

Правительство Российской Федерации
Государственное образовательное бюджетное учреждение высшего
профессионального образования
Национальный Исследовательский Университет
Высшая Школа Экономики
Факультет Информатики, математики и компьютерных наук
Образовательная программа Компьютерная лингвистика
Направление Прикладная математика и информатика
Код направления 01.04.02

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

На тему

Составление корпуса русско-английских эквивалентов интернет-коммуникации и методика работы с ним

Зав. кафедрой
доцент Асеева Н.В.

Выполнила:
Молокова С.В.
гр.13МАГКЛ
Научный руководитель:
кандидат физико-математических наук,
доцент
Карпов Н.В.

г. Нижний Новгород
2015г.

Оглавление

Введение.....	3
ГЛАВА 1. ПОДГОТОВКА К СОЗДАНИЮ КОРПУСА.....	6
1.1. Понятие корпуса.....	6
1.2. Особенности языка интернет-коммуникации.....	12
1.3. Обзор корпусов языка интернет-коммуникации.....	16
1.4. Понятие эквивалента.....	18
Выводы.....	22
ГЛАВА 2. МЕТОДИКА РАБОТЫ С КОРПУСОМ	23
2.1. Конкорданс.....	24
2.2. Скетчи и тезаурус.....	33
2.3. Сравнение скетчей.....	45
Выводы.....	56
ГЛАВА 3. СОСТАВЛЕНИЕ КОРПУСА.....	57
3.1. Описание собственного корпуса	57
3.2. Применение корпуса русско-английских эквивалентов интернет-коммуникации	62
3.3. Усовершенствование корпуса.....	64
Выводы.....	67
Заключение.....	68
Библиографический список.....	70

Введение

Язык интернет-коммуникации, **объект** рассмотрения в настоящей работе, как многогранный и сложный феномен остается одной из мало изученных областей современной лингвистики, несмотря на значительное количество работ, описывающих отдельные его аспекты. В частности, для компьютерной лингвистики язык интернет-коммуникации представляет особый интерес, потому что изучение интернет-общения важно для различных отраслей экономики, язык интернета близок к устной речи, он отражает самые актуальные речевые тенденции и умонастроения, а также электронные тексты наиболее доступны, их легче найти и обработать, чем классические печатные произведения. Современные интернет-технологии способствуют развитию международного общения, в сети стираются границы между странами, открыт доступ к текстам на всех имеющихся языках, что ведет не только к обмену информацией, но и к взаимному влиянию языков друг на друга. Поскольку большое влияние на русский язык оказывается со стороны английского языка, **предметом** нашего исследования стали русско-английские эквиваленты интернет-коммуникации.

Актуальность данной работы обусловлена тем, что взаимодействие систем английского и русского языка в сфере интернет-коммуникации нуждается в изучении и описании, существует необходимость составления двуязычных словарей, определения новых культурных понятий и особенностей общения в сети. Процесс исследования некоторых аспектов, например, лексического, осложняется быстрым темпом смены лексического состава, постоянным появлением новых явлений, а также их устареванием и выходом из употребления. В условиях интернет-коммуникации пользователям приходится сталкиваться с огромным количеством информационных сообщений и реагировать на них, что связано с развитием клипового мышления и ведет к постепенному процессу замены языковых

выражений определенных мыслей, чувств и эмоций шаблонами, например, типовыми картинками, анимацией и фразами-клише.

Цель работы, таким образом, заключается в изучение лексических русско-английских эквивалентов интернет-коммуникации, определение характера связи между ними и особенностей их употребления с использованием инструментов корпусной лингвистики. Поставленная цель определяет выполнение следующих **задач**:

- 1) обзор современного опыта работы с лингвистическими корпусами и принципов их построения,
- 2) уточнение понятия эквивалента,
- 3) определение особенностей языка интернет-коммуникации,
- 4) обзор современных доступных корпусов интернет-коммуникации,
- 5) анализ существующих словарей интернет-языка,
- 6) поиск и сбор материала для исследования с целью последующего его использования в создании собственного корпуса,
- 7) рассмотрение и выбор программного обеспечения для построения корпусов, предлагаемых инструментов для работы,
- 8) создание собственного корпуса и описание методики его построения,
- 9) раскрытие содержания некоторых понятий русского и английского языка интернет-коммуникации на основе корпуса, их сравнение, выявление совпадений и расхождений с применением инструментального анализа, определение характера связи между ними,
- 10) подведение итогов работы и анализ результатов исследования.

Поставленные цели и задачи определили выбор **методов** исследования:

- на этапе сбора материала – метод пассивного наблюдения с последующим обобщением полученных результатов, метод произвольной выборки по определенным критериям,

- на этапе исследования – дистрибутивный метод, метод автоматического анализа, метод количественных характеристик, сопоставительный метод, метод описания по лексико-семантическим группам, инструментальный метод, метод лексикографического описания.

Поэтапное решение поставленных целей и задач определило **структуру** работы: введение, главы 3 исследования, заключение, библиографический список.

В **первой главе** осуществляется предварительная подготовка к созданию корпуса, предусматривающая рассматривается понятие корпуса, определяются особенности языка интернет-коммуникации, производится обзор корпусов, рассматривается понятие эквивалента.

Во **второй главе** описывается методика проведения исследований на основе созданного корпуса посредством инструментов корпус-менеджера, производится анализ полученных результатов.

В **третьей главе** описывается методика создания собственного корпуса, его характеристики, дается оценка корпусу, предлагаются способы его применения и усовершенствования.

В **заключении** подводятся итоги проделанной работы и анализ результатов исследования.

ГЛАВА 1. ПОДГОТОВКА К СОЗДАНИЮ КОРПУСА

Корпусная лингвистика представляет собой на сегодняшний день одно из перспективных и активно развивающихся направлений современной теоретической и прикладной лингвистики, основанное на использовании языковых электронных корпусов и корпусных технологий, что подтверждает большое количество конференций и публикаций по вопросам корпусной лингвистики. В корпусной лингвистике существует два основных направления: создание корпусов и корпусные исследования, то есть лингвистические исследования с использованием корпусных методов. Четкой границы между ними провести невозможно. Сегодня корпусная лингвистика требует как лингвистических, так и технических знаний, прежде всего информационных технологий.

1.1. Понятие корпуса

На сегодняшний день **корпус** представляется как «собрание текстов на данном языке в электронной форме, специальным образом обработанное («аннотированное»), так, чтобы исследователь языка мог быстро и в полном объеме найти в корпусе интересующую его информацию о языке»¹. Основными составляющими корпуса являются языковые данные (тексты, аудиозаписи, видео, картинки), лексические и грамматические ресурсы и технические инструменты (алгоритмы, поисковая система). Аннотация, выполненная по определенному стандарту, позволяет организовать поиск единиц (слов, предложений и текстов) по морфологическим, семантическим и синтаксическим характеристикам, а так же по информации об авторе, дате издания, жанре, тематике, гендерной принадлежности и прочему.

Электронный корпус выгодно отличается от собранных вручную коллекций документов по следующим характеристикам:

¹ Программа фундаментальных исследований президиума РАН <http://www.corplingran.ru/>

- Корпус позволяет получить «живые», настоящие речевые примеры, ориентироваться на узус, а не норму. Корпусная лингвистика использует в рамках своей методологии не прескриптивный, а дескриптивный подход, сформулированный в конце XIX – начале XX веков (В.И. Даль, О. Есперсен).

- Корпус облегчает применение количественных методов исследования, предоставляет новые возможности анализа материала. Инструменты анализа зависят от технических возможностей корпуса.

- Аннотация электронного лингвистического корпуса позволяет осуществлять быстрый поиск единиц по интересующим параметрам. Результат поиска зависит от качества аннотации.

- Электронный корпус создает возможности для изучения языка с формальной точки зрения.

- Объем современных электронных корпусов значительно отличается от тех, что были собраны вручную. Внушительный объем собранных документов способствует адекватному представлению о языковой практике, дает возможность работать с более репрезентативным и сбалансированным собранием языкового материала.

- Электронная форма корпуса обеспечивает быстрый поиск и извлечение материала, избавляет от ручного поиска и обработки иллюстративного материала для исследования. Однако корпус представляет только тот материал, который в нем содержится, поэтому поиск ограничен содержанием и атрибутами аннотации корпуса, требует точных параметров запроса.

В процессе подготовки материала для корпуса неизбежно возникают вопросы, во-первых, по каким принципам проводить выборку материала, каким образом решить проблему объективности представления языка

посредством этой выборки, ведь в идеале корпус должен обеспечивать достоверность представления языка. Во-вторых, полный объем текстов, созданных на том или ином языке, невозможно измерить, он не имеет границ, а корпус, в свою очередь, имеет лимитированный объем.

Для решения ключевых вопросов при составлении качественного лингвистического корпуса следует учитывать такие **требования**, как репрезентативность, объем, сбалансированность, аннотация, наличие определенной тематической ориентации и цели использования, ориентации на пользователя.

Репрезентативность, требование, применяемое к объему и составу корпуса, заключается в статистически достоверном представлении языка или его части, создающееся за счет необходимого объема и жанрового разнообразия текстов. Корпус может быть репрезентативен по отношению к объекту, например, к языку в целом или к его части. Иными словами, относительная частота появления того или иного явления в корпусе должна быть приближена к относительной частоте появления этого явления в языке. Несоблюдение этого требования ведет к искажению объективности представлений о языке. Первые корпуса имели тенденцию к выбору материала, приближенного к норме, отличались литературоцентричностью. В настоящее время корпуса развиваются по пути пропорциональной репрезентативности, согласно которому жанры и типы текстов должны быть представлены пропорционально генеральной совокупности («популяции») текстов на данному языке². Однако достоверно измерить общее количество текстов на языке не представляется возможным, так как оно не ограничено.

Сбалансированность, требование, тесно связанное с репрезентативностью, и также применяемое к составу и объему корпуса, определяет равномерность представления текстов разных типов, за счет чего

² В.А. Плунгян, Д.В. Сичинава Национальный корпус русского языка: опыт создания корпуса текстов современного русского языка

должно достигаться объективное отражение состояния языка. Для удовлетворения этого требования необходимо пропорционально (реально существующему распределению текстов в языке) включать все тексты, представленные на языке, относящиеся к разным жанрам и сферам функционирования (если речь идет о национальном корпусе), или соотносящиеся с разработанной классификацией, отвечающей требованию репрезентативности изучаемой его части. Слабой точкой современных корпусов является включение в корпус большой доли письменных текстов, литературных произведений, а не устных, что не соответствует реальным долям в языке. Создать идеальное соответствие живому языку опять же невозможно, равно как и измерить точные доли видов речи, жанров и стилей. Многие корпуса стремятся исправить сложившуюся ситуацию, создавая пропорциональное разделение внутри письменных жанров и разрабатывая новые подкорпусы устной речи. Например, включаются современные жанры интернет-коммуникации, сферы, сравнительно наиболее приближенной к устной речи, а также мультимедийные корпуса. Например, в НКРЯ на данный момент имеет следующее распределение:

I. Распределение текстов по подкорпусам

Подкорпус	Число текстов	Число предложений	Число словоупотреблений	% словоупотреблений
Основной корпус	76 882	17 574 752	209 198 275	57.3%
- в том числе со снятой омонимией	2 147	516 852	5 944 188	1.6%
Газетный корпус	181 175	8 553 495	113 292 003	31.0%
Диалектный корпус	197	20 273	194 283	0.1%
Обучающий корпус	229	65 666	664 751	0.2%
Параллельный корпус	370	1 609 609	24 022 437	6.6%
Поэтический корпус	41 448	638 861	6 738 474	1.8%
Устный корпус	3 034	1 604 626	10 122 579	2.8%
Мультимедийный корпус	31 741	148 619	648 576	0.2%
Всего:	335 076	30 215 901	364 881 378	100%

II. Распределение текстов основного корпуса по видам и другим метапризнакам

Вид текста	Число текстов	Число предложений	Число словоупотреблений	% словоупотреблений
Художественные письменные тексты	6 308	9 164 402	92 058 363	44.0%
Нехудожественные письменные тексты	70 574	8 410 350	117 139 912	56.0%
Всего:	76 882	17 574 752	209 198 275	100%

Рисунок 1. Распределение текстов по подкорпусам в НКРЯ.³

³ НКРЯ www.ruscorpora.ru/corpora-stat.html

Репрезентативность и сбалансированность – требования, позволяющие адекватно представить существующее разнообразие текстов в относительно равных пропорциях, соответствующих пропорциям в языке. Корпус не может вместить в себя все существующие тексты, но призван адекватно отображать ситуацию, складывающуюся в реальной языковой практике.

Не менее важным требованием является **объем**. Обычно сообщается информация о количестве документов, входящих в корпус, количестве предложений и количестве словоупотреблений. Объем корпуса рассчитывается в зависимости от специфики и целей использования корпуса, например, для адекватности лексикографических исследований требуется корпус в десятки и даже сотни миллионов вхождений, тогда как для изучения пунктуации или грамматических явлений требования меньше. В целом, чем больше объем корпуса, тем он более репрезентативен, применим для частотного анализа и соответствует представлениям об идеальном корпусе, как о вместилище всех текстов, созданных на языке. Информация об объеме корпуса дает возможность абсолютные частоты перевести в относительные и получить более достоверное представление о предмете изучения. Движение в этом направлении особенно характерно для национальных корпусов, тогда как для специализированных корпусов объем может быть ограничен внешними факторами, к примеру, корпус текстов одного автора ограничен объемом созданных им произведений. От объема зависят репрезентативность и сбалансированность.

Электронный (размеченный) корпус отличается от собрания текстов наличием **аннотации (разметки)**, от качества которой зависит эффективность поиска информации в нем. Чем богаче разметка, тем больший интерес представляет корпус для исследователей. Принципы разметки зависят от цели создания и использования корпуса. Аннотирование корпуса

может производиться в ручном, полуавтоматическом или автоматическом режиме, и зачастую подразумевает:

- метаданные о тексте в целом (имя, возраст и пол автора, дату создания текста, название, жанр, тематику, сферу функционирования, источник, размер и прочее),
- акцентологическую, морфологическую и семантическую разметку каждого отдельного звука, морфа, леммы и синтаксическую, соответственно, для словосочетания и предложения.

Качественная разметка, а так же снятие омонимии позволяют предоставлять объективные статистические данные по корпусу, быстро осуществлять поиск по определенным параметрам, получать дополнительную информацию по запросу. Уровень аннотации корпуса напрямую связан с возможностями корпус-менеджера.

Если большинство первых корпусов представляли собой коллекции текстов преимущественно письменной речи, собранным по строгой схеме, с большим вниманием к репрезентативности, то на сегодняшний момент круг интересов составителей корпусов и исследователей значительно расширился, включив в себя корпусы языков разных областей функционирования и подъязыков, например, параллельный корпус текстов ООН <http://www.linguee.com/>, или НКРЯ, постепенно добавляющий в свой состав подкорпусы различных тематик, таких, как мультимедийный, электронный, диахронический подкорпусы.

Интерес к корпусным исследованиям способствовал развитию разнообразия типов корпусов, отвечающих широкому кругу вопросов для изучения.

1.2. Особенности языка интернет-коммуникации

Понятие **интернет-коммуникации** понимается исследователями как «взаимодействие между человеком и человеком преимущественно в текстовом виде посредством связанных в сеть компьютеров или мобильной телефонии»⁴, однако, по причине быстрой скорости развития технических средств и информационных технологий данное понятие нуждается в постоянном переосмыслении, изучении новых аспектов и явлений.

Автор определения делает акцент на том, что коммуникация происходит в основном с помощью письменных текстов, хотя в настоящее время существует видеосвязь, где коммуниканты имеют возможность общаться устно, видеть реакции, мимику, жесты, таким образом, коммуникация становится более полноценной, чем была ранее. Если говорить о письменной коммуникации, то сообщения часто включают в себя видео-ролики, аудио-записи, а так же фото, смайлы, мемы, демотиваторы и прочие картинки для передачи более полного объема информации.

Цели коммуникации в интернете разнообразны, это может быть обмен быстрыми сообщениями, ведение личного блога или дневника, обсуждение актуальных тем и новостей, обращение к органам власти, размещение информации, объявлений и прочее.

Несмотря на то, что сообщения могут быть в письменной форме, их нельзя отнести к сообщениям литературной речи. Исследователи отмечают близость электронного дискурса к устной речи со всеми присущими ей признаками, однако, язык интернет-коммуникации имеет свои особенности и свою конвенциональность, отличающие его от литературного языка. Письменная речь в интернете стремится компенсировать отсутствие личного контакта между коммуникантами, по этой причине происходит обогащение

⁴ Herring, S. A Faceted Classification Scheme for Computer-Mediated Discourse// Language@internet. - 2006. <http://languageatinternet.de/articles/761>.

лексического состава языка, изменение синтаксических конструкций, нарушение норм письменной речи, появлению экстралингвистических вкраплений в сообщения.

В целом, можно выделить следующие свойства интернет-коммуникации:

- высокая степень анонимности,
- шаблонность высказываний и выражения эмоций посредством эмодиконов, картинок, мемов, gif-файлов, картинок, шаблонных фраз,
- существование особой жанровой системы, состоящей из собственных и заимствованных жанров, все жанры являются потенциально диалогическими,
- наличие особой конвенциональности, правил поведения,
- креолизованность текстов, сочетание нескольких семиотических систем.

Как английскому, так и русскому языку интернет-коммуникации характерны такие тенденции, как метафоризация, экономия речевых средств, а так же избыточность. Несмотря на некоторую степень хаотичности и отсутствия строгих норм, интернет-язык стремится к структурированию. Быстрый темп коммуникации, относительная свобода выражения и желание сделать свою речь более оригинальной и отличной от других способствовали появлению особых языковых явлений, свойственных именно интернет-языку: взаимодействие кириллицы и латиницы, фонетическое письмо, нетрадиционное использование знаков препинания, возникновению новых лексических единиц, а так же особых языков внутри языка интернет-коммуникации (lolcat, олбанский язык), различные способы графического оформления сообщений, не свойственные литературной речи.

В связи с тем, что США сохраняет лидирующее положение в области развития информационных технологий, английский язык оказывает одностороннее влияние в частности на русский язык, прежде всего на лексическом уровне, однако, активны заимствования из японского языка попутно с культурой японских комиксов и мультфильмов. Русский язык заимствует англоязычные понятия, адаптирует их, переосмысляет. Русскому и английскому языку характерно наличие аббревиаций, использование фонетического образа слова для написания, использование цифр для замены слов и слогов.

В русском языке можно выделить несколько этапов адаптации заимствований:

1. Фонетическая адаптация. Адаптация может пойти по одному из двух путей:

- a. Слово сохраняет свой исходный фонетический облик. Gamer - геймер, bro – бро, man - мэн, crack – крэк.
- b. Произношение слова отражает не фонетический, а графический облик. Game – гама, game over - гамовер, girl – гирла.

2. Лексико-семантическая адаптация. Некоторые слова приобретают просторечные суффиксы (gif – гифка / гифец, flesh – флешка, label – лэйбак, true – трушный).

3. Грамматическая адаптация. Адаптируясь, лексемы вступают в морфологические парадигмы, характерные для той части речи, которой они являются, начинают участвовать в словообразовании. Game: гама – гамать – погамать – гамер – гамовер. Far – фап – фापать – пофापать – фаповый. Fail – фэйл – фэйловый – зафэйлить – зафэйленный.

Пройдя адаптацию, заимствованная лексика вступает в отношения с уже имеющимся лексическим составом русского языка. С точки зрения

наличия или отсутствия эквивалентных соответствий, можно говорить о существовании первичных и вторичных номинаций. Первичные номинации служат для обозначения новых реалий, которым в языке еще нет названия (friend zone, poker face, facepalm, butthurt). Вторичные номинации обозначают уже знакомые реалии в языке-реципиенте, для которых уже существуют названия (looser – неудачник, gamer – игрок, display - экран). Таким образом, возможно выделить несколько групп:

1). Безэквивалентные наименования (первичная номинация). Twitter – твиттер, твитт, твиттить;

2). Эквивалентные наименования:

а. дублиеты (полные синонимы): debug – дебаг / отладка, up grade – апгрейд / обновление,

б. идеографические синонимы (в языке-реципиенте существуют эквиваленты, однако, заимствованное слово уточняет его значение): девушка / тян, копипаст / копия или плагиат,

с. стилистические синонимы: подеделиться / расшарить, комментировать / комментить, подписчик / фолловер,

д. эмоционально-экспрессивные синонимы: обида – баттхерт,

е. слова активного / пассивного запаса: ненавистник - хейтер.

Лексика, входящая в состав языка интернет-коммуникации фиксируется в электронных словарях, например, в таких, как (для английского языка) Urban Dictionary (www.urbandictionary.com), The Online Slang dictionary (www.onlineslangdictionary.com), www.netlingo.com, thesource4ym.com/SlangDictionary/SlangDictionary.aspx, www.yourdictionary.com, dictionary.reference.com/slang, www.noslang.com/dictionary/, www.wordspy.com и (для русского языка) www.slang.ru и www.slovonovo.ru.

1.3. Обзор корпусов интернет-коммуникации

Область интернет-коммуникации остается недостаточно изученной с лексической точки зрения в силу быстрой скорости изменений в лексическом составе. Существующие словари фиксируют наиболее распространенные единицы, но зачастую их толкование неоднозначно, недаром словари создаются коллективно, существуют рейтинги адекватности и актуальности единиц.

На сегодняшний день существует небольшое количество корпусов, посвященных языку интернет-коммуникации. Одним из наиболее крупных современных корпусов интернет-текстов является **The Corpus of Global Web-Based English (GloWbE)**, состоящий из 1,9 миллиарда слов и 1,8 миллиона веб-страниц, созданных в 20 англоговорящих странах (Великобритания, Ирландия, США, Канада, ЮАР, Нигерия, Кения, Гана, Танзания, Ямайка, Сингапур, Малайзия, Филиппины, Австралия, Гонконг, Индия, Пакистан, Шри-Ланка, Новая Зеландия и Бангладеш). Корпус включает в свой состав the Corpus of Contemporary American English (COCA) и the Corpus of Historical American English (COHA), таким образом, давая возможность изучать особенности диалектов, жанров, исторические изменения, чего нельзя сделать в других корпусах английского языка. В корпусе можно осуществлять поиск по одному, нескольким или всем представленным диалектам, а так же синхронно сравнивать результаты поиска по диалектам, просматривать частотность употребления интересующего вхождения по диалектам, осуществлять поиск слова, словосочетания или грамматической конструкции, изучать синонимы, сочетаемость. Тексты были собраны в декабре 2012 года на основании списка частотности n-граммов корпуса COCA, просматривались страницы для всех интересующих стран, ориентируясь по домену и информации расширенного поиска Google.

Корпус **The WebCorp Live Linguist's Search Engine** также позволяет работать с текстами интернет-коммуникации, но в режиме онлайн. Поиск осуществляется не по заранее собранному закрытому корпусу, а посредством поисковых систем Google, Bing, FAROO, Blekko на 36 языках с возможностью поиска по Google blogs, Google news и FAROO news, а также ограничения по национальным доменам и сайтам популярных новостных изданий, например, BBC News, UK Tabloid Newspapers, French Newspapers, также по Wikipedia, US Academic, UK Academic. В целом, The WebCorp представляет собой небольшой набор инструментов для работы (по большей части поиска) с интернет-текстами, имеющий преимущества перед стандартными корпусами в том, что представляет живой материал и позволяет изучать редкие явления, не всегда содержащиеся в закрытых корпусах. Доступна функция поиска по лемме, по грамматическим элементам, по сочетаниям с *, исключения ненужных результатов (-) и учета верхнего/нижнего регистра, функция составления таблиц коллокаций с указанием частотности вхождений с доступной сортировкой по дате или в алфавитном порядке. На выходе указываются URL-ссылки, дата создания текста. Недостатками корпуса, собранного онлайн, является то, что в результатах поиска присутствуют повторения, хэштеги, что снижает привлекательность корпуса для проведения исследований, поскольку невозможно получить репрезентативные данные.

Таким образом, к составлению корпуса интернет-коммуникации можно подходить двумя способами: собрание фиксированного количества текстов по определенным параметрам или поиск текстов из сети в режиме онлайн. Выбор метода зависит от цели создания и использования корпуса, а также от тематики корпуса.

Первый способ отличается более высоким качеством сбора информации, возможностью заранее исключить повторяющиеся тексты, первично обработать материал, произвести аннотацию, ручную коррекцию,

снять омонимию, решить вопрос сбалансированности и репрезентативности корпуса. Однако материал будет относиться к конкретной дате, что невыгодно для изучения текстов, быстро теряющих свою актуальность. Зато тексты собраны по продуманной схеме, разнообразны по содержанию, стремятся максимально адекватно представить изучаемую область языка.

Второй же способ привлекателен доступностью свежего материала и высокой скоростью составления корпуса, наличием необходимых инструментов для работы, позволяет произвести обзор материала по интересующему вопросу без глубокой аннотации. Поиск осуществляется заранее известной информации или с одним неизвестным с применением *. Тексты собраны случайным образом из разнообразных источников, что, с одной стороны, отражает действительность такой, какая она есть, а с другой, нельзя окончательно быть уверенным в том, что информация собрана адекватно реальной ситуации. Исследователю предстоит самостоятельно оценивать релевантность представленных результатов.

Можно подвести итог, что открытые корпуса подходят для первичного поиска данных, для ориентации в материале для изучения и составления общего представления по изучаемой проблеме. Закрытые корпуса применимы для более точных и детальных исследований, где необходима дополнительная информация и требуются ограничения материала по какому-либо параметру. Такие корпуса эффективны для изучения узких аспектов, особых областей функционирования языка, жанров, уровней.

1.4. Понятие эквивалента

В теории перевода понятие **эквивалентности** рассматривается в ключе нескольких концепций, подразумевает под собой характеристику результата перевода, характер связи исходного текста с его переводом. Достижение эквивалентности считается основной целью перевода, однако,

абсолютная эквивалентность практически невозможна из-за семантических, структурных, прагматических, культурных различий между текстами.

По данным словаря-справочника лингвистических терминов Д.Э.Розенталя⁵, эквивалент – это «единица речи, способная выполнять ту же функцию, что и другая речевая единица». Словарь терминов международной коммуникации⁶ апеллирует к определению эквивалента 1. Я.И. Рецкера (1950) как «регулярной переводной формы, не зависящей от контекста (например, топонимы, антропонимы, термины)». 2. Определяет его как «равнозначную замену исходного слова переводным». Словарь приводит классическое понимание эквивалента вслед за Я.И.Рецкером, описывающее отношения между микроединицами текста, а не межтекстовые отношения, поэтому эквивалент понимался как «постоянное равнозначное соответствие как правило не зависящее от контекста».⁷ Рассматривая эквивалентность с лексикографической точки зрения, автор выделяет полные и частичные эквиваленты, вариантные (контекстные) соответствия, устанавливающиеся между словами, если в языке перевода есть несколько вариантов передачи этого значения, и переводческие трансформации.

В дальнейшем Ю.Найдой⁸ было предложено понятие «динамической эквивалентности», противопоставляемое понятию «формального соответствия» и выдвигающее на передний план значимость коммуникативной ситуации, важность создания максимального ее соответствия между первичным и вторичным текстами. Далее категория эквивалентности получила осмысление на различных уровнях и в разных аспектах.

⁵ Словарь-справочник лингвистических терминов. Изд. 2-е. — М.: Просвещение. Розенталь Д. Э., Теленкова М. А.. 1976.

⁶ Словарь терминов электронной коммуникации. И.Н.Жукова, М.Г.Лебедько, З.Г.Прошина, Н.Г.Юзефович; под ред. М.Г.Лебедько и З.Г.Прошиной. – М. : ФЛИНТА : Наука, 2013. – 632 с.

⁷ Рецкер, Теория перевода и переводческая практика, -М.: Международные отношения, 1974, с.10—11

⁸ Nida E.A., Taber Ch. The theory and practice of translation. - Leiden, 1969, с.202

Тема нашей работы непосредственно связана с понятием эквивалента, поэтому стоит оговорить, что под ним будет подразумеваться. В отличие от одноязычного словаря, например, толкового, где представлены значения слов, в двуязычном словаре представляются **словарные межъязыковые соответствия (эквиваленты)**. В сущности, они должны обозначать совпадающие понятия в одно и другом языке. Объемы значения лексических единиц в разных языках часто не совпадают. В процессе речи значение модифицируется, часть семантических компонентов нейтрализуется, актуализируется один необходимый компонент в заданном контексте, так же он может переосмыслиться⁹. Соответственно, **контекстуальные эквиваленты**¹⁰ устанавливаются путем поиска «идиоматичного (соответствующего нормам сочетаемости и прагматике) выражения» определенного контекстуального значения слова, актуализированного в речи. «Межъязыковая идиоматичность» (несвободная сочетаемость) делает невозможным формальный, не зависящий от контекста перевод, не учитывающий правила синтактики и концептуализации¹¹. Словарные межъязыковые эквиваленты лексических единиц не всегда применимы для перевода, так как их значения лингвоспецифичны. Например, в некоторых словосочетаниях с глаголом 'to pick up'¹² смысл «прикрепленность объекта» нерелевантен и выражается на русский язык соответствующими межъязыковыми эквивалентами, передающими дополнительные смыслы: to pick up flowers – нарвать цветов, to pick up the kids at school – забрать детей из школы, to pick smb up at the office – заехать за кем-либо на работу.

Н.К.Рябцева предлагает следующую классификацию межъязыковых эквивалентов:

⁹ Рябцева Н.К. Прикладные проблемы переводоведения, М.: ФЛИНТА, 2014. – 224 с., с.92

¹⁰ Гак В.Г., Григорьев Б.Б. Теория и практика перевода. Французский язык. – СПб.: Интердиалект+, 2000, с.13

¹¹ Рябцева Н.К. Язык и естественный интеллект. М.: Academia, 2005, 640 с.

¹² Концепция лингвистического континуума: монография / В.М. Савицкий, О.А. Кулаева. Самара: НТЦ, 2004. – 178 с., сс. 34-35

- словарные / контекстуальные,
- однозначные / неоднозначные,
- устойчивые / условные,
- коммуникативно / ситуативно равноценные и др.

Однозначные соответствия – «контекстно независимые константы», значение которых в норме не зависит от контекста: названия, имена собственные, термины и др.

Неоднозначные соответствия – соответствия, перевод которых зависит от контекста, т.е. «контекстно зависимые» соответствия.

Устойчивые соответствия – безальтернативные соответствия, другие варианты передачи на другом языке невозможны (перевод терминов, имен собственных, общепринятых в данной культуре названий, реалий и прочее).

Условные соответствия – приемы передачи различных отклонений от нормативной речи: диалектизмы, жаргонизмы, речь иностранцев и так далее.¹³

Таким образом, межъязыковые эквиваленты могут не являться переводами друг друга в прямом смысле этого слова. На лексическом уровне выделяются словарные межъязыковые эквиваленты, представляющие собой внеконтекстные соответствия слова (словосочетания) на одном языке слову (словосочетанию) на другом языке. Такие соответствия представлены в двуязычном словаре. Также выделяются контекстуальные эквиваленты, не представленные в словаре, но выражающие одинаковый смысл и употребляющиеся для описания аналогичных понятий, ситуаций. Они позволяют передать те смыслы текста оригинала, выражения которых не существует в другом языке, носит концептуальный характер.

¹³ Рябцева Н.К. Прикладные проблемы переводоведения, М. : ФЛИНТА, 2014. – 224 с., с.131

Выводы

Лингвистический корпус представляет собой специально аннотированное собрание текстов в электронной форме, организованное с целью осуществления поиска в нем интересующей информации о языке и получение данных о нем в полном объеме. К современному корпусу предъявляются такие требования, как репрезентативность, сбалансированность, релевантный объем, наличие аннотации, организованный поиск и тематическая ориентация, способствующие достоверному отражению языковой действительности.

Поскольку объектом нашего исследования является язык интернет-коммуникации, то нами были определены его особенности: анонимность, шаблонность высказываний, наличие особой жанровой системы, особой конвенциональности и креолизованность. Ему свойственны метафоризация, отсутствие строгих норм, нетрадиционное использование знаков препинания, фонетическое письмо, наличие уникальных понятий и лексических единиц. В русском языке велика доля англоязычных заимствований.

Для представления языка интернет-коммуникации существуют корпуса как с закрытым, так и с открытым источником. Преимущество на стороне корпусов с закрытым источником, им характерна спланированность, широта объема информации, наличие инструментов для поиска, тогда как корпуса с открытым источником не всегда справляются со своей задачей, представляя результаты с большим количеством шума, повторов, нерелевантных примеров.

Для определения сущности состава нашего корпуса необходимо определить понятие эквивалента. За основу взято определение эквивалента Я.И.Рецкера как «равнозначной замены исходного слова переводным». Эквиваленты могут быть словарными и контекстуальными, подобранными с учетом ассиметрии языковых систем.

ГЛАВА 2. МЕТОДИКА РАБОТЫ С КОРПУСОМ

В рамках нашей работы нами был составлен корпус русско-английских эквивалентов интернет-коммуникации (см. описание гл.3.1) с целью возможности проведения исследований. Корпус был предварительно спланирован, производился отбор материала из сети Интернет по ключевым словам, характерным для языка области функционирования интернет-коммуникации. В корпус отбирались тексты, представляющие речь пользователей в сети: вопросы-ответы, комментарии, сообщения из форумов. Общий объем составляет 2000 текстов. Для работы с составленным корпусом был выбран корпус-менеджер Sketch Engine, представляющий наиболее подходящие инструменты для поиска и анализа материала: Concordance, WordSketches, Thesaurus и Sketch-Diff.

Наша работа с корпус-менеджером началась с того, что мы загрузили собранную коллекцию текстов в два корпуса: корпус русского языка и корпус английского языка, скомпилировали корпуса. Автоматически произведена токенизация, лемматизация и грамматическая аннотация, что в дальнейшем облегчает поиск по корпусу и работу инструментов.

Методика работы с лексическими единицами будет включать в себя несколько шагов: анализ словарных значений слов, просмотр окружения в конкордансе, составление скетчей и по возможности тезауруса, выявление сходства и различия между скетчами. Предварительно рассматриваемые лексические единицы были разделены на несколько групп: выражение эмоций, наименования типажей и чувства и реакции пользователей.

2.1. Конкорданс

*'... is there a limit as to how many "ha's" to put?'*¹⁴

Рассмотрение способов выражения эмоций в интернет-сообщениях с первого взгляда может показаться слишком простой и не заслуживающей внимания темой, да и способы выражения кажутся интуитивно понятными. Однако пользователям не всегда понятны используемые выражения, иногда они даже подобны азбуке Морзе и вызывают разные толкования, что подтверждает большое количество вопросов о том, как эти выражения можно понять или расшифровать.

Для изучения способов выражения эмоции мы воспользуемся таким инструментом, как KWIC конкорданс (поскольку данная тема не нуждается в применении иных ресурсов), для чего взяли из всего множества существующих в английском языке единиц только некоторые из тех, что выражают смех (lol, rofl, bwahaha, ha-ha, haha, HA) и удивление (wtf, damn, omg, WAT). В свою очередь, представленные единицы можно разделить на эмоциональные междометия (bwahaha, ha-ha, haha, HA, WAT) и аббревиации (lol и его производная ololo, rofl, wtf, damn, omg).

• 'HA'

Для английского языка словарь The UrbanDistionary дает следующие толкования:

ha awknowledgement of someones comment or joke, but indicating it wasn't funny.

John: I love girls that wear their PJs all day long, it's like they are ready to jump in bed at a moments notice.

*Jack: ha*¹⁵

¹⁴ <https://answers.yahoo.com/question/index?qid=20131205135737AAXvn2Z>

¹⁵ <http://www.urbandictionary.com/define.php?term=ha>

Словарь Webster's New World College Dictionary дает более широкое толкование:

Ha *interjection* used variously to express surprise, wonder, triumph, anger, etc. *Noun* the sound of this exclamation or of a laugh ¹⁶

По данным словаря Dictionary.com:

Ha *interjection, noun* (used as an exclamation of surprise, interrogation, suspicion, triumph, etc.) ¹⁷

Если следовать толкованиям словарей, то в английском языке междометие 'НА' означает, с одной стороны, выражение поддельного смеха, реакцию на шутку, не вызывающую должного одобрения, а с другой стороны, оно может обозначать различные чувства, такие как удивление, триумф, подозрение, раздражение и так далее. Получается, что 'НА' является обозначением широкого спектра реакций, отличных от смеха.

Для того, чтобы определиться с тем, что же обозначает данное междометие в английском языке, обратимся к нашему корпусу, который предлагает следующие примеры его употребления:

*'Are YOU an innocent person? Bwahaha I was quite the goody two-shoes as a kid, clear up to my high school years. But anymore? **HA!** I hardly think so!'* В представленном примере рассматриваемое междометие выражает скепсис, сомнения, значение смеха отсутствует.

*'Don't you just hate it when people are overly sarcastic and rude? I wish I could just drown certain people in their sarcasm. Update: **Ha** , I love peoples reactions to my sarcasim. YEAH! I'm saracastic but I know when to stop! i actually enjoy dry and sarcastic people but i am a different type of personality than you and*

¹⁶ Webster's New World College Dictionary Copyright © 2010 by Wiley Publishing, Inc., Cleveland, Ohio. <http://www.yourdictionary.com/ha>

¹⁷ Dictionary.com <http://dictionary.reference.com/browse/ha?s=t>

i am an ice princess.’ Здесь ‘НА’ служит для выражения триумфа, чувства достигнутой цели.

*‘Grammatically, does the emoticon go before or after subsequent punctuation? Update: **Ha** , even though I do want an answer to this, I hope you realize that my post was completely sarcastic.’* ‘НА’ выражает сарказм, насмешку над другими.

*‘Trying to understandd this conversation...? guy: no, steph, i do not wanna make out with you (said in a loud sarcastic voice) girl: you sure?? **ha** (still sarcastic) guy: no. (just normal...) Okay, so like does the guy like her?’* Выражение сарказма, сомнения.

*‘But, **ha** ! another thundering crash. The dam has burst.’* Обозначение удивления, реакции на неожиданное событие.

Таким образом, в текстах междометие не употребляется для обозначения смеха, но оно используется для обозначения сарказма, скепсиса, сомнения, триумфа, удивления, в целом это значение можно сравнить с выражением самозащиты, когда другие люди не поддерживают автора, даже осуждают, или мнения собеседников не совпадает, и тогда ‘НА’ выступает в качестве возражения.

Толкование междометия ‘НА’ или «ХА» в русскоязычных словарях отсутствует, поэтому наш русскоязычный корпус поможет выяснить их значение.

«Вот ЕеOneGuy КРАСАВА)) Накопил бабла и живет в Китае ХА)) Ему же много платят))» Выражение усмешки, белой зависти, автор хотел бы себе позволить такой образ жизни, но не может себе позволить.

«А ВЫ ПОЛУЧИЛИ ТАКОЕ ПИСЬМО НА ПОЧТУ? МНОГИМ НА ПОЧТУ ПРИШЛО ТАКОЕ ПИСЬМО И ЛЮДИ УВЕРЕННЫ ЧТО УЖЕ

СТАЛИ МИЛЛИОНЕРАМИ! ХА)))» Выражение скепсиса. Автор не верит в то, что обещанное имеет хотя бы долю вероятности быть правдой.

«Ха , офигенно! Просто в восторге. Подруга сейчас скинула ссылку, на один полезный и прикольный сайт.» Выражение приятного удивления, восторга.

«А он такой еще говорит, мол посиди со мной в машине. пока я следующий вызов подожду. Ха ! еще чего! И я захожу в квартиру, лап-лап по карманам - нет бибики моей.» Выражение протеста, отказа, автор определил коварство и хитрость своего собеседника.

В русскоязычных текстах значения междометия «ХА» может обозначать сарказм, успешку, удавление, протекст, восторг и другие эмоции, не умеющие прямого отношения к смеху, что совпадает со значением междометия 'НА' в английском языке, поэтому 'НА' и «ХА» можно считать однозначными эквивалентами.

- 'На ha'

Двойное 'ha' трактуется в англоязычных словарях следующим образом:

Dictionary.com

ha-ha¹ *interjection, noun* (used as an exclamation or representation of laughter, as in expressing amusement or derision.)¹⁸

Yourdictionary.com:

ha-ha *interjection* used to suggest the sound of laughter, in expressing variously humor, joy, derision, etc.

noun this sound¹⁹

¹⁸ Dictionary.com <http://dictionary.reference.com/browse/haha?s=t>

UrbanDictionary.com

haha Short quick way of letting somebody know you are laughing, most likely at them. *You failed your exam? Haha, er, that's too bad...*²⁰

По информации словарей, междометие 'haha' является быстрым способом обозначения смеха, реакции на юмор, шутку, нечто забавное, а также насмешки над чьим-либо поведением.

Рассмотрим примеры употребления в нашем англоязычном корпусе:

*'For instance, at my job, when someone asks me where they can find something, they ask that normally but when I tell them, they often start laughing. Not laughing laughing like **ha ha**, but they have a laugh in their response if that makes sense. It happened today. I just tried to help her after she asked me and then she laughs. It's hard enough to have low self esteem but then stuff like this happens and brings you down even lower.'* Данный пример описывает понимание рассматриваемого междометия как реакции, близкой к улыбке в ответ на реплику собеседника, но не смеха, как реакции, обозначающей, что один собеседник услышал и понял другого, но ответить ему нечего.

*'Goodluck and dont be so pessimistic! **Haha** i like guys who play the piano because they tend to exude a certain charisma so dont worry becoss u will find someone who appreciates the same thing as u! Stay confident!'* В этом примере '**Haha**' выражает чувство умиления и поддержки, близкое к улыбке. Данную реакцию также невозможно интерпретировать как смех.

*'I'm thinking about buying a nice bike but I'm not sure which is best. (I'm obviously a beginner **haha**) also if you'd like to reccomend a good helmet that would be awesome.'* Выражение чувства неуверенности, ожидание поддержки, но не обозначение смеха.

¹⁹ Yourdictionary.com <http://www.yourdictionary.com/ha-ha>

²⁰ TheUrbanDictionary <http://www.urbandictionary.com/define.php?term=haha>

*'I am 12 and there is this boy in my form who always flirts with me and calls me cute and plays with my hair. He always says he likes me and he would do anything to marry me **haha**.'* Данное выражение можно рассматривать как обозначение сомнения, а также умиления, насмешки над происходящим.

*'That, and the fact he was from the old-school of law lecturing where you were taught the purpose of the law rather than the actual law **haha** ;-).'* Этот случай можно описать как усмешку, подтрунивание над ситуацией.

*'Homefree – I love that!!! And no – I can NOT imagine having 90% of my income free like that! **Haha** wow. What a great affect for sure.'* В данном примере 'Haha' выступает в качестве способа выражения удивления, осознания того, что кажется невероятным.

В результате можно сделать вывод, что междометие 'Haha' умеет мало общего с обозначением смеха, скорее оно подразумевает сложные чувства, которые трудно описать словами, служит для обозначения тех невербальных реакций, которые присутствуют при живом диалоге.

Русскоязычные словари опять же не фиксируют значение данной единицы, поэтому воспользуемся примерами из нашего русскоязычного корпуса.

*«Антон Евстратов@ Ведет влево при резком торможении Заменяли колодки по кругу и диски передние - проблема осталась. Сегодня съездил на диагностику подвески, выяснилось, что подклинивает задний правый суппорт, разбиты стойки и втулки переднего стабилизатора, разбита правая шаровая опора и сайлентблоки переднего рычага. Еще утверждают, что в тормозной системе воздух. Что скажите, в чем может быть дело? Radik@ **xaha**)) вот когда все это устроишь тогда и задавай вопрос)) хотя мне некажется что у тебя будут они)»* Автор находит забавным то, что описывает другой пользователь. Выражение усмешки.

«управдом друг человека@ к чему приведут новые понятия в России?: если можешь воровать-умный. не воруеть-дурак. (коррупция не одну империю сгубила) и Римская и Византийская империя издохли в первую очередь от коррупции, и как следствие -пропасти между верхами и низами. артур якоецко@ Ха , ха.. империя... куча болота и москва сбоку...»
Выражение насмешки, сарказма.

«Просто пример: Яндекс показывает на сайт 191 ссылку из вебмастера, 208 из архива, а ahrefs показывает 599 ссылок. Причем по сути это почти один сквозняк с потенциалом в пару тысяч бэков. Итак. ahrefs показывает 1/4 от общей массы, а сам Яндекс не увидел и 1/8. Кстати всеми любимый YBL показывает и вовсе 33 ссылки. ха ха!» Выражение удивления, усмешки.

«А президента Грузии могу оскорблять как угодно - заслуживает! (Демократический правитель - ха ха - тебе ссылку на ролик как он из "калашников" поливает в парламенте Грузии , при приходе к власти дать ?) -, равно как и других президентов других стран(если заслушивают) - это не национальная рознь .» Выражение насмешки, сарказма.

Примеры из русскоязычного корпуса показывают, что междометие «Хаха» выражает множество оттенков значения, связанного с насмешкой или усмешкой, близко к выражению сарказма, но так же как и в английском языке, далека от обозначения полноценного смеха, так же выражает смешанные и сложные чувства, которые понятны интуитивно в сочетании с мимикой в живом диалоге. Интернет-коммуникация лишена возможности передавать жесты, мимику и другие компоненты живого общения, в ней существуют такие же ограничения и такие же способы выражения эмоций, как и в традиционных формах письменной речи. Англоязычное междометие 'HaHa' и русскоязычное «Хаха» можно считать полными эквивалентами, которые способны передавать аналогичные значения в относительно одинаковых контекстах.

• Hahaha

Данное междометие относится не только к интернет-коммуникации, но и традиционно используется в литературном языке для обозначения смеха. Например, в словаре В.И. Даля зафиксировано такое значение:

«ха-ха-ха! *межд.* громкого хохота, смеха вслух. Хахать *новг.* хахалиться *яросл., костр.* громко смеяться, хохотать. *Ему на боку дыру вертят, а он: ха-ха-ха! Чужой дурак - ха, ха! А свой дурак - ох, ох!»*²¹

Согласно данным Оксфордского словаря²², в английском языке смех обозначается выражением ‘Oh ha-ha’, например, ‘*Oh ha ha, very funny, I said ‘I know you!’ he said. ‘He he,’ she chuckled. ‘Do you?’*”

Рассмотрим употребление междометия ‘Hahaha’ в языке интернет-коммуникации. Англоязычный словарь [Netlingo.com](http://www.netlingo.com) дает такое толкование:

Hahaha it means laughing.²³

[UrbanDictionary.com](http://www.urbandictionary.com) определяет более распространено:

hahaha To express on AIM when something was funny, because just ‘haha’ isn’t that dramatic and can be used as just acknowledging when someone has said something.

Bobby31: ...they call it cheese and quackers!

Annette85: haha

Bobby31: oh come on it was funnier than that

*Annette85: hahaha*²⁴

Таким образом, междометие ‘hahaha’ в английском языке обозначает смех, реакцию на что-либо действительно смешное.

²¹ Даль В.И.. Толковый словарь Даля, 1863-1866.

²² Oxford dictionary <http://www.oxforddictionaries.com/definition/english/ha-ha#ha-ha-1>

²³ NetLingo.com <http://www.netlingo.com/dictionary/h.php>

²⁴ UrbanDicrionary <http://www.urbandictionary.com/define.php?term=hahaha>

Подобным образом мы рассмотрели остальные междометия и можем сделать следующее соотношение:

HA – bwahaha - - ХА – бугага (бугагашеньки)

Нaha – lol - - хаха – лол (ололо)

Нahaha – lmao – rofl - - хахаха (ахаха, азаза) – рофл – ржунимагу

Wtf – WAT - - взф – ВАТ

Damn – dam - - дэм

С помощью анализа словарных значений и рассмотрения конкорданса нам удалось выяснить значение междометий, выражающих эмоции. Толкования, представленные в словарях, не всегда объективно и полно отражают реальную языковую картину. Тему выражения эмоций англоязычные словари языка интернет-коммуникации освещают в более широко, чем русскоязычные. Примеры из конкордансов наших англоязычного и русскоязычного подкорпусов позволили произвести обзор речевых примеров, показывающих, что эмоциональные междометия имеют широкий спектр значений и оттенков значений, которые не всегда можно в полной мере выразить словами, они служат для того, чтобы компенсировать недостаток компонентов живого диалога, которые не могут быть представлены в сети. Наш корпус не обладает большим объемом, но позволяет решить задачу изучения таких междометий.

2.2. Скетчи и Тезаурус

Инструмент Word Sketch предназначен для анализа текстового материала корпуса, с помощью которого можно определить, в какие семантические отношения вступает искомая единица. Все случаи употребления классифицируются согласно выполняемой синтаксической роли, а также есть дополнительная функция кластеризации результатов.

Существует возможность просмотра частотности употребления и просмотра фрагмента текста по каждому случаю.

Инструмент Thesaurus подразумевает построение тезауруса на основании дистанции употребления в контексте слов и частотности близкого употребления. Принцип работы инструмента таков: производится поиск однотипных словосочетаний, в которых одно и другое слово выполняют одинаковые синтаксические роли, например, выступают объектами после одного глагола. Затем им присваиваются коэффициенты, и вычисляется дистанция между одним и другим словом.

Попробуем использовать данные инструменты для рассмотрения такого понятия как ‘troll’ и соответствующий ему словарный эквивалент русского языка. Но начнем с анализа словарных значений. Англоязычные словари дают следующие толкования слова ‘**troll**’:

TheUrbanDictionary.com:

troll *noun* One who posts a deliberately provocative message to a newsgroup or message board with the intention of causing maximum disruption and argument.²⁵



Рисунок 2. Изображение тролля.²⁶

²⁵ UrbanDictionary <http://www.urbandictionary.com/define.php?term=troll>

²⁶ UrbanDictionary http://res.cloudinary.com/urbandictionary/image/upload/a_exif.c_fit,h_200,w_200/v1421526525/tj7gukeidrt8giwgivpl.jpg

English Wiktionary на Yourdictionary.com:

Troll *Noun* (plural *trolls*)

1. (fantasy) A supernatural being of varying size, now especially a grotesque humanoid creature living in caves or hills or under bridges. [from early 17th c.]
2. (slang) An ugly person of either sex, especially one seeking sexual experiences.

Verb (third-person singular simple present *trolls*, present participle *trolling*, simple past and past participle *trolled*)

8. (intransitive, Internet slang) In an online community or discussion, (to post inflammatory material so as) to attempt to lure others into combative argument for purposes of personal entertainment and/or gratuitous disruption. [from late 20th c.]
9. (Internet slang) By extension, to incite anger (outside of an internet context); to provoke, harass or annoy.²⁷

Русскоязычный словарь Slanger.ru дает определение, в целом совпадающее с англоязычными словарями:

Троль - провокатор. Чаще всего применяется в интернете, но иногда и к провокаторам в оффлайне.²⁸

По данным из словарей, изначально 'troll' является литературным персонажем, представляющим из себя монстра разных размеров, живущего в пещерах, холмах или под мостами, а в языке интернет-коммуникации под данным названием подразумевается пользователь (типаж пользователя), размещающий сообщения провокативного характера с целью над всеми подшутить, вызвав дискуссии, споры и ругань, и получающий от этого удовольствие. Либо это может быть несимпатичный человек

²⁷ English Wiktionary <http://www.yourdictionary.com/troll>

²⁸ Slanger.ru http://slanger.ru/?mode=library&sl_id=1818

противоположного пола, ищущий знакомств с целью завести сексуального партнера.

Для того чтобы узнать особенности употребления данного слова и выяснить более подробные характеристики 'troll' как персонажа, применим инструменты Word Sketch и Thesaurus.

Результаты Word Sketch с произведенной кластеризацией для схождения 'troll', noun:

Troll

Object of

Feed 14

Encounter 2 hate 15 ignore 6 spot 1

Battle 12

Attack 4 chase 1 defeat 5 fight 2 hunt 1 kill 5

Annoy 21

Discourage 2 motivate 3 scare 1 slow 1

Troll 8

Smell 3

Tolerate 2

Win 10

Subject of

Attack 8

Hate 2 ruin 4

Tend 3

Attempt 1

Infest 1

Overrun 1

Lurk 1

Guard 1

Modifier

Anonymous 8

Patent 7

Orc 10

Dwarve 3 elf 4 goblin 7 ogre 5

Leftist 3

Liberal 1 right-wing 1

Hideous 1

Homestuck 1

Christian 1

Internet 3

Modifies

Ogre 3

Dwarve 1 goblin 1 elf 1 ogre 1

Druid 2

Mage 1 shaman 1

Spammers 2

Idiot 2

Fool 1

Troublemaker 1

And/or

Orc 12

Dwarve 3 elf 2 goblin 4 ogre 1

Spammers 3

Hacker 1 spam 1

Hater 2

Racist 1

Vandal 1

Moron 2
 Undead 1
 Gnome 1
 Lurker 2
 Unicorn 1
 Bastard 1
 Monster 1
 Zombie 1
 Dragon 2
 Bigot 1

pp obj of-i 4

tribe 1
 army 2
 horde 1

pp obj with-i 4

deal 3
 argue 1

pp obj of-i 2

afraid 2

pp obj to-i 2

respond 1
 reply 1

pp obj by-i 2

attack 1
 overrun 1

На основе анализа полученных результатов можно прийти к выводу, что слово ‘troll’ в английском языке интернет-коммуникации имеет 2 значения: с одной стороны, оно обозначает выдуманного персонажа, скорее всего встречающегося в компьютерных играх, а с другой стороны, типаж пользователя, ведущего себя провокационным образом. Соответственно, представленные лексические единицы можно разделить на 2 группы:

1. Единицы, относящиеся к выдуманному персонажу:

Похожие персонажи: elf, ogre, orc, goblin, unicorn, gnome, monster, dwarves, moron, zombie, druid, undead, mage, shaman, dragon,

Объединения: tribe, army, horde,

присваиваются качества этого персонажа, характеристики действий, а также аналогичными являются способы борьбы с такими пользователями, поскольку их действия не одобряются, таких пользователей «ненавидят» и «боятся», стараются не попадать в его ловушку, с ними страшно вступать в спор и отвечать на их сообщения. Такие пользователи ассоциируются с ‘spammers’, ‘hacker’, ‘hater’, ‘racist’, ‘vandal’, ‘moron’, ‘bastard’, ‘idiot’, ‘fool’, ‘troublemaker’.

На русскоязычной корпусе Sketch Engine показывает следующие результаты с применением кластеризации для вхождения «тролль», сущ.:

Тролль

Subject of

Троллить 3

Наводнять 2

Завестись 5

 Попасться 1

Набежать 4

Усмехнуться 1

Возмущаться 1

Появиться 1

A modifier

Форумный 4

Пещерный 5

Болотный 3

Злобный 7

 Алчный 1 глупый 1 коварный 1 бездарный 1 наглый 1 подлый 1
 злой 3 тупой 1 кровожадный 1 мерзкий 1 хитрый 1 алчный 1 ужасный 1
 гнусный 1

 Проплаченный 2

 Американский 3

Госдеповский 2

И/или

Гоблин 5

Гном 4

Маг 1 колдун 1 эльф 1 огр 1 дракон 1 вампир 1 хоббиты 1

Орк 6

Лжец 1

Провокатор 1 хам 1 грубиян 1

Флудер 1

Спамер 1

Неадекват 1

Сударь 1

Gen modifies

Логово 2

Обиталище 1

Полчище 3

Набег 1 нашествие 1 натиск 1

Дубина 1

Сборище 1

Кучка 1

Игнорирование 1

Провокация 3

Нападки 1 выпад 1

Орда 1

Набор 1

Object2 of

Кормить 5

Убивать 3

Атаковать 4 встретить 3 убить 1 ударить 1

Побеждать 2

Распознать 2

Игнорировать 3

Pr obj c

Сражаться 3

Бороться 1

Расправляться 2

Спорить 4

Схватка 2

Битва 1 сражение 1

Pr obj для

Кормушка 1

Pr obj у

Плен 1

По полученным результатам так же можно разделить лексику на две группы согласно двум типам обозначаемых объектов, персонажа из компьютерных игр и человека, принимающего стратегию поведения этого персонажа, склонного к провокациям и спорам:

1. Характеристики: злобный, пещерный, болотный, алчный, глупый, коварный, бездарный, наглый, подлый, злой, тупой, кровожадный, мерзкий, хитрый, алчный, ужасный, гнусный,

Подобные персонажи: гоблин, гном, маг, колдун, эльф, огр, дракон, вампир, хоббит, орк,

Действия субъекта: наводнять, завестись, попасться, набежать,

Действия над объектом: кормить, убивать, атаковать, убить, ударить, побеждать, сражаться, бороться, расправляться,

Жилище: логово, обиталище,

Оружие: дубина,

Объединения: полчище, сборище, кучка, орда,

Род занятий: битва, схватка, сражение, набег, нашествие, натиск,

2. Действия субъекта: троллить, усмехнуться, возмущаться, появиться, проковации, нападки, выпады,

Действия над объектом: распознать, игнорировать, спорить с,

Характеристики: форумный, американский, госдеповский, проплаченный,

Синонимы: лжец, провокатор, хам, грубиян, флудер, спамер, неадекват, сударь.

С помощью инструмента Thesaurus можно построить облако ассоциаций, семантически связанных понятий:

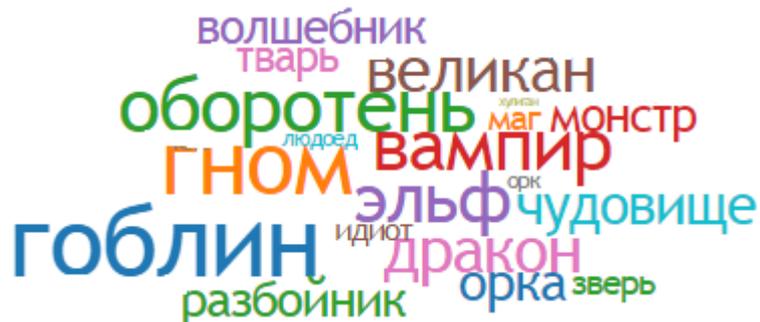


Рисунок 4. Графическое представление тезауруса к «троль».

По результатам анализа, прослеживается аналогия с англоязычным представлением о персонаже и о пользователе, именуемом словом «троль», как о выдуманном герое компьютерных игр, отличающимся грубостью, хамством, наглостью и лживостью, чьи поступки сравнимы с вандализмом. По роду своей деятельности персонаж предназначен для битв, сражений, набегов стаями и атак, живет в логове, болоте или пещере, питается из кормушки, его может кто-либо содержать для боевых целей. В языке интернет-коммуникации происходит перенос качеств данного существа на пользователей, ведущих себя неделикатно, провоцирующих споры, атакующих своими репликами, которого важно распознать, чтобы не попасть к нему в кормушку или плен. Поскольку тролля может кто-либо содержать,

то в русскоязычной картине интернет-коммуникации у тролля появляются дополнительные признаки в связи с политической обстановкой, применимые для людей, отстаивающих позицию правительства США, по представлениям пользователей, за определенную плату: «проплаченный», «госдеповский», «американский». В англоязычной же картине, присутствует компонент навязывания политические и религиозных взглядов, что характерно для некоторых распространенных в западных странах религиозных течений: ‘Christian’, ‘leftish’, ‘right-wing’.

Проведя на основе нашего корпуса анализ интернет-понятий ‘troll’ в английской языке и «тролль» в русском, мы может установить эквивалентные связи между некоторыми выражениями, учитывая, что эти понятия обозначают аналогичные объекты действительности:

Troll – тролль

To feed a troll – кормить / покормить / накормить тролля

Horde of trolls – орда троллей

Tribe of trolls – стало / полчище троллей

A bunch of t. – кучка / сборище троллей

Patent troll – патентный т.

To kill a t. – убить т.

To win a t. – победить т.

To encounter a t. – встретить / наткнуться на т.

To attack a t. – атаковать т.

To argue with a t. – спорить с т.

To ignore a t. – игнорировать т.

To overrun – кишеть / наводнять

To be afraid of t. – бояться т.

Homestuck t. – домашний т.

Рассматриваемые инструменты, Word Sketch и Thesaurus, успешно применимы для анализа сочетаемости и употребления лексических единиц, они помогают раскрыть содержание понятия, обнаружить коллокации,

которые не всегда очевидны, позволяют проследивать семантическую близость понятий с использованием средств визуализации.

2.3. Сравнение скетчей

Инструмент Sketch-Diff предназначен для того, чтобы обеспечить возможность лаконично сравнить сочетаемость двух вхождений, просмотреть имеющиеся у них общие и уникальные комбинации с коэффициентом близости, а также можно сравнить употребление одного и того же вхождения в разных частях корпуса и посмотреть вхождения, употребленные через союз и / или. Данные функции подходят для изучения близких понятий в языке интернет-коммуникации, например, ‘friendship’ и ‘friendzone’ / «дружба» и «френдзона».

Англоязычный словарь the Random House Dictionary, представленный на сайте Dictionary.com, так описывает значение слова ‘friendship’:

Friendship, *noun*

1. the state of being a friend; association as friends: to value a person's friendship.
2. a friendly relation or intimacy.
3. friendly feeling or disposition.

Synonyms: harmony, accord, understanding, rapport.²⁹

friend-zone

noun

1. *Slang*. A friendship in which one person, typically male, is romantically or sexually attracted to the other, but the attraction is not mutual: *He's obviously in love with her but she keeps him I the friend zone.*
2. *verb (used with object), friend-zoned, friend-zoning*.
Slang. to put (someone) in the friend zone: *He wonders why he's always getting friend-zoned by women he likes.*³⁰

²⁹ The Random House Dictionary <http://dictionary.reference.com/browse/friendship?s=t>

³⁰ The Random House Dictionary <http://dictionary.reference.com/browse/friend%20zone?s=t>

Словарь The UrbanDictionary представляет несколько значений ‘friendship’, два из них звучат следующим образом:

Friendship a relationship between two individuals that share a mutual bond of trust and liking as well as commitment and personal feelings and experiences, which allows them to benefit from each other.

"A friend is one who walks in when others walk out" - a friendship quote³¹

Friendship is something between two or more people who can trust each other and that will never turn their back on each other no matter what.

Friendship can last forever if its with the right people. People you can trust.³²

Friend Zone What you attain after you fail to impress a woman you're attracted to. Usually initiated by the woman saying, "You're such a good friend". Usually associated with long days of suffering and watching your love interest hop from one bad relationship to another. *Verb tense is "Friend-ed"*.

"I spent all that money on a date, just to find out she put me in the Friend Zone(said with eerie echo)."

"You know that hot girl I've been talking to? She just Friend-ed me."³³

Получается, что понятие ‘friendship’ в английском языке можно определить как взаимную связь между людьми, основанную на доверии и симпатии, общих интересах и совместном жизненной опыте, от которой эти люди получают поддержку и пользу. На друга можно рассчитывать, он не отвернется ни при каких обстоятельствах.

Понятие ‘friendzone’ подразумевает результат отношений после неудачной попытки произвести впечатление на понравившуюся женщину. Обычно это касается мужчин, долго страдающих после того, как женщина

³¹ UrbanDictionary <http://www.urbandictionary.com/define.php?term=friendship>

³² UrbanDictionary <http://www.urbandictionary.com/define.php?term=friendship>

³³ UrbanDictionary <http://www.urbandictionary.com/define.php?term=Friend+Zone>

предлагает остаться друзьями, тогда как они рассчитывали на большее и не получили взаимности.

Для того чтобы выяснить разницу между этими понятиями, воспользуемся инструментом Sketch-Diff на нашем англоязычном корпусе.

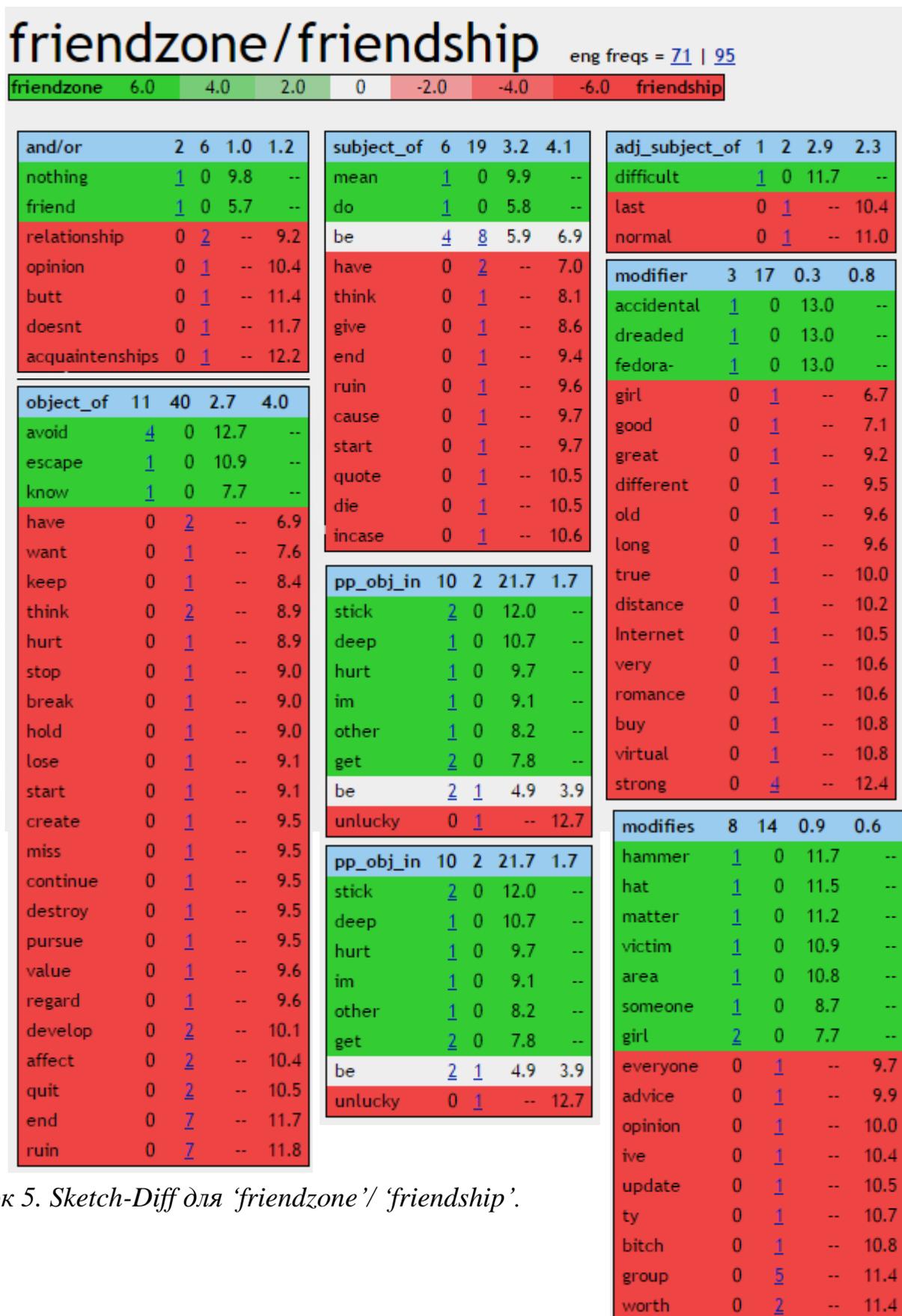


Рисунок 5. Sketch-Diff для 'friendzone' / 'friendship'.

Если рассмотреть полученные результаты, то становится ясным, что понятие ‘friendzone’ близко с такими словами, как ‘nothing’, ‘friend’, что подтверждает суждение о том, что желаемый результат не достигнут, ничего не получилось. Данное слово выступает в роли объекта в словосочетаниях ‘avoid of f.’, ‘escape of f.’, ‘get out of f.’, что говорит о том, что попадания в данное положения избегают или стараются из него выбраться. Это слово так же может быть объектов в выражениях ‘get in f.’, ‘hurt in f.’, ‘stick in f.’, ‘deep in f.’, ‘be in f.’, указывающих на то, что попадание в ‘friendzone’ приносит страдание, оно не приносит удовольствия, затягивает и тяготит. Такое положение воспринимается как трудное (‘difficult’), человек чувствует себя жертвой (‘victim’), и обычно это связано с девушкой или женщиной (‘girl’).

Понятие же ‘friendship’ близко к понятиям ‘relationships’ и ‘acquaintances’, выступает объектом в отношении к глаголами, которые можно разбить на несколько семантических групп:

1. бытийные глаголы: ‘have’, ‘want’, ‘think’, ‘miss’, ‘pursue’, ‘value’, ‘affect’, ‘hurt’,
2. глаголы, обозначающие начало чего-либо: ‘start’, ‘create’,
3. глаголы, обозначающие продолжительность чего-либо: ‘keep’, ‘hold’, ‘continue’, ‘develop’,
4. глаголы, обозначающие завершение: ‘stop’, ‘loose’, ‘quit’, ‘end’,
5. глаголы, обозначающие разрушение: ‘destroy’, ‘ruin’, ‘break’.

Также ‘friendship’ принимает роль субъекта действия с глаголами ‘give’, ‘end’, ‘ruin’, ‘start’, ‘die’, ‘incase’.

Таких отношений как ‘friendship’ хотят, их ценят, о них думаю, их не хватает. Сами же отношения можно начать, продолжать и развивать их, а можно разрушать и наносить отношениям вред. Происходит присвоение понятию ‘friendship’ качеств объектов, которые можно построить и разрушить, например, здания. Значит, над такими отношениями работают, их выстраивать в длительном процессе, но также они могут быть разрушены,

что вызывает негативные эмоции. Также понятию ‘friendship’ присваиваются свойства живого организма, который способен умирать, давать. Такие отношения могут начинаться и заканчиваться сами, самостоятельно разрушаться и укрепляться.

У понятия ‘friendship’ выявлены характеристики, которые можно отнести к нескольким группам:

1. по половой принадлежности: ‘boy’, ‘girl’,
2. по качеству: ‘good’, ‘great’, ‘true’, ‘strong’, ‘old’, ‘long’, ‘romance’,
3. по типу контакта: ‘distance’, ‘Internet’, ‘virtual’.

Можно отметить, то понятие ‘friendship’ имеет положительные характеристики по качеству: это могут быть хорошие, длительные, крепкие отношения, старые и с присутствием романтики, разграничивается понятие женской и мужской дружбы, контакт может быть виртуальным и дистанционным, опосредован техническими средствами связи.

Рассмотрев особенности двух понятий, ‘friendzone’ и, мы можем прийти к выводу, что отношения, обозначаемые словом ‘friendship’, расцениваются положительно, люди создают такие отношения, работают над тем, чтобы были крепкими, долгими, приносили удовольствие и пользу, однако их они могут быть разрушены или могут закончиться. Такие отношения могут существовать виртуально, без живого контакта. Понятие же ‘friendzone’ имеет противоположные характеристики, такие отношения нежелательны, их стараются избегать, от них нужно избавляться, находить способ от них освободиться, потому что они приносят только отрицательные эмоции. Такие отношения не имеют продолжения, они являются результатом неудачной мужской попытки завязать более серьезные отношения, и такое предложение дружбы не ценится. Этот тип связи происходит в реальном мире, не в сети, он не приносит пользу, не требует верности, воспринимается мужчинами как ловушка.

Рассмотрим значение понятий «френдзона» и «дружба» в русском языке интернет-коммуникации, начав с прочтения толкований в словарях.

Teenslanger.su

Френдзона, *-ы, ж.* дружба, когда парень любит девушку, а она принимает его только как друга. *Самое главное - если вы попали во френдзону - осознать это самостоятельно, и понять, что у вас теперь есть проблема.* • *Френдзона - это ты признался девушке, сказал все как есть, а она в ответ, извини, мы друзья, иди так или никак.* • *Настоящая френдзона - это унижение.* • *Иногда френдзона – это не следствие, а приговор, причем мгновенный и не обсуждаемый.*³⁴

Луркморье

Зона дружбы — это одна из старейших и наиболее распространённых ловушек, которые женщины расставляют для ввержения хороших, годных мальчиков в вечную пучину ментального рабства с дальнейшим превращением их в лузеров.³⁵

Русскоязычные словари интернет-языка описывают значение «френдзоны» как отношения между мужчиной и женщиной, характерными чертами которых являются отсутствие взаимности со стороны женщины и негативное эмоциональное восприятие таких отношений мужчиной, ассоциирующихся с пленом и ловушкой.

Понятие «дружба» не представлено в интернет-словарях, поэтому обратимся к классическим словарям:

Толковый словарь Ожегова

ДРУ́ЖБА, *-ы, жен.* Близкие отношения, основанные на взаимном доверии, привязанности, общности интересов. *Давнишняя д. Д.*

³⁴ Teenslanger.su <http://teenslang.su/id/16405>

³⁵ Lurkmore http://lurkmore.to/Friend_Zone

одноклассников. Не в службу, а в дружбу (не по обязанности, а из дружеского расположения; разг.). прил. дружеский, -ая, -ое.³⁶

Большой толковый словарь

ДРУЖБА, -ы; ж. 1. Отношения, основанные на взаимной привязанности, доверии, духовной близости, общности интересов и т.п. *Многолетняя, старая, давняя д. Крепкая, прочная, верная д. Горячая, искренняя нежная д. Братская, мужская, солдатская, фронтовая д. Завязать дружбу с кем-л. Жить в любви и дружбе. Сделать что-л. из дружбы, в знак дружбы. Водить дружбу с кем-л. (поддерживать такие отношения). Сделать что-л. не в службу, а в дружбу (не по обязанности, а из расположения).* 2. Отношения между народами, странами, государствами, основанные на доверии, взаимопонимании и невмешательстве во внутренние дела друг друга. *Д. народов. Политика мира и дружбы. Договор о дружбе и взаимопомощи. Встреча глав государств проходит в обстановке дружбы и взаимопонимания. Нерушимая, непоколебимая д. Посадить дерево дружбы (как символ таких отношений между народами). Поезда, рейсы дружбы (в СССР: туристические поездки в дружественные страны с какой-л. культурной или общественно-политической целью).³⁷*

Словарь синонимов

Дружба, дружество, дружелюбие, доброжелательство, лад, мир, согласие, панибратство, короткое знакомство, побратимство, (благо)приязнь, амикошонство, любовь, братание, единение, общение; дружба искренняя, лицемерная, собачья, тесная. Сделать что по дружбе. Лады всего дороже. Обольщенный моею славою, он стал было искать моего дружества. *Пушк.*

³⁶ Толковый словарь Ожегова <http://dic.academic.ru/dic.nsf/ogegova/53126>

³⁷ Большой толковый словарь

<http://www.gramota.ru/slovari/dic/?all=x&word=%25E4%25F0%25F3%25E6%25E1%25E0>

Прот. Вражда. Ср. Любовь, Согласие и Союз. быть в дружбе, водить дружбу, разорвать дружбу, сводить дружбу.³⁸

По данным словарей, «дружбой» считаются отношения между мужчиной и женщиной, основанные на доверии, согласии, привязанности, духовной близости и общности интересов. Также дружба понимается как отношения стран, основанные на принципе доверия, сотрудничества, поддержки и невмешательства.

Инструмент Sketch-Diff дает результаты, по которым понятие «френдзона» отличается от понятия «дружба» тем, что вступает в отношения со следующими словами:

В качестве субъекта действия с глаголами: «нравится», «быть»,

В качестве обстоятельства места с предлогом «в»: «попасть» 5, «бывать» 2, «сидеть» 4, «поместить» 1, «поселяться» 1, «отправить» 1, «кидать» 1, «перевод» 1, «заходить» 1, «засунуть» 1, «забираться» 1, «швырять» 1,

В качестве обстоятельства места с предлогом «из»: «выйти» 2, «вытаскивать» 1, «выходить» 1,

В качестве объекта: «вызывать» 1, «предлагать» 1, «водить» 1,

В качестве объекта в Род.п.: «житель» 1, «разрушитель» 1,

С определением: «безнадежный» 1.

Данные показывают, что понятие «френдзона» представляется не столько как отношения, сколько как место, куда попадают, из которого нужно выбираться самому или вытаскивать кого-либо из него. Это место имеет признаки помещения, в котором можно поселиться и которое нужно

³⁸ Словарь синонимов

<http://www.gramota.ru/slovari/dic/?all=x&word=%25E4%25F0%25F3%25E6%25E1%25E0>

разрушить. Вероятно, происходят ассоциации с понятием «зона» как с тюремным заключением. Встретившиеся лексические единицы имеют яркую негативную эмоциональную окраску, что говорит о том, что данное понятие вызывает с отрицательной точки зрения, затрагивает чувства.

Для понятия «дружба» получены следующие данные:

Выступает в роли субъекта: «завязаться», «крепнуть», «зародиться», «длиться», «сломаться», «продолжаться», «проверяться», «перерасти», «рушиться», «слабеть», «победить», «соединить», «возобновиться»,

С определениями: «крепкий», «давний», «тесный», «нерушимый», «нерушимый», «интернациональный», «российско-китайский», «долгий», «взаимный», «старый», «мужской», «истинный», «подлинный», «женский», «близкий», «вечный», «многолетний», «неразрывный»,

В качестве объекта в Род.п.: «народ», «дорожить»,

В качестве обстоятельства места с предлогом «в»: «класться», «верить», «жить», «верность», «перерасти», «преданный»,

В качестве обстоятельства места с предлогом «из»: «вырасти», «рождаться», «вытекать», «выгода».

Выходит, что понятие «дружба» включает в себя два компонента: отношения между людьми и отношения между государствами. Такие отношения можно охарактеризовать как крепкие, взаимные, они могут быть прочными, длительными и истинными. Разграничиваются понятия женской и мужской дружбы. Дружба может зародиться, крепнуть или слабеть, что говорит о присвоение дружбе свойств живого организма, может ломаться, длиться, перерасти в нечто иное. Актуальными на сегодняшний день дружескими интернациональными отношениями являются российско-китайские.

С помощью инструмента Sketch-Diff нам удалось произвести сравнение четырех понятий из английского и русского языка, обнаружить их общие и уникальные свойства. Результаты проведенного контрастного анализа привели нас к выводам, что между рассмотренными понятиями прослеживаются эквивалентные связи, но есть и асимметрия. Если сравнивать понятия «френдзона» и «дружба» в русском языке, то первое имеет негативное восприятие, означает невзаимные отношения между мужчиной и женщиной, неудачную попытку перейти на более серьезный этап отношений, они приносят мужчине страдания, в которые он погружается и может в них находиться долгое время, поэтому его нужно из них спасать и реабилитировать. Второе же понятие имеет противоположное значение, подразумевает приятные взаимные отношения между людьми, в которые нужно верить, им свойственна верность, взаимопонимание, взаимная выгода. Существует надежда на то, что они могут длиться долго, быть вечными и прочными.

В целом, понятия «дружба» и ‘friendship’ эквивалентны, они описывают аналогичную форму отношений, которые воспринимаются одинаково положительно. Такие отношения имеют одинаковую основу, одинаковых участников и предъявляют одинаковые требования к участникам. Однако в английском языке присутствует представление о том, что такие отношения нужно строить, над ними нужно работать, тогда как в русском языке в такие отношения нужно верить, это больше духовные отношения, чем деятельностные, они требуют преданности и единства, общности взглядов. Также в русском языке выражено представление об интернациональной дружбе, о братстве народов. Таким отношениям присваиваются свойства личных отношений.

Сравнение понятий «френдзона» и ‘friendzone’ показало, что они могут считаться эквивалентными, одинаково обозначают негативно воспринимаемые отношения, которые не подлежат дальнейшему развитию. В

таких отношениях мужчина оказывается жертвой отсутствия взаимных намерений со стороны женщины, что можно считать отрицательным результатом попытки завязать серьезные отношения. Это состояние болезненно для мужчин, вызывает сильные эмоции. Френдзона воспринимается как заключение, как ловушка и плен, близкое к понятию тюремного заключения в русском языке, из которого нужно освобождать человека, помогать ему выбраться, тогда как в английском языке больше актуально восприятие попадания во френдзону как результат женской коварности и хитрости, попадания под действие которой следует избегать. Попадание в френдзону в русском языке воспринимается как результат оказанного неопределенноличного действия на мужчину, которого «кинули», «засунули» в это пространство, освобождать из которого его тоже нужно с посторонней или коллективной помощью, а в английском языке это результат самостоятельного личного действия, собственной неудачи, выходить из этого положения предполагается своими силами, самостоятельно.

Выводы

В данной главе была продемонстрирована методика работы с созданным корпусом русско-английских эквивалентов интернет-коммуникации в корпус-менеджере Sketch Engine, включающая четыре способа анализа понятий, характерных для русского и английского языка интернет-коммуникации: использование конкорданса, скетчей, тезауруса и сравнение скетчей.

С помощью конкорданса на русском и английском подкорпусах были рассмотрены значения междометий, выражающие эмоции: был проведен анализ словарных значений, построен конкорданс для каждого вхождения, представляющий примеры употребления интересующей единицы в контексте. В результате, получен вывод, что словарные значения не всегда объективно приводят толкования междометий.

С помощью инструментов Word Sketch и Thesaurus были изучены понятия 'troll' и «тролль»: для каждого слова обнаружены сочетания на основе семантических связей, классифицированы согласно их типам, затем кластеризованы по семантической близости. Результаты сопровождаются информацией о частотности употребления в корпусе. На основании данных о близости значений составлены тезаурусы с графическим представлением. Анализ полученных результатов позволил установить эквивалентные связи между компонентами значений.

Инструмент Sketch-Diff позволил произвести сравнение двух близких понятий в русском языке, «дружба» и «френдзона», и аналогичных понятий в английском языке, 'friendship' и 'friendzone', зачем произвести их сравнение. На основании статистических данных о сочетаемости единиц построена таблица, демонстрирующая сравнение характерных употреблений одной единицы и другой. Сначала произведено сравнение близких понятий внутри одного языка, затем сравнение понятий между языками.

ГЛАВА 3. СОСТАВЛЕНИЕ КОРПУСА

3.1. Описание собственного корпуса

Построенный нами корпус создан с целью изучения русско-английских эквивалентов интернет-коммуникации на примере текстов из сети Интернет. С помощью данного корпуса возможно выполнить лексикографические и лексико-семантические исследовательские задачи, выявлять сочетаемость единиц, их окружение, определять совпадения и расхождения между ними. Выполнить данные задачи важно именно посредством составления корпуса, так как сделать это вручную достаточно трудоемко и требует больших временных затрат, а корпус позволяет применить автоматизированные инструменты, что ускорит работу, и получить точные результаты на основе статистических данных. Поставленные цели определили тип корпуса, корпус является двухязычным и сравнительным, но не параллельным. Сравнительные корпуса дают возможность изучать языки в компаративном аспекте, строятся из коллекции несвязанных между собой текстов на разных языках, тексты не являются переводами, но обладают общими чертами, которые возможно сравнить. Такие корпуса выгодно отличаются от переводных корпусов тем, что могут использоваться для выявления различий и совпадений в представленных языках.

Корпус предназначается для личного пользования. Его можно охарактеризовать скорее как инструмент для исследования, который можно модифицировать в зависимости от целей, расширять объем, по необходимости структурировать данные. Возможности нашего корпуса не позволяют поставить его в один ряд с представительскими корпусами, созданных для выполнения множества задач, поскольку у него небольшой объем и представлены не все жанры языка интернет-коммуникации.

Для достижения объективного результата исследования мы сделали выбор в пользу сбора доступных текстов из сети, максимально отдаленных от традиционной литературной речи, например, из проектов Otvet.mail.ru на русском языке и Answers.yahoo.com на английском языке. В корпус не входят тексты, содержащие литературные произведения, новости, записи из блогов литературного характера, обзоры и аналитические статьи. Словом, в корпус входят комментарии, сообщения из форумов, личные заметки из блогов, то есть тексты, максимально относящиеся к общению между пользователями. Данные тексты интересны тем, что они отличаются от нормированной письменной речи, речь пользователей уникальна и представляет образец речи именно интернет-коммуникации. Для каждого текста указываются метаданные: ссылка на источник и псевдоним автора. Личность автора и данные о нем достоверно установить практически невозможно, поэтому в аннотации представлены только эти данные.

Прежде чем приступить к сбору текстов, нами был проведен анализ словарей интернет-языка на русской и английском языке, были выбраны наиболее известные и не вызывающие сомнения (учитывая особенности интернет-словарей и методику включения единиц в них) лексические единицы и разбиты на тематические группы. По результатам анализа словарей был создан общий список слов, из которого затем были исключены слова, вышедшие из употребления, потерявшие свою актуальность и встречающиеся достаточно редко. Основной идеей выбора было наличие аналогичного понятия как в русском, так и в английском языке, те, между которыми можно установить эквивалентные связи. Найденные слова используются в качестве ключевых для целевого поиска текстов.

Весь объем слов, представленных в словарях, рассмотреть в рамках нашей работы не удастся, поэтому в список наших ключевых слов попали только следующие группы:

- выражение эмоций: lol, rofl, omg, damn, bwahaha, ha-ha, haha, wtf, HA, WAT,

- наименования типажей пользователей: troll (+ to troll, trolling), epic hero, epic fail, hater, slowpoke, noob, looser,

- чувства и реакции пользователей: facepalm, butthurt, troll face, poker face, 'thanks, Obama'/ Спасибо Путину за это, Not Bad, true story, WAT, copypaste, ... detected, epic win, epic fail, mimimi, нуа/ ня/ няша/няшно, friend zone/френдзона.

На каждое из перечисленных слов в корпус попало по несколько текстов, отражающих разнообразные варианты их употребления, повторения по возможности исключались. Всего в корпус попало по 1000 текстов для каждого языка. Каждый текст включает в себя вопрос и все комментарии к нему. Тексты относятся к различным тематикам. Наш корпус получился относительно небольшого размера по сравнению с современными известными корпусами, но поскольку он имеет узкую специализацию и направлен на индивидуальное исследование, такой размер вполне достаточен. Он не отражает в полной мере всю языковую картину интернет-коммуникации, зато в его состав вошли специально отобранные современные тексты без повторов, что позволит составить представление об изучаемой проблеме, поэтому корпус можно считать относительно репрезентативным. Состав корпуса сбалансирован, содержит однородные по структуре и жанру тексты, взятые из однотипных источников.

Выбор корпус-менеджера напрямую связан с целями построения корпуса и с возможностью его применения для исследований с точки зрения тех функций и инструментов, что доступны для работы. Основными критериями отбора можно считать свободный доступ для использования, понятный интерфейс, наличие автоматической обработки текстов, широкий набор функций помимо построения конкорданса и поиска в формате KWIC.

Поскольку цель нашего корпуса – это сравнение эквивалентов интернет-коммуникации, то корпус-менеджер также должен предоставлять возможность построения сравнительных корпусов и быть способным работать как с английским, так и с русским языком.

Для работы с нашим корпусом было выбрано приложение **Sketch Engine**³⁹, приложение, созданное для построения корпусов и работы с ними. Оно предлагает функции KWIC конкорданса для заранее созданных и загруженных собственных корпусов или для корпусов, автоматически создаваемых в данном приложении. Приложение дает возможность работы с более чем 40 языками, включает в себя крупные корпусы на этих языках, например, BNC, Brown Corpus, BASE, TenTen Web Corpora, BAWE, ARC, Susanne Corpus, WaC Web Corpora, Europarl Parallel Corpora, OPUS Parallel Corpora и остальные (200+). Доступ к сервису платный для использования на срок более месяца.

Sketch Engine включает в себя следующие инструменты:

1. Concordancer, построение конкорданса по запросу для просмотра правого и левого контекста, количества интересующих вхождений с применением фильтров поиска и сортировки,
2. Word sketches, просмотр употребления искомого вхождения, грамматического поведения и сочетаемости при помощи распределения результатов поиска на группы исходя из синтаксических отношений (выступает ли слово в качестве объекта, субъекта, определения или обстоятельства), также на выходе предоставляется информация о частотности употребления.
3. Word sketch difference, визуальное сравнение сочетаемости искомым единицам, их контекста, результаты разбиты на группы по синтаксическим отношениям (использование в роли определения и

³⁹ Sketch Engine <https://the.sketchengine.co.uk>

обстоятельства), по употреблению через сочинительный и противительный союзы. Результаты выделяются цветом.

4. Corpus Architect, возможность создания собственных корпусов из текстов в формате TXT, PDF, PS, DOC, HTML, VERT с последующим поиском по ним.
5. WebBootCat, инструмент для создания собственного корпуса из текстов сети Интернет за достаточно короткое время, для чего нужно ввести примерно 6 единиц, применимых в изучаемой области, затем сервис собирает страницы по данным запросам из любых источников или из указанных заранее.
6. Wordlists, составление списков частотности не только по слову, но и по морфологическому тэгу, коллокациям, N-граммам, грамматическим формам, и, например, вхождениям с какими-либо буквами.
7. Keywords, выделение ключевых слов на основании частотности их появления в корпусе, возможно сопоставить данные с частотностью употребления в сравниваемом корпусе.
8. Terms, инструмент для работы с устойчивыми словосочетаниями как, например, 'sustainable development', 'natural gas', 'New York', 'sea level', 'The United States', 'et al.' и прочие.
9. Distributional thesaurus, инструмент для создания тезаурусов для имен существительных, прилагательных и глаголов (зависит от языка) с указанием частотности употребления.
10. Parallel corpora, эта функция позволяет загружать параллельные корпуса на двух или более языках для дальнейшего изучения, предварительного выравнивания не требуется, но подразумевается работа с аналогичными текстами, например, с текстами-переводами, для их сравнения.

11. Compare corpora, данный инструмент дает возможность сравнить два и более корпуса между собой, строится таблица с коэффициентами их близости.

Приложение Sketch Engine отличается от других тем, что подходит для работы с мультязычными корпусами, позволяет делать сравнение единиц, а также сравнение корпусов в целом, составлять тезаурусы, производит автоматический статистический анализ результатов, с их графическим представлением. Данное приложение отвечает нашим требованиям, позволяет просматривать окружение и сочетаемость единиц, строить тезаурусы, просматривать статистическую информацию на каждом этапе работы. С его помощью нам удалось провести исследования, описанные в главе 2, с применением выбранных инструментов.

3.2. Применение корпуса русско-английских эквивалентов интернет-коммуникации

В настоящее время существует довольно-таки небольшое количество лингвистических корпусов языка интернет-коммуникации, представленных в свободном доступе для русского и английского языка, но они необходимы для получения эмпирических данных и проведения не только лингвистических исследований, но и исследований в области психологии, социологии, рекламы и других наук. Поскольку язык интернет-коммуникации имеет свои характерные особенности, то необходимо постоянное его изучение, описание и дальнейшее применение результатов в прикладных целях, например, для создания и усовершенствования словарей, грамматик, инструментов анализа интернет-сообщений, на основе которых работают приложения, предназначенные для обработки текстов из сети Интернет.

Учитывая, что характерной чертой языка интернет-коммуникации является все же творческое переосмысление языковых единиц и обилие заимствований, как лексических, так и грамматических, следовательно,

разработка способов работы с такими единицами становится актуальной задачей. В процессе проведения исследования на нашем корпусе мы столкнулись с неудачной попыткой распознавания некоторых вхождений, и вследствие, невозможности применения инструментов для анализа. Например, слово «детектед», пришедшее из английского языка, не поддается анализу, потому что имеет особую форму, нетипичную для русского языка. В английском языке оно функционирует как форма причастия прошедшего времени в пассивном залоге, придя в русский язык, оно также выполняет роль страдательного причастия прошедшего времени, например, в выражении «тролль детектед».

Для устранения подобных недостатков, в частности, в плане лексики, на основе корпуса интернет-коммуникации можно составить списки слов, затем вычленить неологизмы и включить в словарь, найти типичные коллокации с последующим описанием правил их применения (в зависимости от цели), параллельно провести этих процедур на двуязычных подкорпусах и сравнить результаты с установлением эквивалентных связей (по необходимости).

В процессе работы под корпусом возникали частные случаи, которые можно решить предложенным способом, но за единичными ситуациями кроются более серьезные проблемы, а именно несовершенство технологии создания корпуса, не учитывающей особенности языка рассматриваемой области: спорадический сбор текстов из сети, изначально исключающий возможность рассмотреть интернет-явления в полном объеме, и ограниченный объем корпусов, что противоречит сущности интернет-коммуникации как непрерывного потока обмена информацией. Следовательно, для того, чтобы корпус служил инструментом для получения достоверных данных в полном масштабе, возникает задача развития технологий создания корпуса интернет-коммуникации.

3.3. Усовершенствование корпуса

Современные корпуса интернет-коммуникации создаются в основном закрытым способом (как и наш корпус): собирается объемная репрезентативная коллекция текстов, обрабатывается, аннотируется и объединяется в корпус. Но, исходя из того факта, что лексический состав языка интернет-коммуникации обновляется с быстрым темпом, постоянно создаются новые тексты и новые понятия, то такие корпуса актуальны только на очень короткий промежуток времени после своей публикации. Необходимы постоянные обновления, и составление и поддержание таких корпусов требует масштабных затрат труда, времени, энергии, ресурсов для хранения и обработки информации.

Корпуса с закрытым источником имеют множество видимых достоинств:

- спланированность,
- структурированность,
- тщательный отбор данных,
- предварительная обработка текстов,
- снятие омонимии,
- наличие метаданных.

Недостатки корпусов с закрытым источником:

- ограниченный объем,
- ресурсоемкость,
- быстрая потеря актуальности,
- неполнота представления данных,
- отставание обновлений от темпа развития языка.

С позиции решения проблемы актуальности текстов и ресурсоемкости, возможно создание метода построения корпуса с открытым источником, который бы позволял автоматически собирать данные в режиме онлайн. То есть, возможно создать универсальный инструмент для создания

и анализа корпусов, совместив идеи традиционных корпусов с закрытым источником и идеи The WebCorp Live Linguist's Search Engine по поиску текстов в режиме онлайн, а также добавить инструменты Sketch Engine для анализа данных.

Для реализации создания такого инструмента можно предложить рассмотрение следующих задач:

- определение методики поиска материала: например, сбор из проверенных источников, на которых происходит непрерывное общение пользователей, без содержания инородной информации, например, новостей или литературных произведений,
- определение объема выборки: объем должен быть относительно крупным, чтобы обеспечить репрезентативность выборки, но возникает проблема технического характера, разработка алгоритмов, позволяющих за короткий период времени обработать большое количество информации,
- выбор методики сбора и предварительная обработки информации: выгрузка текстов, очищение от меток, повторений, хэштэгов, и лишней информации, так, чтобы на выходе оставался только текст и псевдоним автора,
- основная обработка: проведение сегментации от уровня предложения до уровня леммы, грамматическая разметка,
- разработка способа представления данных,
- разработка инструментов для анализа: словник, ключевые слова, конкорданс (фильтры, сортировка), параллельный конкорданс, коллокации, скетчи (классификация, кластеризация), тезаурус с графическим представлением, сравнение скетчей (близость употребления в сочетаниях с другими вхождениями), построение онтологии, частотный анализ, возможность

установления связей между вхождениями корпуса на одном языке и на другом.

Такой инструмент позволит получать материал для строительства корпуса на актуальный момент, получать результаты хорошего качества без лишнего шума (который присутствует в Web Corp и не позволяет доверять результатам поиска) и произвести анализ данных с помощью встроенных инструментов.

Выводы

В данной главе описываются характеристики спланированного и построенного нами двуязычного сравнительного корпуса русско-английских эквивалентов интернет-коммуникации, соответственно состоящего из двух подкорпусов. Корпус собран из текстов по заранее продуманному алгоритму. Корпус создавался для личных целей как инструмент для проведения исследований.

Для работы с корпусом был выбран корпус-менеджер Sketch Engine, представляющий наиболее подходящие инструменты для работы со сравнительным корпусом с закрытым источником. Предлагаемые функции создания корпуса с открытым источником не дали результата, отвечающего нашим потребностям.

Построенный корпус может быть пригоден для выполнения иных целей, например, для создания и усовершенствования словарей интернет-коммуникации, для разработки и развития словарей и грамматик, на основе которых работают приложения для анализа интернет-данных, также на базе корпуса возможно проведение исследований по определению способов работы со специфическими явлениями языка интернет-коммуникации.

В процессе создания и работы с корпусом был сделан вывод о том, что необходимо создание универсального инструмента для построения корпуса с открытым источником и встроенными инструментами для анализа данных. В частности, для работы со сравнительным корпусом необходимо наличие параллельного конкорданса, инструментов для обнаружения, классификации и кластеризации коллокаций, для построения дистрибутивного тезауруса для запрашиваемого вхождения, для сравнения сочетаемости и для установления связи между найденными эквивалентами.

Заключение

Основной целью нашей работы было составление корпуса русско-английских эквивалентов интернет-коммуникации и описание методики работы с ним. Выполнение данной цели повлекло за собой постановку нескольких задач, решение которых изложено в настоящей работе.

Изначально перед нами стояла задача определения сущности лингвистического корпуса и современных требований к его характеристикам, определение особенностей языка интернет-коммуникации, поскольку он послужил объектом нашего исследования, также потребовалось провести обзор уже существующих корпусов интернет-коммуникации для составления представления о том, каким должен быть наш корпус, и рассмотрение теоретических данных о том, каким образом понимается термин «эквивалента» в современном переводоведении, чтобы определить сущность данных, которые должны быть представлены в нашем корпусе.

После проведения предварительной подготовки, нами был спланирован и построен корпус, позволяющий произвести на нем исследование межъязыковых соответствий с применением инструментов, предложенных корпус-менеджером Sketch Engine. А именно, нам удалось провести исследование эмоциональных междометий с помощью составления конкорданса на русском и английском подкорпусах и установить эквивалентные связи между этими выражениями. Наш корпус дал возможность сравнить содержание понятий 'troll' и «тролль» и установить эквивалентные связи между их компонентами. И наконец, нам удалось раскрыть понятия 'friendzone' и 'friendship', «френдзона» и «дружба» и сравнить эти пары между собой. Таким образом, наш корпус отвечает поставленным целям.

После проделанной работы нами было рассмотрено возможное дальнейшее применение корпуса по решению актуальных задач создания и совершенствования словарей интернет-коммуникации и разработке средств для анализа текстов интернет-сообщений.

В итоге, можно сделать вывод, что в ходе работы были решены поставленные цели, мы достигли желаемого результата. Однако обнаружена необходимость развития технологий создания корпусов применительно к языку интернет-коммуникации, а именно создание корпусов с открытым источником, позволяющих получать свежие данные и производить их анализ.

Библиографический список

- 1) *Baisa V.* (2010) Towards disambiguation of word sketches // Text, Speech and Dialogue. // Springer Berlin Heidelberg, P. 37-42.
- 2) *Bell M. J., Schäfer M.* (2013) Semantic transparency: challenges for distributional semantics // Proceedings of IWCS 2013 Workshop Towards a Formal Distributional Semantics, P. 1-10.
- 3) *Cook P. et al.* (2013) A lexicographic appraisal of an automatic approach for detecting new word senses // Electronic lexicography in the 21st century: thinking outside the paper: proceedings of the eLex 2013 conference, 17-19 October 2013, Tallinn, Estonia, P. 49-65.
- 4) *Hasselgård H., Ebeling J., Ebeling S.O.* (2013) Corpus Perspectives on Patterns of Lexis // John Benjamins Publishing.
- 5) *Ferraresi A.* Constructing and evaluating Web corpora: ukWaC. (http://acorn.aston.ac.uk/conf_papers/paper%20-%20Adriano%20Ferraresi.doc)
- 6) *Guevara E.* (2010) NoWaC: a large web-based corpus for Norwegian // Proceedings of the NAACL HLT 2010 Sixth Web as Corpus Workshop. Association for Computational Linguistics, P. 1-7.
- 7) *Gurrutxaga A. et al.* (2013) Automatic Comparable Web Corpora Collection and Bilingual Terminology Extraction for Specialized Dictionary Making // Building and Using Comparable Corpora. Springer Berlin Heidelberg, P. 51-75.
- 8) *Inumella A. et al.* (2009) Associating Collocations with Dictionary Senses // Proceedings of 6th Biennial Conference of the Asian Association for Lexicography, P. 102-113.

- 9) *Kilgarriff A.* (2010) Corpora by Web Services // Web Services and Processing Pipelines in HLT: Tool Evaluation, LR Production and Validation, T. 3077, P. 45.
- 10) *Kilgarriff A. et al.* (2010) A quantitative evaluation of word sketches // Proceedings of the 14th EURALEX International Congress, Leeuwarden, The Netherlands.
- 11) *Kilgarriff A.* THE SKETCH ENGINE AS A COMMON PLATFORM FOR SHOWCASING LANGUAGE RESOURCES // FASSBL6, T. 153.
- 12) *Kilgarriff A., Kosem I.* (2012) Corpus tools for lexicographers // Electronic lexicography. P. 31-56.
- 13) *Kilgarriff A.* Web as Corpus (http://www.itri.bton.ac.uk/~Adam.Kilgarriff/wac_cfp.html)
- 14) *Kruger A., Wallmach K., Munday J.* (2011) Corpus-Based Translation Studies: Research and Applications. Bloomsbury Publishing.
- 15) *Lewandowska-Tomaszczyk B., Thelen M.* (2010) Meaning in Translation. Peter Lang.
- 16) *MULLER-SPITZER C., Kopleinig A., TOPEL A.* (2012) Online dictionary use: Key findings from an empirical research project // Electronic Lexicography.
- 17) Multimodal Parallel Russian Corpus (MultiPARC): Main Tasks and General Structure // LREC 2012 Workshop on Best Practices for Speech Corpora in Linguistic Research. Istanbul, Turkey, May 21, 2012. ELRA, 2012, c.13—16.

- 18) *Parsing D.* (2014) Synthesis Lectures on Human Language Technologies // *Sandra Kubler, Ryan McDonald, Joakim Nivre.* P. 115.
- 19) *Petasis G., Petasis D.* (2010) BlogBuster: A Tool for Extracting Corpora from the Blogosphere // LREC.
- 20) *Rapp R.* (2014) Corpus-based computation of reverse associations // Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14).
- 21) *Schäfer, R.; Barbaresi, A.; Bildhauer, F.* The Good, the Bad, and the Hazy: Design Decisions in Web Corpus Construction. / Proceedings of the 8th Web as Corpus Workshop (WAC-8), 2013 (<http://www.lt3.ugent.be/media/uploads/publications/2013/wac8-proceedings.pdf>)
- 22) *Sharoff S.* (2006) Open-Source Corpora: Using the Net to Fish for Linguistic Data. (<http://corpus.leeds.ac.uk/serge/publications/ijcl06-11-4-proof.pdf>)
- 23) *Sharoff, S.* Classifying Web corpora into domain and genre using automatic feature identification. In Proc. of Web as Corpus Workshop, Louvain-la-Neuve, September, 2007.
- 24) *Sitchinava D.* Parallel corpora within the Russian National Corpus // *Prace Filologiczne*, LXIII, 2012. – p. 271—278.
- 25) *Д.А. Салимова, А.А. Тимерханов.* Двужычие и перевод: теория и опыт исследования. – М. : ФЛИНТА : Наука, 2012. – 280 с.
- 26) *Добровольский Д. О.* Корпус параллельных текстов в исследовании культурно-специфичной лексики // Национальный

корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. – с. 383—401.

- 27) Интернет-коммуникация как новая речевая формация : науч. ред. *Т.Н. Колокольцева, О.В. Лутовинова*. – М. : ФЛИНТА : Наука, 2012. – 328 с.
- 28) Ментальные структуры и их репрезентация лексическими средствами в германских и романских языках [Электронный ресурс]: *О.Н. Прохорова, И.В. Чекулай, Ж.Багана, И.А. Куприева*. – 2-е изд., стер. – М. : ФЛИНТА, 2013. – 332 с.
- 29) Основные понятия переводоведения (Отечественный опыт). Терминологический словарь-справочник / Отд. языкознания; Отв. Редактор канд. филол. наук *Раренко М. Б.* – М., 2010. – 260 с.
- 30) Параллельные корпуса в составе Национального корпуса русского языка как инструмент лексико-грамматической типологии // *Rafael Guzmán Tirado, Irina A. Votyakova (ed.)*. : - Tipología Léxica. Granada, 2013, с. 77—82.
- 31) *Плунгян В. А., Сичинава Д. В.* Национальный корпус русского языка: опыт создания корпусов текстов современного русского языка // *Л.Н. Беляева и др.* Труды Межд. конференции «Корпусная лингвистика-2004».–СПб.: СПбГУ. – 2004. – С. 216-238.
- 32) Поливариантные параллельные тексты в составе НКРЯ // Труды международной конференции «Корпусная лингвистика-2013», СПб., 2013.
(<http://corpora.phil.spbu.ru/Works2013/%D0%A1%D0%B8%D1%87%D0%B8%D0%BD%D0%B0%D0%B2%D0%B0.pdf>)

- 33) *Рябцева Н.К.* Прикладные проблемы переводоведения, - М.: ФЛИНТА, 2014. – 224 с.
- 34) *Рябцева Н.К.* Язык и естественный интеллект. - М.: Academia, 2005. – 640 с.
- 35) *Шаров, С.А.* Представительный корпус русского языка в контексте мирового опыта // НТИ, Сер. 2. 2003. №6. - с. 9-17.