описания словарной единицы, включая грамматические, лексические, исторические, концептуальные, культурологические характеристики, что делает его универсальным. Это свойство называют объемностью.

В-третьих, электронный формат дает возможность включения дополнительных наглядных средств, видео и аудиофайлов в словарную статью. Это свойство словаря в лексикографии называют мультимедийностью.

В-четвертых, электронный словарь в полной мере отвечает запросам пользователя, что делает его удобным и максимально ориентированным на

любого читателя. Такое свойство словаря в лексикографии называют интерактивностью.

В-пятых, компьютерные словари непрерывно пополняются, расширяются и обновляются, что делает его актуальным.

Все вышеперечисленные преимущества электронных словарей неизбежно влекут за собой появление новых сложностей при их создании.

1. Прежде всего, современные электронные словари быстро устаревают за счет ускоренных темпов развития компьютерных технологий. В результате, лексикографам приходится активно обновлять не только сами словарные базы, но и компьютерные программы.

2. Кроме того, сегодня назрела необходимость создавать «глобальные» словари, объединяющие многочисленные исследования ученых и научных школ. Проблема заключается в сложности объединения словарных баз различных периодов истории языка в связи с их огромными размерами. Поэтому такая работа требует дополнительного финансирования, а также помощи со стороны государства.

3. В современных условиях создание электронных словарей может стать средством коммерциализации, что превратит компьютерную лексикографию в средство получения прибыли. Чтобы этого не произошло, и такие работы были востребованы, важно выработать у потребителя необходимость в использовании словарей.

Несмотря на вышеперечисленные недостатки, компьютерный словарь сможет вытеснить печатные издания в ближайшее время. Поэтому разработка оптимальных в использовании, полных, универсальных электронных словарей — насущная задача современной компьютерной лексикографии.

1.3.3 Способы разработки электронных словарей

Создание словаря является трудоемким процессом. Существует несколько подходов к составлению электронных словарей. Основными являются оцифровка какого-либо бумажного словаря либо создание словаря с нуля.

В настоящее время в России наиболее распространены следующие электронные словари: Lingvo, МультиЛекс, Мультитран, Polyglossum и другие. Несмотря на свою универсальность, эти словари обладают уникальными особенностями.

Lingvo [21] — один из самых известных электронных словарей на просторах российского Интернета, который заслуженно получил популярность среди своих пользователей за актуальность и новизну лексической базы, удобство и скорость в использовании. Разработчик – компания Abbyu дополняет новые версии словаря современной лексикой, а также вносит правки и исправления. Именно поэтому словарь Lingvo близок к языковой практике. Что интересно, Lingvo не является точным аналогом печатных версий словарей. Для создания узконаправленного словаря, например, Lingvo Economics, используется несколько словарей разных авторов. Еще один немаловажный факт — основная лексика в Abbyu Lingvo озвучивается диктором с классическим английским произношением.

МультиЛекс [6] — электронный словарь, созданный на основе оцифрованных копий известных печатных изданий современности. Словарь обладает многогранными средствами поиска, а также предоставляет объемную словарную информацию (транскрипцию, ударения, объяснения значений, варианты употребления). В МультиЛекс встроен синтезатор звука, который воспроизводит произношение слов. Однако существует вероятность искажения произношения, поэтому необходимо проверять транскрипцию слова. К тому же, МультиЛекс привязан к бумажному оригина-

лу, что лишает возможности вносить исправления и пополнять электронный словарь.

Мультитран [7] — база электронных двуязычных словарей, являющаяся одной из наиболее полных в России. Мультитран создан путем оцифровки большого количества печатных изданий словарей и их объединения в электронном формате. Более того, в нем предусмотрена функция усовершенствования базы. Каждый зарегистрированный пользователь может добавить на сайт свои термины или варианты перевода, а также оставить сообщение об обнаруженной ошибке, что значительным образом упрощает процесс внесения корректировок и обновления.

Polyglossum [22] — база электронных словарей, оснащенная современной поисковой системой для осуществления быстрого и полномасштабного поиска в интересующем направлении. Эта база содержит электронные словари, созданные на основе бумажных изданий, с возможностью внесения правок и без них, а также созданные с нуля электронные словари. Polyglossum подходит инженерам, ученым и техническим переводчикам для осуществления их профессионального труда. Эта база обладает большим количеством технических и отраслевых словарей.

Создание словаря с нуля возможно на основе разных источников. Одним из них является онтология предметной области как способ представления знаний.

Ключевым определением онтологии в этом значении является определение американского ученого Томаса Грубера: «явная спецификация концептуализации», где концептуализация понимается как «абстрактная, упрощенная модель мира, представляемая нами с определенной целью» [20; Т. 5].

Онтология содержит несколько элементов, основными из которых являются концепты, отношения и атрибуты. Под концептами понимаются концептуализации всех представителей какой-либо сущности или явления. Они могут содержать дочерние концепты, описывающие более конкретные

сущности или явления. Отношения устанавливают связи между концептами, в то время как атрибуты определяют свойства концептов [12].

Под онтологией предметной области, в свою очередь, понимается такая онтология, в которой представлены знания определенной предметной области: например, медицины, авиации, ИТ, лингвистики и т. д. Такие онтологии могут повторно применяться для различных задач в рамках предметной области, для которой они были созданы. Среди онтологий предметной области можно выделить базовые онтологии, содержащие основные, наиболее общие концепты предметной области, и узконаправленные онтологии, ориентированные на отдельные ее аспекты [16; Т. 14].

Онтология предметной области «Терроризм», используемая в качестве основания для разработки электронного словаря во второй главе настоящей работы, является базовой, поскольку она содержит основные концепты предметной области: «Тип теракта», «Объект теракта», «Исполнитель теракта», «Террористическая организация», «Оружие» и т. д. [23].

К одному концепту может присоединяться много слов и словосочетаний, которые подходят по семантическому наполнению. Например, концепт «террорист» — человек, осуществляющий террористическую деятельность; связанные с ним лексемы: террорист, боевик, нападавший, стрелок и т.д.

Таким образом, на основе онтологии можно разработать словарь, в котором лексика будет поделена на категории, важные для предметной области.

ВЫВОДЫ ПО ГЛАВЕ 1

Таким образом, в пункте 1.1 мы вывели обобщенный вариант понятия словаря на основе определений словаря из разных источников, рассмотрели основные периоды развития и функции лексикографии, виды словарной работы, а также роль словаря в лексикографии.

В пункте 1.2 представлена классификация типов словарей на основе различных принципов и параметров.

В пункте 1.3 характеризуются электронные словари, рассматриваются их преимущества и недостатки в сравнении с печатными экземплярами, а также освещаются способы разработки электронных словарей на примере наиболее популярных электронных словарей России.

ГЛАВА 2. ДАННЫЕ ДЛЯ РАЗРАБОТКИ ЭЛЕКТРОННОГО ФРАНКО-АНГЛО-РУССКОГО СЛОВАРЯ ПРЕДМЕТНОЙ ОБЛАСТИ «ТЕРРОРИЗМ»

2.1 Сбор и подготовка материала

Настоящее исследование базируется на онтологическом ресурсе предметной области «Терроризм», описанном в работе С.О. Шереметьевой и А.Ю. Зиновьевой [23]. В работе, на которую мы опираемся, описывается разработка онтологического ресурса предметной области «Терроризм» на материале новостных сообщений о террористической деятельности. Однако стоит отметить, что этот материал является довольно широким, нам же хотелось рассмотреть предметную область более конкретно, остановившись именно на терактах, вследствие чего в качестве материала данного исследования нами были выбраны псевдопараллельные (сопоставимые) тексты статей из открытой интернет-энциклопедии Wikipedia, из которых было составлено три одноязычных корпуса объемом в 30 000 словоупотреблений каждый: «Стрельба в мечетях Крайстчерча», «Стрельба в синагоге Питтсбурга», «Террористический акт в Беслане», «Террористические акты 11 сентября 2001 года», «Теракт на Дубровке», “Christchurch mosque shootings”, “Pittsburgh synagogue shooting”, “Beslan school siege”, “September 11 attacks”, “Moscow theater hostage crisis”, “Attentats de Christchurch”, “Fusillade dans une synagogue de Pittsburgh”, “Prise d’otages de Beslan”, “Attentats du 11 septembre 2001”, “Pris d’otages du théâtre de Moscou”.

В качестве языков исследования выбраны русский, английский и французский, поскольку они покрывают довольно большое количество текстов СМИ. В частности, английский язык является языком международного общения, поэтому многие новости дублируются на английском даже в тех странах, где он не является официальным. Французский в свою очередь является официальным языком Франции, одним из официальных

языков ООН, языком франкоязычного населения Бельгии, Швейцарии, Канады, Африки, Карибского бассейна, Французской Гвианы и т. д.; при этом в Африке совершается значительное количество терактов. Наконец, русский язык является языком, используемым на территории бывшего СССР, где теракты также нередки. Соответственно, на каждом из этих языков ежедневно генерируется огромное количество материалов для перевода.

Онтология была выбрана в качестве основы исследования для того, чтобы обеспечить разделение слов в словаре на значимые категории, которые упростили бы для переводчика процесс ориентирования в нашем словаре. Предложенная в работе С.О. Шереметьевой и А.Ю. Зиновьевой онтология является очень обширной и содержит множество концептов. Такое количество категорий, на наш взгляд, только запутало бы переводчика, поэтому мы решили отобрать концепты, являющиеся наиболее важными и употребимыми в предметной области, и сформировать на их основе следующие категории, показывающие, на наш взгляд, наиболее полную вариативность аспектов предметной области:

- *Террорист* — лицо, осуществляющее террористическую деятельность в составе террористической группы/организации, либо в одиночку.
- *Террористическая организация* — группа людей, осуществляющая террористическую деятельность.
- *Объект теракта* — люди, имущество, объекты культурного значения, в отношении которых производится террористический акт.
- *Средство теракта* — объект или событие, использованное для осуществления террористического действия.
- *Тип теракта* — вид нанесенного ущерба в связи с использованием определенных средств теракта.
- *Последствия* — причинение опасности жизни и здоровью людей, а также нанесение материального ущерба.

- *Цель теракта* — задачи, на выполнение которых нацелена террористическая деятельность.
- *Другая террористическая деятельность* — любая другая деятельность, совершаемая террористом или террористической организацией вне норм закона.
- *Антитеррористическая деятельность* — применение военной тактики, методов и стратегии, которые используются правительственными, военными, правоохранительными, коммерческими и разведывательными органами для борьбы с терроризмом или его предотвращения.
- *Реакция общественности* — ответ социума на последствия террористической деятельности, в частности терактов.

2.2 Подготовка частотных списков

Собрав трехязычный корпус текстов, общий объем которого составил примерно 90 000 словоупотреблений, мы пришли к выводу о необходимости его минимизации, чтобы сделать корпус текстов пригодным для дальнейшего анализа. С этой целью нами были составлены частотные списки каждого подкорпуса при помощи базового функционала автоматического экстрактора ключевых слов Lana-Key [17; Т. 12]. Частотные списки состояли из слов и словосочетаний, которые посредством были извлечены из текста при помощи данной программы с указанием частоты встречаемости. При обработке мы не ставили целью разделение глагольных и именных групп, поэтому в результирующем списке были как существительные, так и прилагательные, глаголы и предложные группы.

Анализ полученных частотных списков осуществлялся в несколько этапов.

1. Сокращение лексического материала.

Важной задачей на начальном этапе работы со словарем было его сокращение, поскольку полученные списки были слишком велики и неин-

формативны. Так, русскоязычный список состоял из 36 тысяч лексических единиц, большая часть которых имела частоту встречаемости 1, английский и французский список в свою очередь состояли из 53 и 52 тысяч единиц соответственно. Считается, что лаконичность словаря прямо пропорциональна его информативности. Как полагают Н. Н. Петрушевская и Л. Н. Беляева, при составлении словаря определенной тематики нужно учитывать не все слова, а только 85%, имеющие наивысший процент покрытия текста [1; Т. 2; 10], то есть 85% процентов частотного словаря, отсортированного по убыванию частоты. Следуя этому принципу, на данном этапе работы мы оставили только 85% лексики, поскольку оставшиеся 15% имеют очень низкую частоту встречаемости и, следовательно, малополезны для словаря. Благодаря этому мы оставили только те слова и словосочетания, которые являются максимально подходящими для нашего словаря.

2. *Удаление неправильно выделенных лексических единиц и стоп-слов.*

Результатом предыдущего этапа стали списки высокочастотных лексических единиц, однако не все они были релевантными, в списках содержалось много слов, которые обычно называют стоп-словами (для французского и английского языка это, в частности, артикли и предлоги; для русского — местоимения). Кроме того, часть последовательностей были выделены неправильно, и, следовательно, они не могут считаться правильными лексическими единицами: например, в русскоязычном списке были такие последовательности, как *в были, также к, по и, с одна* и др. В англоязычном списке к таким последовательностям можно отнести *of the, to the, by the*; во французском — *à la, par les, du septembre, World Trade* (незаконченная именная группа от *World Trade Center*) и др.

Чтобы избавиться от неправильно выделенных лексических единиц и стоп-слов, мы вручную произвели выборку слов из частотных списков, в результате чего все неподходящие лексические единицы были удалены.

3. Удаление имен собственных.

Мы приняли решение о необходимости удалять большинство имен собственных, за исключением наименований террористических организаций, поскольку наименования (в частности, названия объектов терактов и имена террористов) относят нас к определенным событиям, в то время как мы хотели сделать словарь универсальным в рамках предметной области. Таким образом, проведя дополнительный анализ списков лексических единиц, мы удалили из них имена террористов, их жертв, названия стран и конкретных объектов террористических актов.

4. Удаление нерелевантных лексических единиц.

Этот шаг, ставший одним из наиболее трудоемких в настоящей работе, заключался в том, чтобы удалить из полученных частотных списков нерелевантные для предметной области лексические единицы. Релевантность лексических единиц определялась на основе выделенных нами ранее девяти категорий.

На этом этапе мы столкнулись с некоторыми проблемами.

Во-первых, как оказалось, некоторые слова могут относиться одновременно к разным категориям. Сама тема «Терроризм» является довольно узкой, и авторы часто используют одинаковые слова, но это не означает их схожесть по контексту. Например, нам встретилась такая лексема, как *осужденный*. Было не ясно, в какую категорию ее определить: ПОСЛЕДСТВИЯ или РЕАКЦИЯ ОБЩЕСТВЕННОСТИ. Так мы пришли к выводу о том, что если *осужденный* — это субстантивированное прилагательное, то оно относится к категории ПОСЛЕДСТВИЯ, а если это форма глагола *осудить* — то к категории РЕАКЦИЯ ОБЩЕСТВЕННОСТИ.

То же самое можно сказать и о лексеме *подозреваемый*. Если это субстантивированное прилагательное, то оно относится к категории ПОСЛЕДСТВИЯ, если форма глагола *подозревать* — к категории РЕАКЦИЯ ОБЩЕСТВЕННОСТИ.

Также в английском списке нам встретилось слово *hostage* ‘заложник’ (и его аналог во французском *otage*), которое тоже создало некоторые трудности в понимании и неясность, к какой категории его отнести. Здесь все зависит от того, что мы считаем терактом. Если сам факт нападения, то заложники — это ПОСЛЕДСТВИЯ. Но мы понимаем данную лексему как осаду с заложниками и относим к категории ТИП ТЕРАКТА, в результате которого заложники либо погибнут, либо будут освобождены.

Во-вторых, мы столкнулись с проблемой лексической многозначности слова и частеречной омонимии. Так, слово *shot* несколько раз используется в текстах, но в одном случае оно использовано как существительное *shot* ‘выстрел’, а в другом это форма страдательного причастия от глагола *shoot* ‘стрелять’.

Благодаря работе с контекстом, нам удалось успешно решить данные проблемы и правильно определить, в какую группу должно попасть слово. тщательно проверили каждое слово из наших частотных списков по контексту, удалили нерелевантные и проверили встречаемость слов для каждой категории. Трехязычные списки лексических единиц с указанием частоты встречаемости, полученные в результате этого этапа, представлены в Приложении А.

5. Сопоставление лексики разных языков.

В целях сопоставления лексики разных языков мы взяли русский частотный список за основу как самый длинный, после чего к каждому слову в нем нашли эквиваленты на английском и французском языках. Когда русский частотный список закончился, мы продолжили искать эквиваленты к оставшимся английским, а затем французским лексемам.

Здесь мы также встретились с некоторыми трудностями. Во-первых, некоторые лексические единицы совсем не имели эквивалентов на других языках, некоторые — только на одном из языков. Особо явно это прослеживается во французском частотном списке. Мы полагаем, что это связано с тем, что подкорпусы разных языков оказались довольно далеки друг

от друга в содержательном плане, несмотря на то, что являлись и сопоставимыми (псевдопараллельными). Вполне возможно, что во французских энциклопедических статьях некоторые теракты освещены более кратко, потому что они происходили не во Франции или франкоговорящих странах. Поэтому отсутствующие эквиваленты мы заменили собственным переводом.

Во-вторых, в некоторых случаях лексическая единица, напротив, имела несколько эквивалентов. Так, в английском списке присутствовало слово *rifle*, которое может переводиться как *винтовка* и как *автомат*. При этом в русском списке присутствовали об этих слова, и мы приняли решение обоим словам поставить в соответствие английское слово *rifle*. Похожая ситуация наблюдалась с русскими глаголами совершенного и несовершенного вида, отличающиеся отсутствием или наличием приставки, например *требовать* и *потребовать*, которым в обоих случаях соответствует английской глагол *to demand*. Кроме того, в русскоязычном списке часто встречались синонимичные глаголы, соответствующие одному английскому или французскому: например, *погибнуть* и *скончаться* соответствуют английскому глаголу *to die*. Конечно, в подобных случаях мы могли бы поставить в соответствие глаголу *погибнуть* глагол *to die*, а для глагола *скончаться* подобрать другой эквивалент, отсутствующий в корпусе, например *to perish*, но такие действия нам кажутся излишними. На наш взгляд, прибегать к переводу лексических единиц при сравнении частотных списков сопоставимых корпусов следует только в крайних случаях (в частности, когда эквивалента совсем нет, как в предыдущей рассмотренной нами проблеме), иначе теряется смысл использования таких корпусов.

После проведения вышеперечисленных манипуляций со списками у нас осталось 136 лексических единиц на каждом из языков, которые являются наиболее актуальными для предметной области и, следовательно, могут стать основой трехязычного электронного словаря. Данные лекс и-

ческие единицы, как уже было сказано ранее, отнесены нами к различным категориям:

- ТЕРРОРИСТ: 7 лексических единиц;
- ТЕРРОРИСТИЧЕСКАЯ ОРГАНИЗАЦИЯ: 0 лексических единиц;
- ОБЪЕКТ ТЕРАКТА: 39 лексических единиц;
- СРЕДСТВО ТЕРАКТА: 18 лексических единиц;
- ТИП ТЕРАКТА: 11 лексических единиц;
- ПОСЛЕДСТВИЯ: 25 лексических единиц;
- ЦЕЛЬ ТЕРАКТА: 1 лексическая единица;
- ДРУГАЯ ТЕРРОРИСТИЧЕСКАЯ ДЕЯТЕЛЬНОСТЬ: 3 лексические единицы;
- АНТИТЕРРОРИСТИЧЕСКАЯ ДЕЯТЕЛЬНОСТЬ: 23 лексических единицы;
- РЕАКЦИЯ ОБЩЕСТВЕННОСТИ: 10 лексических единиц.

По результатам анализа частотных списков можно сделать следующие выводы.

1. Самой объемной получилась категория ОБЪЕКТ ТЕРАКТА. Именно в этой категории описываются объекты и предметы, в отношении которых осуществляется террористический акт. Далее в списке следует категория ПОСЛЕДСТВИЯ. Большой объем данной категории связан с большим количеством слов, которые описывают результаты террористических действий. На третьем месте расположена категория АНТИТЕРРОРИСТИЧЕСКАЯ ДЕЯТЕЛЬНОСТЬ. Она освещает способы борьбы с террористами. Крайне маленькими являются категории ДРУГАЯ ТЕРРОРИСТИЧЕСКАЯ ДЕЯТЕЛЬНОСТЬ и ЦЕЛЬ ТЕРАКТА. Другая террористическая деятельность, вероятно, не получила широкого освещения в рассмотренном корпусе вследствие того, что содержащиеся в нем тексты

описывают именно теракты, а не какую-либо другую террористическую деятельность (финансирование, рекрутирование и т. д.). Следовательно, можно сказать, что мы успешно справились с задачей выбора материала исследования. Категория ЦЕЛЬ ТЕРАКТА в свою очередь может являться малопредставленной по двум причинам: во-первых, часто террористы прямо не декларируют цель теракта, при этом она может быть выведена из курса деятельности террористической организации; во-вторых, цель теракта может быть представлена не одним словом, а словосочетанием, предложением или даже целым абзацем, из-за чего она просто могла не попасть в частотный список. С другой стороны, цели, которые формулируются в несколько словосочетаний или предложений, вряд ли могут являться предметом лексикографии.

Категория ТЕРРОРИСТИЧЕСКАЯ ОРГАНИЗАЦИЯ оказалась вообще не представлена в частотных списках. Возможными причинами могут быть следующие:

- 1) отсутствие упоминаний террористических организаций в корпусе (маловероятно);
- 2) малая частота встречаемости конкретных словоформ;
- 3) количество компонентов в наименованиях террористических организаций более четырех, из-за чего они не попали в частотные списки.

2. Чаще всего глаголы были представлены в формах прошедшего времени и в страдательном залоге. Например, английский глагол *to arrest* употреблялся в четырех формах: *arrested* (простое прошедшее время, действительный залог), его омоформа *arrested* (причастие прошедшего времени, страдательный залог), *was arrested* и *had been arrested*. Для французского же языка более характерно прошедшее завершённое время (*passé composé*). Глаголы в настоящем и будущем времени на всех трех языках встречаются довольно редко. Это, весьма вероятно, обусловлено тем, что речь в таких текстах идет о прошедших событиях.

2.3 Прототип электронного словаря

На основе полученного на предыдущем этапе трехязычного частотного списка, представленного в Приложении Б, мы создали прототип электронного словаря. Он представляет собой набор категорий, каждая из которых состоит из списка эквивалентных лексических единиц на трех языках с указанием частоты встречаемости (в случае, если единица является переводной, ее частота в корпусе принимается равной нулю). Для существительных на русском и французском языке для удобства переводчика мы указали род, а для глаголов — наиболее часто встречающиеся формы употребления в корпусе. Примеры категорий с лексическим наполнением представлены в таблицах ниже.

Таблица 1 – Лексическое наполнение категории «Террорист»

Категория	Террорист
Определение	Лицо, осуществляющее террористическую деятельность в составе террористической группы/организации, либо в одиночку
Лексема 1	захватчики [5] – hostage-taker [18] - envahisseurs[0]
Лексема 2	attacker 10 – атакующий 0 – assaillant 5
Лексема 3	shooter 10 – стрелок 0 – tireur 6
Лексема 4	hijacker 42 – угонщик 0 - pirate de l'air 0
Лексема 5	meurtrier 5 – убийца 0 - murderer 0

Таблица 2 – Лексическое наполнение категории «Последствия»

Категория	Последствия
Определение	Причинение опасности жизни и здоровью людей, а также нанесение материального ущерба
Лексема 1	заложник [116] – hostage [61] – otage [27]
Лексема 2	скончаться [9] – die [30] – mourir [0]
Лексема 3	arrested 14 - арестован 10 - détenu 0
Лексема 4	injured 13 - пострадавшие 11 - victime 40
Лексема 5	dommage 10 - ущерб 0 - damage 10

Таблица 3 – Лексическое наполнение категории «Объект теракта»

Категория	Объект теракта
Определение	Люди, имущество, объекты культурного значения, в отношении которых производится террористический акт

Продолжение таблицы 3

Лексема 1	школа [46] – school [59] - école [10]
Лексема 2	человек 64 – human 8 – personne 81 / homme 8
Лексема 3	building 46 – здания 93 – bâtiment 20 / immeuble 17
Лексема 4	family 15 – семьи 15 – famille 13
Лексема 5	citoyen 7 – граждане 11 – citizen 10

Таблица 4 – Лексическое наполнение категории «Средства теракта»

Категория	Средства теракта
Определение	Объект или событие, использованное для осуществления действия
Лексема 1	винтовка [6] – rifle [6] – fusil [0]
Лексема 2	взрывное устройство 8 - explosive device 0 - engin explosif 0
Лексема 3	bomb 6 – взрывчатка 11 – bombe 7
Лексема 4	car 13 – автомобиль 0 – voiture 0
Лексема 5	gaz 6 – газ 0 – gaz 0

Таблица 5 – Лексическое наполнение категории «Тип теракта»

Категория	Тип теракта
Определение	Вид нанесенного ущерба в связи с использованием определенных средств
Лексема 1	атака [12] – attack [174] – attaque [33]
Лексема 2	расстрел [9] – shooting [28] – fusillade [0]
Лексема 3	explosion 15 – взрыв 0 – explosion 6
Лексема 4	crash 12 – врезаться 6 – heurter 0
Лексема 5	massacre 6 – резня 0 – massacre 0

Таблица 6 – Лексическое наполнение категории «Антитеррористическая деятельность»

Категория	Антитеррористическая деятельность
Определение	Применение военной тактики, методов и стратегии, которые используются правительственными, военными, правоохранительными, коммерческими и разведывательными органами для борьбы с терроризмом или его предотвращения
Лексема 1	разговор[5] – conversation [6] - conversation [0]
Лексема 2	переговоры 15 – negotiation 5 – négociation 0
Лексема 3	rescue 13 – освобождена 21 – secours 9
Лексема 4	construction 5 – расследование 16 – investigation 0
Лексема 5	résistance 14 – сопротивление 0 – resistance 0

Таблица 7 - Лексическое наполнение категории «Другая террористическая деятельность»

Категория	Другая террористическая деятельность
Определение	Любое действие, совершенное против норм закона
Лексема 1	требовать [10] – demand [0] - requérir [0]
Лексема 2	потребовать 7 - demand [0] - requérir [0]
Лексема 3	conspiracy 10 – заговор 0 – complot 11

Таблица 8 - Лексическое наполнение категории «Реакция общественности»

Категория	Реакция общественности
Определение	Ответ социума на последствия девиантного поведения
Лексема 1	обвинен [14] – condemned [11] – condamne [7]
Лексема 2	память 7 – remembrance 0 – hommage 5
Лексема 3	memorial 5 – мемориал 0 – mémorial 0
Лексема 4	sentenced 5 - заключенный 0 – condamné 0
Лексема 5	justice 8 – справедливость 0 – justice 0

ВЫВОДЫ ПО ГЛАВЕ 2

В данной главе описана разработка франко-англо-русского электронного словаря на основе онтологической модели предметной области «Терроризм». В целях сбора данных для словаря были использованы псевдопараллельные корпуса энциклопедических текстов предметной области на русском, английском и французском языках. Онтология была выбрана в качестве основы исследования для разделения слов в словаре на значимые категории, что должно значительно упростить использование нашего словаря переводчиком.

Вначале мы прибегли к минимизации корпусов текстов для упрощения их анализа в дальнейшем, получив, таким образом, частотные списки. Затем мы выполнили чистку полученных списков, избавившись от стоп-слов, неправильно извлеченных и нерелевантных для предметной области лексических единиц. Лексические единицы из очищенных списков были разнесены по заранее отобраным категориям и сопоставлены друг с другом. В случае отсутствия эквивалентов в корпусе на каком-либо из языков мы подбирали эквивалент самостоятельно.

Затем нами были проведены подсчеты количества словоупотреблений в каждой категории, выявлены наиболее объемные категории, рассмотрены наиболее часто используемые формы глаголов, описаны некоторые сложности, с которыми мы столкнулись в процессе работы. После этого мы продемонстрировали прототип электронного словаря и описали его.

В данной главе мы прибегли к количественному и сравнительно-сопоставительному анализу текстовых корпусов на русском, английском и французском языках для разработки данных электронного словаря.

Полученные данные фактически представляют собой наполнение электронного трехязычного словаря предметной области «Терроризм» и могут использоваться переводчиками в качестве справочного материала, а также при обучении переводу. В будущем также можно будет создать

программную оболочку словаря, чтобы поместить в нее полученные данные, что сделает процесс пользования словарем намного удобнее.

ЗАКЛЮЧЕНИЕ

Мы живем в эру ускоренного прогресса, где скорость поиска информации играет немаловажную роль. Люди привыкли активно пользоваться продуктами информационных технологий. Еще недавно популярные бумажные словари утратили свою значимость в современном мире.

Сегодня целесообразно заменить печатные варианты словарей усовершенствованными, общедоступными электронными моделями. Более того, такие словари обладают рядом преимуществ по сравнению с печатными версиями. Например, высокая скорость поиска, удобство в использовании, актуальность информации, большой объем лексических единиц, варианты их употребления и перевода на другие языки.

Именно поэтому наше исследование заключается в поиске, обработке и анализе материалов для последующего составления электронного словаря предметной области на основе онтологии предметной области «Терроризм» для переводчиков.

В результате данного исследования нами был проделан большой объем работы. Мы рассмотрели понятие словаря, его функции и особенности. Затем охарактеризовали электронные словари, разобрались в их свойствах.

Следующим шагом был отбор материалов для составления псевдопараллельных (сопоставимых) корпусов из текстов энциклопедических статей, выделение значимых категорий словаря на основе онтологии предметной области «Терроризм».

После этого мы минимизировали полученные корпуса текстов, что сделало их более пригодными для анализа. Таким образом, в результате отбора лексики по частоте встречаемости в текстах, мы получили частотные списки.

Следующим пунктом нашей работы было сопоставление лексических единиц разных языков друг с другом, а также выделение грамматических и семантических особенностей данной лексики.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Беляева Л.Н. Извлечение лексикографической информации средствами информационных технологий / Л.Н. Беляева // Иностранные языки в дистанционном обучении: материалы III Международной научно-практической конференции (Пермь, 23–25 апреля 2009 г.). – Пермь: Изд-во Пермского гос. тех. ун-та, 2009. – Т. 2. – С. 151–158.
2. Большая российская энциклопедия : [в 35 т.] / гл. ред. Ю. С. Осипов. – М. : Большая российская энциклопедия, 2004–2017. // Большая российская энциклопедия. Том 30. – М., 2015. – С. 424–425.
3. Брокгауз и Ефрон: энциклопедия: в 86 томах с иллюстрациями и дополнительными материалами. – Электрон. дан. – М. : ИДДК, 2004.
4. Гак В.Г. Словарь // Лингвистический энциклопедический словарь / Главный редактор В.Н. Ярцева. – М.: Советская энциклопедия, 1990. – 685 с.
5. Даль В.И. Толковый словарь живого великорусского языка: избр. ст. / В. И. Даль; совмещ. ред. изд. В. И. Даля и И. А. Бодуэна де Куртенэ; [науч. ред. Л. В. Беловинский]. – М. : ОЛМА Медиа Групп, 2009. – 573 с.
6. МультиЛекс // электронный словарь: URL: <http://www.lingvo-plus.ru/electronnie-slovari/>
7. Мультитран // электронный словарь: URL: <https://www.multitrans.com>
8. Новый энциклопедический словарь. – Москва : Большая Российская энциклопедия : Рипол классик, 2006. – 1455 с.
9. Ожегов С.И. О трех типах толковых словарей русского языка // Ожегов С.И. Лексикология. Лексикография. Культура русской речи. Учебное пособие для вузов. – М., 1974. – С. 158–82.
10. Петрушевская Н.Н. Опыт лингвостатистического отбора лексики для обучения чтению в техническом вузе / Н.Н. Петрушевская // Ана-

лиз содержания курса иностранного языка. – Томск, 1976. – Вып. 3. – С. 48–50.

11. Рублева О.Л. Лексикология современного русского языка / О. Л. Рублева. – Владивосток: Изд-во Дальневосточного ун-та, 2004.

12. Соловьев В.Д., Добров Б.В., Иванов В.В., Лукашевич Н.В. Онтологии и тезаурусы: модели, инструменты, приложения: учебное пособие / – М.: Бином. Лаборатория знаний, 2013. – 176 с.

13. Толковый словарь русского языка: 72500 слов и 7500 фразеологических выражений / С. И. Ожегов Н. Ю. Шведова ; Российская АН, Ин-т рус. яз., Российский фонд культуры. - 2-е изд., испр. и доп. – Москва : Азъ, 1994. – 907 с.

14. Ушаков Д. Н. Большой толковый словарь русского языка: современная редакция / Д. Н. Ушаков. – Москва : Дом Славянской кн., 2008. – 959 с.

15. Фесенко О.П. Электронные словари как продукт современной лексикографии / О.П. Фесенко, С.С. Лаухина // Омский научный вестник. – 2015. – № 4 (141). – С. 46–48.

16. Шереметьева С.О. К вопросу о разработке онтологической модели предметной области «терроризм» / С. О. Шереметьева, А. Ю. Зиновьева // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. – 2017. – Т. 14, № 4. – С. 48–54.

17. Шереметьева С.О. Методы и модели автоматического извлечения ключевых слов / С.О. Шереметьева, П.Г. Осминин // Вестник ЮУрГУ. Серия Лингвистика. – 2015. – Т. 12 (1). – С. 76–81.

18. Щерба Л.В. Опыт общей теории лексикографии. Основные типы словарей // Изв. АН СССР. ОЛЯ. – 1940. № 3. – С. 103–110.

19. Энциклопедический словарь Русского библиографического института Гранат. – 11-е стереотип. изд. – Москва : Рус. библиогр. инст. Гранат, 1940–1941.

20. Gruber T. R. A Translation Approach to Portable Ontology Specifications / T. R. Gruber // Knowledge Acquisition. – 1993. – Vol. 5 (2). – P. 199–220
21. Lingvo // электронный словарь: URL: <https://www.lingvo.ru>
22. Polyglossum // электронный словарь: URL: <http://www.ets.ru/pg/pg/ru/index.htm>
23. Sheremetyeva S. On Modelling Domain Ontology Knowledge for Processing Multilingual Texts of Terroristic Content / S. Sheremetyeva, A. Zinovyeva // Proceedings of the Third International Conference Digital Transformation and Global Society. – Springer, 2018. – P. 368–379.

ПРИЛОЖЕНИЕ А

Список русскоязычных лексических единиц

Лексическая единица	Частота	Категория
заложник	116	последствия
здания	93	объект теракта
человек	64	объект теракта
школа	46	объект теракта
погиб	44	последствия
захват	42	антитеррористическая деятельность
прорыв	36	антитеррористическая деятельность
мечеть	31	объект теракта
башня	30	объект теракта
освобождена	21	антитеррористическая деятельность
расследование	16	антитеррористическая деятельность
территория	16	объект теракта
разрушены	16	последствия
переговоры	15	антитеррористическая деятельность
помещение	15	объект теракта
семьи	15	объект теракта
обвинен	14	реакция общественности
подозреваемый	14	последствия
арест	13	антитеррористическая деятельность
осада	13	антитеррористическая деятельность
атака	12	тип теракта
пострадавшие	11	последствия
взрывчатка	11	средство теракта
граждане	11	объект теракта
районы	11	объект теракта
захват заложников	10	последствия
арестован	10	последствия
выстрел	10	тип теракта
освобождение заложников	10	антитеррористическая деятельность
требовал	10	другая террористическая деятельность
автомат	9	средство теракта
был арестован	9	антитеррористическая деятельность
расстрел	9	тип теракта
скончался	9	последствия
взрывное устройство	8	средство теракта
в результате теракта	7	последствия
память	7	реакция общественности
задержан	7	последствия
потребовал	7	другая террористическая деятельность
судебный процесс	6	антитеррористическая деятельность
винтовка	6	средство теракта

врезался	6	тип теракта
гранатомёт	6	средство теракта
иностранцы	6	объект теракта
огнестрельное оружие	6	средство теракта
приговор	6	антитеррористическая деятельность
снаряд	6	средство теракта
убили	6	последствия
в ходе штурма	5	последствия
захватчики	5	террорист
застрелен	5	последствия
захват театрального центра	5	тип теракта
огнемёты	5	средство теракта
опознан	5	последствия
осудил	5	реакция общественности
переговоры с террористами	5	антитеррористическая деятельность
предложения	5	антитеррористическая деятельность
разговор	5	антитеррористическая деятельность
расстреливал	5	тип теракта
сопротивление	5	антитеррористическая деятельность/последствия
тяжело ранен	5	последствия
эвакуация заложников	5	антитеррористическая деятельность

Список англоязычных лексических единиц

Лексическая единица	Частота	Категория
attack	174	тип теракта
people	95	объект теракта
hostage	61	последствия
school	59	объект теракта
building	46	объект теракта
World Trade Center	42	объект теракта
hijacker	42	террорист
mosque	42	объект теракта
killed	40	последствия
victim	40	последствия
child	37	объект теракта
died	30	последствия
shooting	28	тип теракта
support	26	реакция общественности
death	23	последствия
tower	21	объект теракта
shot	20	тип теракта
passenger	19	объект теракта
hostage-taker	18	террорист
explosion	15	тип теракта
family	15	объект теракта
murder	15	тип теракта

arrested	14	последствия
suspected	14	последствия
car	13	средства теракта
injured	13	последствия
rescue	13	антитерроритическая деятельность
aircraft	12	объект теракта
crashed	12	тип теракта
gunman	12	террорист
station	12	объект теракта
House	11	объект теракта
assault	11	антитерроритическая деятельность
collapse	11	последствия
collapsed	11	последствия
condemned	11	реакция общественности
hall	11	объект теракта
worker	11	объект теракта
Airport	10	объект теракта
attacker	10	террорист
citizen	10	объект теракта
conspiracy	10	другая террористическая деятельность
damage	10	последствия
shooter	10	террорист
weapon	10	средства теракта
were killed	10	последствия
crew	9	объект теракта
vehicle	9	объект теракта
District	8	объект теракта
battle	8	антитерроритическая деятельность
death toll	8	последствия
gun	8	средства теракта
human	8	объект теракта
kill	8	последствия
theatre	8	объект теракта
to prevent	8	антитерроритическая деятельность
Defense	7	антитерроритическая деятельность
fell	7	последствия
injury	7	последствия
synagogue	7	объект теракта
worshipper	7	объект теракта
Boeing	6	средства теракта
basement	6	объект теракта
bomb	6	средства теракта
bombing	6	тип теракта
conversation	6	антитерроритическая деятельность
negotiation	6	антитерроритическая деятельность
rifle	6	средства теракта

suicide	6	последствия
tank	6	средства теракта
was arrested	6	последствия
Church	5	объект теракта
High school	5	объект теракта
aviation	5	объект теракта
condemned the attack	5	реакция общественности
construction	5	антитерроритическая деятельность
genocide	5	цели
grenade	5	средства теракта
international Airport	5	объект теракта
memorial	5	реакция общественности
negotiate	5	антитерроритическая деятельность
penalty	5	реакция общественности
pilot	5	объект теракта
resolution	5	антитерроритическая деятельность
sentenced	5	реакция общественности
truck	5	средства теракта

Список франкоязычных лексических единиц

Лексическая единица	Частота	Категория
personne	81	объект теракта
attentat	79	террорист
victime	40	последствия
World Trade Center	36	объект теракта
attaque	33	тип теракта
avion	28	объект теракта
otage	27	последствия
bâtiment	20	объект теракта
immeuble	17	объект теракта
enfant	15	объект теракта
colonne	14	объект теракта
résistance	14	антитеррористическая деятельность
synagogue	14	объект теракта
théâtre	14	объект теракта
famille	13	объект теракта
Boeing	11	средства теракта
complot	11	другая террористическая деятельность
femme	11	объект теракта
ville	11	объект теракта
dommage	10	последствия
école	10	объект теракта
destruction	9	последствия
espace aérien	9	объект теракта

base	9	объект теракта
secours	9	антитеррористическая деятельность
Port	8	объект теракта
homme	8	объект теракта
justice	8	реакция общественности
bombe	7	средства теракта
chasseur	7	средства теракта
citoyen	7	объект теракта
condamne	7	реакция общественности
explosion	6	тип теракта
gaz	6	средства теракта
appareil	6	средства теракта
massacre	6	тип теракта
peine	6	реакция общественности
tireur	6	террорист
tuerie	6	тип теракта
assaillant	5	террорист
base aérienne	5	объект теракта
décombe	5	последствия
hommage	5	реакция общественности
meurtrier	5	террорист
protection	5	антитеррористическая деятельность
témoignage	5	реакция общественности

ПРИЛОЖЕНИЕ Б

Трехязычный список лексических единиц с разделением по категориям

РУССКИЕ ЛЕ	ЧАСТОТ А	КАТ .	АНГЛИЙСКИЕ ЛЕ	ЧАСТОТ А	КАТ .	ФРАНЦУЗСКИЕ ЛЕ	ЧАСТОТ А	КАТ .
<i>Авиационная база</i>	0	ОТ	<i>air base</i>	0	ОТ	<i>base aérienne</i>	5	ОТ
<i>авиация</i>	0	ОТ	<i>aviation</i>	5	ОТ	<i>aviation</i>	0	ОТ
автомат	9	СТ	<i>rifle</i>	6	СТ	<i>fusil</i>	0	СТ
<i>автомобиль</i>	0	СТ	<i>car</i>	13	СТ	<i>voiture</i>	0	СТ
арест	13	АД	<i>arrest</i>	0	АД	<i>arrestation</i>	0	АД
арестован	28	П	<i>arrested</i>	26	П	<i>arrêté</i>	4	П
арестован	19	АД	<i>arrested</i>	26	АД	<i>arrêté</i>	4	АД
атака	12	ТТ	<i>attack</i>	174	ТТ	<i>attaque/attentat</i>	33/79	ТТ
<i>аэропорт</i>	0	ОТ	<i>airport</i>	10	ОТ	<i>aéroport</i>	0	ОТ
<i>база</i>	0	ОТ	<i>base</i>	0	ОТ	<i>base</i>	9	ОТ
башня	30	ОТ	<i>tower</i>	21	ОТ	<i>tour</i>	0	ОТ
<i>боец</i>	0	СТ	<i>fighter</i>	0	СТ	<i>chasseur</i>	7	СТ
<i>Бомба</i>	0	СТ	<i>bomb</i>	6	СТ	<i>bombe</i>	7	СТ
в результате теракта	7	П	<i>as a result of a terrorist at- tack</i>	0	П	<i>à la suite de l'attaque terroriste</i>	0	П
в ходе штурма	5	П	<i>during the assault</i>	0	П	<i>pendant l'assaut</i>	0	П
<i>верующий</i>	0	ОТ	<i>worshipper</i>	7	ОТ	<i>worshipper</i>	0	ОТ
<i>Вести переговоры</i>	0	АД	<i>negotiate</i>	5	АД	<i>négociier</i>	0	АД
<i>взрыв</i>	0	ТТ	<i>explosion</i>	15	ТТ	<i>explosion</i>	6	ТТ
взрывное устройство	8	СТ	<i>explosive device</i>	0	СТ	<i>engin explosif</i>	0	СТ
взрывчатка	11	СТ	<i>bomb</i>	6	СТ	<i>bombe</i>	7	СТ
винтовка	6	СТ	<i>rifle</i>	6	СТ	<i>fusil</i>	0	СТ

РУССКИЕ ЛЕ	ЧАСТОТ А	КАТ .	АНГЛИЙСКИЕ ЛЕ	ЧАСТОТ А	КАТ .	ФРАНЦУЗСКИЕ ЛЕ	ЧАСТОТ А	КАТ .
<i>Воздушное пространство</i>	0	ОТ	<i>airspace</i>	0	ОТ	<i>espace aérien</i>	9	ОТ
<i>вооружённый преступник</i>	0	Т	<i>gunman</i>	12	Т	<i>tireur</i>	0	Т
врезаться	6	ТТ	<i>crash</i>	12	ТТ	<i>se heurter</i>	0	ТТ
выстрел	10	ТТ	<i>shot</i>	20	ТТ	<i>coup de feu</i>	0	ТТ
<i>газ</i>	0	СТ	<i>gaz</i>	0	СТ	<i>gaz</i>	6	СТ
<i>Геноцид</i>	0	Ц	<i>genocide</i>	5	Ц	<i>génocide</i>	0	Ц
<i>Город</i>	0	ОТ	<i>city</i>	0	ОТ	<i>ville</i>	11	ОТ
гражданин	11	ОТ	<i>citizen</i>	10	ОТ	<i>citoyen</i>	7	ОТ
<i>Граната</i>	0	СТ	<i>grenade</i>	5	СТ	<i>grenade</i>	0	СТ
гранатомёт	6	СТ	<i>grenade launcher</i>	0	СТ	<i>lanceur de grenades</i>	0	СТ
<i>грузовик</i>	0	СТ	<i>truck</i>	5	СТ	<i>camion</i>	0	СТ
<i>Дом</i>	0	ОТ	<i>house</i>	11	ОТ	<i>maison</i>	0	ОТ
<i>Женщина</i>	0	ОТ	<i>woman</i>	0	ОТ	<i>femme</i>	11	ОТ
<i>жертва</i>	0	П	<i>victim</i>	40	П	<i>victime</i>	0	П
<i>завалы</i>	0	П	<i>rubble</i>	0	П	<i>décombre</i>	5	П
<i>заговор</i>	0	ДТД	<i>conspiracy</i>	10	ДТД	<i>complot</i>	11	ДТД
задержан	7	П	<i>arrested</i>	14	П	<i>détenu</i>	0	П
<i>Заключенный</i>	0	РО	<i>sentenced</i>	5	РО	<i>condamné</i>	0	РО
заложник	116	П	<i>hostage</i>	61	П	<i>otage</i>	27	П
застрелен	5	П	<i>shot</i>	10	П	<i>abattu</i>	0	П
захват	42	АД	<i>capture</i>	0	АД	<i>piégeage</i>	0	АД
захват заложников	10	П	<i>hostage-taking</i>	0	П	<i>prise d'otages</i>	0	П
захват театрального центра	5	ТТ	<i>theatre</i>	0	ТТ	<i>capturer le théâtre</i>	0	ТТ
захватчик	5	Т	<i>hostage-taker</i>	18	Т	<i>envahisseur</i>	0	Т
<i>защита</i>	0	АД	<i>protection</i>	0	АД	<i>protection</i>	5	АД
здание	93	ОТ	<i>building</i>	46	ОТ	<i>bâtiment/immeuble</i>	20/17	ОТ

РУССКИЕ ЛЕ	ЧАСТОТ А	КАТ .	АНГЛИЙСКИЕ ЛЕ	ЧАСТОТ А	КАТ .	ФРАНЦУЗСКИЕ ЛЕ	ЧАСТОТ А	КАТ .
иностранец	6	ОТ	<i>foreigner</i>	0	ОТ	<i>étranger</i>	0	ОТ
колонна	0	ОТ	<i>column</i>	0	ОТ	<i>colonne</i>	14	ОТ
команда	0	ОТ	<i>crew</i>	9	ОТ	<i>équipage</i>	0	ОТ
крах	0	П	<i>collapse</i>	11	П	<i>effondrement</i>	0	П
люди	0	ОТ	<i>people</i>	95	ОТ	<i>gens</i>	0	ОТ
Массовое убийство	0	ТТ	<i>killling</i>	0	ТТ	<i>tuerie</i>	6	ТТ
Международный аэропорт	0	ОТ	<i>international airport</i>	5	ОТ	<i>aéroport international</i>	0	ОТ
Мемориал	0	РО	<i>memorial</i>	5	РО	<i>mémorial</i>	0	РО
мечеть	31	ОТ	<i>mosque</i>	42	ОТ	<i>mosquée</i>	0	ОТ
Мечеть	31	ОТ	<i>mosque</i>	42	ОТ	<i>mosquée</i>	0	ОТ
наказание	0	РО	<i>penalty</i>	5	РО	<i>pénalité</i>	0	РО
нападавший	0	Т	<i>attacker</i>	10	Т	<i>assaillant</i>	5	Т
нападающий	0	Т	<i>attacker</i>	10	Т	<i>assaillant</i>	5	Т
обвинен	14	РО	<i>condemned</i>	11	РО	<i>condamne</i>	7	РО
оборона	0	АД	<i>defense</i>	7	АД	<i>défense</i>	0	АД
огнемёты	5	СТ	<i>flamethrower</i>	0	СТ	<i>lance-flammes</i>	0	СТ
огнестрельное оружие	6	СТ	<i>firearm</i>	0	СТ	<i>armes à feu</i>	0	СТ
опознан	5	П	<i>identified</i>	0	П	<i>identifié</i>	0	П
оружие	0	СТ	<i>weapon</i>	10	СТ	<i>arme</i>	0	СТ
осада	13	АД	<i>siege</i>	0	АД	<i>siège</i>	0	АД
освободить	21	АД	<i>rescue</i>	13	АД	<i>libérer</i>	0	АД
освобождение заложников	10	АД	<i>hostage rescue</i>	0	АД	<i>libération des otages</i>	0	АД
осудить	5	РО	<i>condemn</i>	5	РО	<i>condamner</i>	0	РО
Осудить нападение	0	РО	<i>condemn the attack</i>	5	РО	<i>condamné l'attentat</i>	0	РО
память	7	РО	<i>remembrance</i>	0	РО	<i>hommage</i>	5	РО
Пассажир	0	ОТ	<i>passenger</i>	19	ОТ	<i>passager</i>	0	ОТ

РУССКИЕ ЛЕ	ЧАСТОТ А	КАТ .	АНГЛИЙСКИЕ ЛЕ	ЧАСТОТ А	КАТ .	ФРАНЦУЗСКИЕ ЛЕ	ЧАСТОТ А	КАТ .
переговоры	20	АД	negotiation	11	АД	<i>négociation</i>	0	АД
<i>Пилот</i>	0	ОТ	pilot	5	ОТ	<i>pilote</i>	0	ОТ
<i>пистолет</i>	0	СТ	gun	8	СТ	<i>arme à feu</i>	0	СТ
погибать	44	П	die	30	П	<i>mourir</i>	0	П
<i>Подвал</i>	0	ОТ	basement	6	ОТ	<i>sous-sol</i>	0	ОТ
<i>Поддержка</i>	0	РО	support	26	РО	<i>soutien</i>	0	РО
подозреваемый	14	П	suspected	14	П	<i>suspect</i>	0	П
помещение	15	ОТ	hall	11	ОТ	<i>investissement</i>	0	ОТ
<i>порт</i>	0	ОТ	<i>port</i>	0	ОТ	<i>port</i>	8	ОТ
пострадавший	11	П	injured	13	П	<i>victime</i>	40	П
потребовать	7	ДТД	<i>demand</i>	0	ДТД	<i>requérir</i>	0	ДТД
предложения	5	АД	<i>offers</i>	0	АД	<i>ouvertures</i>	0	АД
<i>предотвратить</i>	0	АД	to prevent	8	АД	<i>prévenir</i>	0	АД
<i>прибор</i>	0	СТ	<i>device</i>	0	СТ	<i>appareil</i>	6	СТ
приговор	6	АД	<i>sentence</i>	0	АД	<i>arrêt</i>	0	АД
прорыв	36	АД	<i>breakthrough</i>	0	АД	<i>percée</i>	0	АД
<i>работник</i>	0	ОТ	worker	11	ОТ	<i>travailleur</i>	0	ОТ
разговор	5	АД	conversation	6	АД	<i>conversation</i>	0	АД
<i>разрушение</i>	0	П	<i>destruction</i>	0	П	<i>destruction</i>	9	П
разрушены	16	П	collapsed	11	П	<i>ruiné</i>	0	П
районы	11	ОТ	district	8	ОТ	<i>arrondissement</i>	0	ОТ
расследование	16	АД	construction	5	АД	<i>investigation</i>	0	АД
расстрел	9	ТТ	shooting	28	ТТ	<i>fusillade</i>	0	ТТ
расстрелять	5	ТТ	<i>shoot</i>	0	ТТ	<i>fusiller</i>	0	ТТ
<i>ребенок</i>	0	ОТ	child	37	ОТ	<i>enfant</i>	15	ОТ
<i>резня</i>	0	ТТ	<i>massacre</i>	0	ТТ	<i>massacre</i>	6	ТТ

РУССКИЕ ЛЕ	ЧАСТОТ А	КАТ .	АНГЛИЙСКИЕ ЛЕ	ЧАСТОТ А	КАТ .	ФРАНЦУЗСКИЕ ЛЕ	ЧАСТОТ А	КАТ .
<i>Резолюция</i>	0	АД	resolution	5	АД	<i>résolution</i>	0	АД
<i>рухнуть</i>	0	П	fell	7	П	<i>tomber</i>	0	П
<i>самолет</i>	0	ОТ	aircraft	12	ОТ	avion	28	ОТ
<i>Свидетельские показания</i>	0	РО	<i>testimony</i>	0	РО	<i>témoignage</i>	5	РО
семья	15	ОТ	family	15	ОТ	famille	13	ОТ
<i>синагога</i>	0	ОТ	synagogue	7	ОТ	synagogue	14	ОТ
скончаться	9	П	die	30	П	<i>mourir</i>	0	П
<i>Смерть</i>	0	П	death	23	П	<i>décès</i>	0	П
снаряд	6	СТ	<i>projectile</i>	0	СТ	<i>projectile</i>	0	СТ
<i>Сопrotивление</i>	0	АД	<i>resistance</i>	0	АД	<i>résistance</i>	14	АД
<i>спасение</i>	0	АД	rescue	13	АД	secours	9	АД
<i>Список жертв</i>	0	П	death toll	8	П	<i>nombre de morts</i>	0	П
<i>справедливость</i>	0	РО	<i>justice</i>	0	РО	justice	8	РО
<i>Сражение</i>	0	АД	battle	8	АД	<i>bataille</i>	0	АД
<i>Станция</i>	0	ОТ	station	12	ОТ	<i>station</i>	0	ОТ
<i>Старшая школа</i>	0	ОТ	high school	5	ОТ	<i>lycée</i>	0	ОТ
<i>стрелок</i>	0	Т	shooter	10	Т	tireur	6	Т
судебный процесс	6	АД	<i>legal proceeding</i>	0	АД	<i>contentieux</i>	0	АД
<i>Суицид</i>	0	П	suicide	6	П	<i>suicide</i>	0	П
<i>Танк</i>	0	СТ	tank	6	СТ	<i>tank</i>	0	СТ
<i>Театр</i>	0	ОТ	theatre	8	ОТ	théâtre	14	ОТ
<i>теракт с использованием бомбы</i>	0	ТТ	bombing	6	ТТ	<i>attentat à la bombe</i>	0	ТТ
территория	16	ОТ	<i>territory</i>	0	ОТ	<i>territoire</i>	0	ОТ
<i>Травма</i>	0	П	injury	7	П	<i>blesure</i>	0	П
<i>Транспортное средство</i>	0	ОТ	vehicle	9	ОТ	<i>véhicule</i>	0	ОТ

РУССКИЕ ЛЕ	ЧАСТОТ А	КАТ .	АНГЛИЙСКИЕ ЛЕ	ЧАСТОТ А	КАТ .	ФРАНЦУЗСКИЕ ЛЕ	ЧАСТОТ А	КАТ .
требовать	10	ДТД	<i>demand</i>	0	ДТД	<i>requérir</i>	0	ДТД
тяжело ранен	5	П	<i>seriously injured</i>	0	П	<i>grièvement blessé</i>	0	П
<i>убийство</i>	0	ТТ	murder	15	ТТ	<i>meurtre</i>	0	ТТ
<i>убийца</i>	0	Т	<i>murderer</i>	0	Т	meurtrier	5	Т
убить	6	П	kill	40	П	<i>tuer</i>	0	П
<i>угонщик</i>	0	Т	hijacker	42	Т	<i>pirate de l'air</i>	0	Т
<i>ущерб</i>	0	П	damage	10	П	dommage	10	П
<i>Церковь</i>	0	ОТ	church	5	ОТ	<i>église</i>	0	ОТ
человек	64	ОТ	human	8	ОТ	personne/homme	81/8	ОТ
школа	46	ОТ	school	59	ОТ	école	10	ОТ
<i>штурм</i>	0	АД	assault	11	АД	<i>assault</i>	0	АД
эвакуация заложников	5	АД	<i>hostage evacuation</i>	0	АД	<i>évacuation des otages</i>	0	АД