

Федеральное государственное
автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт филологии и языковой коммуникации
Кафедра лингвистики и межкультурной коммуникации
45.03.02 Лингвистика

УТВЕРЖДАЮ
Заведующий кафедрой ЛиМКК
Л.В. Куликова

« _____ » _____ 2016 г.

БАКАЛАВРСКАЯ РАБОТА
МАРКЕРЫ ЛИНГВИСТИЧЕСКОЙ ИСКУССТВЕННОСТИ
ИНТЕРНЕТ-КОММУНИКАЦИИ
(НА МАТЕРИАЛЕ АНГЛОЯЗЫЧНЫХ ЧАТ-БОТОВ)

Выпускник

Т.А. Тихонова

Научный руководитель

канд.пед.наук, доц. Е.В. Ерёмина

Нормоконтролер

М.К. Мжельских

Красноярск 2016

РЕФЕРАТ

Выпускная квалификационная работа по теме «Маркеры лингвистической искусственности интернет-коммуникации (на материале англоязычных чатботов)» содержит 80 страниц текстового документа, 4 приложения, 63 использованных источника.

ИНТЕРНЕТ, КОММУНИКАЦИЯ, ИНТЕРНЕТ-КОММУНИКАЦИЯ, ЧАТБОТ, ТЕСТ ТЬЮРИНГА

Цель работы – выявление основных маркеров лингвистической искусственности коммуникации с программой чатбот.

Задачи работы: проанализировать теоретические основы изучения интернет-коммуникации и выявить ее основные характеристики; описать особенности коммуникации с чатботом; определить чатбот тестом Тьюринга; провести анализ чатбота на лексическом, синтаксическом и текстовом уровне и установить признаки искусственности коммуникации.

Актуальность и новизна данной работы определены слабой изученностью интернет-коммуникации, низким уровнем проработанности данной темы, несмотря на то, что интернет все прочнее входит в нашу жизнь не только как средство работы или учебы, но и как средство общения. Особенности взаимодействия с чатботами практически не изучаются российскими учеными, степень изучения данного вопроса зарубежными исследователями также невысока.

Выделены следующие группы маркеров на лексическом уровне: умение распознавать лишь повседневную лексику и неспособность работы с SMS-сокращениями и сленгом. На синтаксическом уровне: неумение распознавать грамматические неправильно построенные предложения, использование стратегии перефразирования высказывания собеседника, неспособность распознавания местоимений, кроме I, you, it, my, its и your, неумение распознавать разделительный тип вопроса. На текстовом уровне: неспособность работать с высказываниями, состоящими из более чем одного предложения, несоблюдение регистров общения, а также неумение удерживать в памяти весь разговор.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИЗУЧЕНИЯ ИНТЕРНЕТ-КОММУНИКАЦИИ	7
1.1. Теоретические подходы к изучению интернет-коммуникации.....	7
1.2. Основные характеристики интернет-коммуникации	17
1.3. Особенности коммуникации с чатботом	24
ВЫВОДЫ ПО ГЛАВЕ 1	35
ГЛАВА 2. ЛИНГВИСТИЧЕСКИЕ ПОКАЗАТЕЛИ ИСКУССТВЕННОСТИ ОБЩЕНИЯ С ПРОГРАММОЙ «CLEVERBOT»	39
2.1. Определение программы чатбот тестом Тьюринга	39
2.2. Анализ чатов на лексическом уровне.....	50
2.4. Анализ чатов на синтаксическом уровне.....	54
ВЫВОДЫ ПО ГЛАВЕ 2	69
ЗАКЛЮЧЕНИЕ	70
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	71
ПРИЛОЖЕНИЯ	78

ВВЕДЕНИЕ

На сегодняшний день интернет играет важную роль в жизни современного человека. Компьютерная техника, смартфоны, глобальная сеть предлагают невероятные возможности получать, хранить информацию, обмениваться ей, общаться в удобных условиях. Согласно работам многих современных лингвистов, например, А.В. Палковой, О.В. Лутовиновой, Е.И. Горошко, А.Е. Войскунского, Т.Ж. Бальжировой, современное общество является миром коммуникации, в котором постоянно появляющиеся технические средства обуславливают успешность обмена информации. Развитие цифровых и компьютерных технологий послужило открытию нового типа взаимодействия – «человек – компьютер», а появление и распространение интернета расширило эту схему, сделав компьютер посредником коммуникации. Доступность интернета благоприятствует развитию видов общения и появлению новых типов общественных отношений.

Одной из наиболее востребованных областей интернета в 2016 году является написание и настройка чатботов – автоматизированных программ, имитирующих общение с реально существующим человеком. Данные программы активно используются в различных областях. К примеру, многие онлайн-магазины (М-Видео, РТБ), службы доставки еды (Burger King) размещают на своих сайтах чатботы для ускорения приема заказов, и подобные онлайн-консультанты привлекают клиентов оперативностью работы. Даже компания Facebook готовит платформу для размещения коммерческих чатботов, способных забронировать для пользователя столик или оформить покупку товара таким образом, как будто клиент был обслужен первоклассным менеджером. Однако чатботы также могут представлять собой угрозу, так как, притворяясь человеком, программа может просить у собеседника важные пароли или перевод денежных средств. Некоторые сервисы для онлайн-знакомств (Мамба, Тутла)

используют чатботы для привлечения новых пользователей и вымогания оформить платную подписку на некоторые услуги. К примеру, для продолжения разговора с «человеком», подходящим по интересам, возрасту, чувству юмора, для получения личной информации о нем пользователю предлагается оформить премиум-доступ. В связи с этим поднимается проблема изучения коммуникации с чатботами.

Данная выпускная квалификационная работа посвящена маркерам лингвистической искусственности интернет-коммуникации. **Цель** исследования – выявление основных маркеров лингвистической искусственности коммуникации с программой чатбот. Поставленная цель потребовала решения следующих задач:

- проанализировать теоретические основы изучения интернет-коммуникации и выявить ее основные характеристики;
- описать особенности коммуникации с чатботом;
- определить чатбот тестом Тьюринга;
- провести анализ чатбота на лексическом, синтаксическом и текстовом уровне и установить признаки искусственности коммуникации.

Объектом данного исследования выступает общение с программным коммуникантом (чатботом) в сети интернет. **Предметом** исследования являются маркеры лингвистической искусственности коммуникации с чатботом. **Материалом** исследования послужил архив диалогов чатбота Cleverbot с реальными людьми.

Актуальность и новизна данной работы определены слабой изученностью интернет-коммуникации, низким уровнем проработанности данной темы, несмотря на то, что интернет все прочнее входит в нашу жизнь не только как средство работы или учебы, но и как средство общения. Особенности взаимодействия с чатботами практически не изучаются российскими учеными,

степень изучения данного вопроса зарубежными исследователями также невысока.

Методологическую и теоретическую базу настоящего исследования составляют труды отечественных и зарубежных ученых-исследователей, посвященные: 1) интернет-коммуникации (Н.А. Ахренова, Е.Н. Галичкина, О.В. Дедова, Е.А. Вишнякова, В.Г. Костомаров, Л.Г. Ковальская, Т.Н. Колокольцева, С.В. Титова, Л.Ю. Щипицина и др.), 2) особенностям чатов (работы Н.С. Андриановой, В.Н. Базылева, А.Б. Бушева, Е.Н. Вавилова, Е.И. Горошко, Л.Ю. Иванова, Л.А. Капанадзе, Л.Ф. Компанцевой, Н.А. Лапшеевой, М.Л. Макарова, С.Н. Михайлова, Е.Ю. Распопниной, Н.Б. Рогачевой, А.А. Селютина, Н.Г. Трофимовой, О.Ю. Усачевой, Л.Ю. Щипициной и др.), 3) технологиям виртуальных собеседников (Н.Н. Зильберман, А.И. Провотар, К.А. Клочко и др.).

В качестве основных методов исследования в работе были использованы: анализ научной отечественной и зарубежной литературы, монографий и статей, обзорный, описательный и сравнительный методы, контекстуальный анализ, а также аналитический анализ и метод лингвистического описания.

Данная работа состоит из введения, двух глав, заключения, библиографического списка и приложений. В первой главе рассмотрены основные подходы к изучению интернет-коммуникации, определены основные характеристики интернет-коммуникации, выявлены особенности коммуникации с чатботом, в том числе методы порождения текстов в системе чатбота. Во второй главе предоставлен анализ эффективности теста Тьюринга для определения уровня искусственности чатбота, анализ программы чатбота на лексическом и синтаксическом уровнях, а также на уровне текста. В заключении представлены основные выводы и результаты теоретической и практической части исследования, отмечена практическая значимость, а также намечены перспективы для дальнейшего изучения.

ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИЗУЧЕНИЯ ИНТЕРНЕТ-КОММУНИКАЦИИ

1.1. Теоретические подходы к изучению интернет-коммуникации

В двадцать первом веке интернет является неотъемлемой частью жизни современного человека. Глобальная сеть вошла в повседневность, навсегда изменив ее: она сделала возможным получение доступа к невероятному количеству информации, даже если мы находимся на работе, учебе, дома или даже на улице. Интернет ускорил процесс приема и передачи данных, тем самым увеличив эффективность человеческой деятельности. В добавление ко многим другим положительным изменениям в нашей жизни интернет породил совершенно новые виды межличностного общения. Сеть сделала нашу жизнь намного удобнее.

Исходя из термина, представленного в документе «Internet Freedom and Broadcast Deployment Act of 2001», который был принят на второй сессии 107-го заседания Конгресса Соединенных Штатов Америки, интернет является совокупностью «...большого числа связанных компьютеров и средств телекоммуникации, включающих в себя оборудование и программное обеспечение, которые составляют всемирную Сеть, использующую протоколы TCP/IP (или любые предшествующие или последующие версии этих протоколов) для передачи информации посредством проводной или беспроводной связи» [Internet Freedom and Broadcast Deployment Act of 2001, 2002]. Иными словами, интернет первоочередно является сетью телекоммуникационных и компьютерных средств, в то время как в гуманитарных трудах интернет описывается через его функции. Например, в работе «Общая теория социальной коммуникации» А.В. Соколова представлены следующие определения интернета относительно его функций:

- глобальный канал коммуникации, поддерживающий передачу мультимедийных сообщений вне зависимости от расстояния (учитывая коммуникационно-пространственную функцию);
- место хранения информации, доступное всем пользователям (коммуникационно-временная функция);
- служебный механизм самовыражения и социализации личности и социальной группы (коммуникационно-социализирующая функция) [Соколов, 2002].

Мы абсолютно согласны с данными определениями интернета относительно его функций. С нашей точки зрения, его коммуникационно-пространственная функция является самой значимой с точки зрения лингвистических исследований и оттого ценной для данной работы. Интернет объединяет своих пользователей по всему миру, позволяет общаться без временных задержек и получать максимальное количество информации, так как при данном виде коммуникации задействована передача аудио- и видеоматериалов. Исходя из этого, интернет значительно влияет на общество, на их уклад жизни и характер их коммуникации. Его социальную функцию определяют следующие особенности:

- способность замены естественного межличностного общения тет-а-тет его искусственной формой, что ведет к изменению повседневного взаимодействия индивидов и социальных групп;
- появляется особый вид социальных образований – интернет-сообщества, значительно разнящиеся по тематике: сообщества поэтов, автолюбителей, закрытые «пиратские» торрент-трекеры, форумы религиозных движений, научно-исследовательские интернет-группы и многие другие [Поправко, 2011];
- передача ценностей, знаний, норм, установок и так далее;

- увеличение возможностей принимать, обрабатывать, усваивать и передавать информацию на высокой скорости [Горошко, 2012].

Интернет формируется и реализуется самостоятельно, он обладает бесконечными и быстроспроизводимыми ресурсами памяти, которые по своей системе сходны к человеческому темпу мышления: полиэкранность и нелинейность освоения информации, а также возможность прямого, творческого, свободного и интерактивного участия в массовой коммуникации. Осуществление данных действий невозможно для человека без языковой идентификации. Значит, интернет развивается как особая область функционирования языка и речи [Вишнякова, 2011].

Определение функций, особенностей и возможностей глобальной сети привели к появлению большого интереса к исследованию интернет-коммуникации, данной теме посвятили свои работы Н.А. Ахренова, Е.Н. Галичкина, О.В. Дедова, Е.А. Вишнякова, В.Г. Костомаров, Л.Г. Ковальская, Т.Н. Колокольцева, С.В. Титова, Л.Ю. Щипицина и многие другие.

По нашему мнению, стоит также обратить внимание на написание термина «интернет». Некоторые исследователи, например, Е.А. Вишнякова, считают, что Интернет является именем собственным, и потому его нужно писать с заглавной буквы. В «Русском орфографическом словаре Российской академии наук» представлено такое же написание: Интернет [Русский орфографический словарь Российской академии наук, 2007]. Другие ученые придерживаются того мнения, что в современном русском языке интернет перестал обозначать имя собственное и уже перешел в число нарицательных существительных. В «Толковом словаре иноязычных слов» присутствует следующее пояснение: «ИНТЕРНÉТ [тэ́, нэ́], а, м., с прописной буквы [англ. Internet <inter(national) международный + net сеть][Крысин, 2008]. Тем не менее, обе группы ученых считают, что в составных словах, например, интернет-сообщество, интернет-друг, интернет-библиотека, первая часть выступает в роли прилагательного и потому пишется со строчной

буквы. Русский орфографический словарь даёт следующее толкование: сегодня возможно оба вида написания, и с прописной, и со строчной буквы. Например: доступ к Интернету и доступ к интернету, пользоваться Интернетом и пользоваться интернетом. [Русский орфографический словарь, 2015]. Мы считаем данный термин нарицательным существительным и потому пишем его со строчной буквы и склоняем по правилам русского языка.

В современной лингвистике термин «коммуникация» имеет большое количество определений, так как исследователи обращают внимание на различные характеристики коммуникации в процессе ее описания. В широком смысле под коммуникацией подразумевается обмен информацией между индивидами, пунктами или доменами посредством общей системы символов. Механистический подход описывает этот термин как процесс кодирования и передачи информации от источника и прием информации получателем сообщения. Деятельностный подход представляет коммуникацию как совместную деятельность участников коммуникации (коммуникантов), в ходе которой вырабатывается общий взгляд на вещи и действия с ними [Glossary Commander, 2016].

О.С. Ахманова в «Словаре лингвистических терминов» утверждает, что коммуникацией является процесс передачи некоторого мысленного содержания при помощи языка [Ахманова, 1966]. Однако мы считаем это определение несколько узким, так как коммуникация описывается через призму языка как единственного ресурса мысленного содержания. По нашему мнению, другие системы, например, жесты, изображения, эмодзи и эмодзи также передают мысленное содержание и обязательно должны учитываться. Тем не менее, данное определение не ограничивает участников коммуникации, и можно предположить, что ими могут являться не только реально существующие люди, но и программы, имитирующие их, то есть чатботы.

С.Г. Тер-Минасова представляет в книге «Язык и межкультурная коммуникация» разные определения данного феномена из разных источников.

- Коммуникация – акт общения, связь между двумя или более индивидами, основанная на взаимопонимании; сообщение информации одним лицом другому или ряду лиц [Словарь иностранных слов, 1979].
- Коммуникация – сообщение, общение [С.И. Ожегов, 1972].
- Communication. Act of imparting (esp. news); information given; intercourse. Коммуникация. Акт обмена (особенно новостями); данная информация; общение (перевод С.Г. Тер-Минасовой). [Oxford Dictionary, 1964], [Тер-Минасова, 2010].

В рамках социолингвистического направления В.П. Конечкой представлено следующее определение коммуникации: «...социально обусловленный процесс передачи и восприятия информации, как в межличностном, так и в массовом общении по разным каналам при помощи различных вербальных и невербальных средств» [Грушевицкая, Попков, Садохин, 2002]. В «Психологической энциклопедии» межличностная коммуникация представлена несколько иначе, как «процесс передачи информации неким источником другому конкретному целевому объекту или членам идентифицируемой группы. Эти коммуникации обычно осуществляются в ходе личных взаимодействий, однако они могут также реализовываться с использованием почты, телефона или других электронных средств» [Психологическая энциклопедия, 2016], [Морозова, Базылева, 2011]. Стоит обратить внимание на то, что в данных определениях нет указания того, что получателем или отправителем сообщения непременно должен быть человек. Более того, учтены электронные каналы коммуникации, являющиеся неотъемлемой частью коммуникации с чатботом.

М.Л. Макаров, рассматривая коммуникацию с позиции конструктивизма, говорит о том, что она представляет собой «частный случай интеракции,

опирающейся на интерпретацию, предполагающей координацию действий во взаимодействии, а также удовлетворяющей потребности выражения внутренних состояний» [Макаров, 1997]. Будучи созданным изначально для оказания собеседникам психологической помощи, чатбот функционально отвечает требованиям данного определения: взаимодействуя с человеком, он удовлетворяет его потребность в выражении желаний, целей, настроения.

С точки зрения А.А. Брудного, «коммуникация – одна из форм всеобщей связи между явлениями, а именно направленная связь, которая выражается в передаче сигналов (последнее не обязательно осознанный, осмысленный процесс)» [Брудный, 1977; Прокопец, 2015]. Данное определение фокусируется на коммуникации как на не всегда осознаваемом процессе, что относится к коммуникации с чатботом со стороны программы.

Исходя из данных определений, можно сделать вывод, что основной характеристикой коммуникации является передача и прием данных (обмен ими). В определении из «Психологической энциклопедии» говорится, что коммуникация может осуществляться с помощью электронных каналов связи, а М.Л. Макаров опирается в своем определении на выражение внутренних состояний партнеров по коммуникации как на важную особенность данного процесса. А.А. Брудный выделяет осознанность коммуникации как необязательную черту. Следовательно, общение с чатботом может по праву считаться коммуникацией: оно опосредованно техническими и электронными средствами связи, существует для передачи своего настроения человеком, а также является осознанным лишь со стороны человека. Более того, чатбот основан на передаче и приеме данных, что является общей характеристикой коммуникации, подчеркнутой в большей части определений.

В современных лингвистических исследованиях авторы придерживаются разных терминов, которые обозначают межличностное отношение с помощью глобальной сети. В научных работах встречаются следующие термины:

виртуальный дискурс (О.В. Лутовинова), компьютерно-опосредованная коммуникация (Л.Ю. Щипицина, И.Н. Розина), электронное общение/коммуникация (Е.Н. Галичкина), интернет-коммуникация (Т.Н. Колокольцева) и так далее. Л.Ю.Щипицина в книге «Компьютерно-опосредованная коммуникация: Лингвистический аспект анализа» исследует частотность употребления данных терминов.

Таблица 1. Частотность употребления терминов

Термин	Количество ссылок, 24.07.2008	Количество ссылок, 15.12.2009	Научные работы с использованием термина
Базовое слово «коммуникация»			
Электронная коммуникация	1890	4600	[Смирнов, 2004; Горошко, 2005; Рязанцева, 2010]
Интернет-коммуникация	1520	6530	[Горошко, 2006]
Компьютерно-опосредованная коммуникация	1090	38900	[Розина, 2002, 2005; Раскладкина, 2005; Волкова, 2007]
Виртуальная коммуникация	1040	2410	[Травин, 2001; Бергельсон, 2002; Потапова, 2004]
Компьютерная коммуникация	774	3730	[Бакулев, 2005]
Сетевая коммуникация	531	56600	[Жичкина, Белинская, 1999; Обухов, 2008]
Базовое слово «дискурс»			
Компьютерный дискурс	207	1470	[Леонтович, 2000; Галичкина, 2001; Кондратов, 2004]
Сетевой дискурс	2011	790	[Моргун, 2002]
Виртуальный дискурс	170	1240	[Лутовинова, 2007; Компанцева, 2006, 2007]
Интернет-дискурс	128	515	[Компанцева, 2004; Громова, 2007; Егорова, 2008]
Электронный дискурс	97	498	[Аврамова, 2004; Соколинская, 2004]
Компьютерно-медийный дискурс	4	8	[Ковальская, 2003; Коретникова, 2006]

Данная статистика показывает, что количество обозначений такого типа коммуникации, при котором задействован интернет, со временем возрастает, что подтверждает увеличение ее значимости в жизни современного человека, корпусов текстов в сети по этой теме и росту интереса к исследованиям в этом направлении [Щипицина, 2010]. Новизну исследований в области интернет-коммуникации также подтверждает отсутствие устоявшегося терминологического аппарата.

При изучении данного феномена во время нашей работы наиболее часто встречался термин «интернет-коммуникация», и для более точного определения частотности в русскоязычных запросах терминов, обозначающих коммуникацию в глобальной сети, на сегодняшний день мы провели статистический анализ с помощью сервиса Яндекс.Статистика. Данная таблица демонстрирует частотность употребления терминов в запросах пользователей за месяц [Яндекс.Статистика].

Таблица 2. Частотность употребления терминов относительно коммуникации в интернете на сегодняшний день

Термин	Количество запросов на момент 04.06.2016
Базовое слово «коммуникация»	
Электронная коммуникация	5094
Интернет-коммуникация	9575
Компьютерно-опосредованная коммуникация	134
Виртуальная коммуникация	1560
Сетевая коммуникация	1707
Базовое слово «дискурс»	
Компьютерный дискурс	353
Сетевой дискурс	119
Виртуальный дискурс	165
Интернет-дискурс	869
Электронный дискурс	220
Компьютерно-медийный дискурс	2

Данная статистика подтверждает то, что термин «интернет-коммуникация», относящийся к новому виду общения, порожденного глобальной сетью, употребляется чаще других. Именно этот термин мы и будем

использовать в данной работе. Значительное количество запросов по данной тематике показывает ее актуальность и значимость в лингвистических исследованиях.

Тем не менее, несмотря на очевидную популярность термина, «интернет-коммуникация» для обозначения коммуникации в сети некоторые ученые, например, О.Н. Морозова, считают его сомнительным, так как одна часть термина – интернет – представляет собой среду, в то время как другая – коммуникация – является сутью этой среды. Однако использование данного определения оправдывается тем, что ещё недавно главными функциями интернета были получение и хранение информации, а теперь их вытесняет роль канала коммуникации, что стимулирует новые социокультурные процессы. Глобальная сеть становится средой общения, которая играет важную роль во всех сферах деятельности современного человека. Функция канала коммуникации в совокупности с растущей доступностью к интернету способствует появлению новых перспективных видов коммуникации, которые приводят к возникновению инновационных видов отношений внутри общества [Морозова, 2010].

Е.А. Землякова, чье мнение разделяет Е.И. Горошко, представляет классификацию, обозначающих феномен коммуникации в интернете. По ее мнению, термин «электронная коммуникация» является родовым понятием, так как обозначает коммуникативное взаимодействие с помощью любого электронного канала. Данное понятие включает в себя «компьютерную коммуникацию», под которой понимаем любое коммуникативное взаимодействие, осуществляемое с помощью компьютера. И уже в этот термин входит «интернет-коммуникация», то есть вербальное и невербальное взаимодействие коммуникантов в интернете [Землякова, 2010], [Горошко, 2012]. По нашему мнению, эта классификация наиболее точно представляет структуру типов коммуникации.

Определение Л.Ю. Щипициной основано на существующих определениях коммуникации и компьютерно-опосредованной коммуникации: интернет-коммуникация является одним из «исторических этапов в развитии коммуникационных средств и существует наряду с неопосредованным (устным) общением, общением, опосредованным бумажными носителями (рукописным и печатным, включающим непериодические издания (книга) и периодические издания (газеты, журналы)), а также общением, опосредованным электронными приборами (факс, радио, телевидение)» [Щипицина, 2010]. Данное определение не фокусируется на возможных типах коммуникантов, оно лишь указывает на канал связи, благодаря чему общение с чатботом может по праву считаться одним из видов интернет-коммуникации, так как оно происходит с помощью компьютера и интернета.

С.В. Бондаренко считает интернет-коммуникацию, или виртуальное общение, общение двух и более лиц с помощью компьютера, при котором невидимые коммуниканты письменно посылают сообщения и имеют возможность мгновенного получения ответа и права доступа к информации, которая хранится в компьютере коммуниканта [Бондаренко, 2004]. Исходя из данного определения, общение с чатботом также является одним из видов интернет-коммуникации, так как оно происходит с помощью компьютера, коммуниканты не видят друг друга, но при этом имеют мгновенно отправлять и получать информацию. Мы считаем, что это определение термина интернет-коммуникация наиболее точно отражает суть данного феномена, и в данной работе придерживаемся именно данной трактовки этого понятия.

С.В. Бондаренко также представляет классификацию типов интернет-коммуникации. По ее мнению, в зависимости от типа субъекта, различают:

- общение реального субъекта с реальным партнером (электронная почта, группы новостей, ICQ и другие интернет-пейджеры, SMS-сообщения, дискуссионные форумы, электронные рассылки и т.д.);

- общение реального субъекта с субъективированным объектом как иллюзорным партнером (взаимодействия с почтовыми роботами, компьютерные игры, веб-сайты и т.д.);
- общение воображаемых партнеров (взаимодействие компьютерных интеллектуальных агентов, взаимодействия с DNS-серверами и т.д.).

Данная классификация подтверждает, что чатботы являются одним из видов интернет-коммуникации, так как относятся к одному из выделенных ее видов – общение реального субъекта с субъективированным объектом как иллюзорным партнером.

1.2. Основные характеристики интернет-коммуникации

Исходя из работы «Лингвистические характеристики интернет-чатов как вида коммуникации» Е.А. Подгорной и К.А. Демиденко, чат является одной из форм межличностного общения в интернете с наиболее высокой степенью интерактивности, для которой характерен компьютерно-опосредованный синтетический тип устно-письменной коммуникации с особенностями электронного канала передачи сообщений [Демиденко, Подгорная, 2014].

Многие лингвисты занимаются сегодня изучением чат-коммуникации как интернет-жанров, их характеристики и особенности раскрываются в работах таких ученых как Л.М. Гриценко, Д.А. Данилец, М.В. Кузьмина, Н.А. Лепшеева, К.В. Овчарова и Е.М. Чухарев. Наиболее часто чаты рассматриваются как самостоятельный электронный жанр, например, в работах Н.С. Андриановой, В.Н. Базылева, А.Б. Бушева, Е.Н. Вавилова, Е.И. Горошко, Л.Ю. Иванова, Л.А. Капанадзе, Л.Ф. Компанцевой, Н.А. Лапшеевой, М.Л. Макарова, С.Н. Михайлова, Е.Ю. Распопниной, Н.Б. Рогачевой, А.А. Селютина, Н.Г. Трофимовой, О.Ю. Усачевой и Л.Ю. Щипициной.

Коммуникация в интернете также может представлять собой совершенно новую форму общения, а именно письменную разговорную речь. Большая часть информации в интернете представлена в письменном виде, но темп и тип речи приближен к устной коммуникации. Интернет-коммуникация является живым человеческим мышлением, выраженным в виде не только текста, но и цвета, звука, графики и анимации, так как она не может существовать без активной человеческой деятельности по ее созданию и развитию [Шилина, 2012].

С одной стороны, для пользователей чата наиболее важной и распространенной проблемой является недостаток средств, в том числе паралингвистических, для выражения эмоций, например, тембр голоса, темп речи, дикция, жесты, мимика, акцентирование части высказывания. С другой стороны, собеседники представлены друг другу только через текст, и порожденное высказывание и его автор являются тождественными, ведь ничего кроме текста не существует. В тет-а-тет беседе всегда участвует сам человек, тело, которое, соединяясь с высказыванием, рождает образ личности. Данная нехватка эмоциональных средств частично компенсируется за счет добавления к тексту типизированных эмоциональных реакций, выраженных с помощью знаков препинания – эмотиконов, с помощью типизированных картинок - эмодзи и детально проработанных изображений лиц и животных – стикеров [The Guardian, 2015]. Однако несмотря на присутствие данных эмоциональных средств, ощущается неполная передача настроения собеседника, за определенным эмотиконом не закреплено только одно значение, и он может только направлять коммуникантов, добавляя степень их эмоций, показывать, что текст эмоционально окрашен [Виноградова, 2004].

В добавление к эмотиконам, эмодзи и стикерам в интернет-чатах также используется «капс» – написание текста с помощью зажатой клавиши Caps Lock, позволяющей писать только заглавными буквами, что при виртуальном общении расценивается как повышение голоса. Повышенная эмоциональность также

передается с помощью множества восклицательных знаков [Бахтин, 1986]. Точки в конце высказывания ставятся крайне редко – для быстроты письма, по той же причине начинаются предложения чаще всего со строчной буквы. Поставленную точку в конце высказывания собеседники часто трактуют как переход к серьезной теме диалога, сухость общения, нежелание проявлять эмоции.

К другим особенностям чат-коммуникации можно отнести сленг, который, возникая в глобальной сети, распространяется и на общеупотребительную лексику. В английском языке появляются новые сокращения, которые быстро входят в обиход и редко подлежат трансформации. Русскоязычные же пользователи чатов используют два способа порождения сленга:

- обозначают новые реалии лексическими средствами родного языка: кирпич (процессор), железо (технические компьютерные средства), мыло (электронная почта). Одним из наиболее часто употребляемых способов обогащения словарного интернет-запаса является аффиксация, с помощью которой образованы следующие слова: отладчик, загрузка, оцифровка и так далее. Подобная аффиксация в большой мере находится под влиянием компьютерного жаргона. Как считает И.А. Стернин, компьютерный жаргон состоит из обозначений действий, выполняемых пользователем, деталей оборудования, названий его программного обеспечения, сбоев работы системы и так далее;
- трансформируют или заимствуют англоязычные термины, иногда снижая коннотацию: твитнуть (отправить сообщение в Twitter), реплой (от reply, ответ на сообщение в почте), усер (пользователь, не очень хорошо разбирающийся в компьютерах) сервер, хакер, гама (компьютерная игра), вишасы (добрые пожелания), хрюша (Windows XP);
- переводят англоязычный сленг или используют для его обозначения русскую раскладку клавиатуры: З.Ы. (P.S.), СМСМ, имхо (согласно моему скромному мнению – ИМНО, in my humble opinion) [Стернин, 2000].

Вообще по способу словообразования технический (компьютерный) сленг и терминологию можно условно разделить на следующие группы: сокращения, синонимия, перевод, калькирование и фонетическая мимикрия. Возможно, некоторые лексемы рождаются благодаря чувству юмора пользователей, желанию сделать работу с компьютером проще, а также языковой игре, появляющейся из компьютерной терминологии.

Особое место в чатах занимает ономастическая лексика, так как все пользователи имеют свои собственные имена, настоящие или вымышленные – никнеймы (ники, псевдонимы). Чаще всего это имена яркие, экспрессивные, запоминающиеся: Wan2Sree, Ha10 (в многопользовательской онлайн-игре Halo), Ягодка (Одноклассники), Виктория Чужая (ВКонтакте). С точки зрения морфологии, язык интернет-коммуникации отличается постоянным использованием императива первого лица единственного числа вместо формы вежливости во множественном числе: «Хочешь красивое тело к лету? Жми сюда!», «Я жду тебя, напиши мне!» (ВКонтакте). Подобные рекламные объявления часто располагаются вокруг поля чата, привлекая внимание собеседников.

Н.А. Кубракова представляет в своей работе «Игровая составляющая интернет-коммуникации (на примере чат-коммуникации)» особенности данного вида общения, сближающие его с игрой. По ее мнению, анонимность и «бестелесность» собеседников, чувство безопасности за жизнь раскрепощают их и позволяют представить себя стилистически таким образом, каким это невозможно сделать с реальной жизни [Кубракова, 2015]. Е.И. Шейгал и Ю.М. Иванова также считают это свойством игры, возможностью не отвечать за свои поступки, что помогает вывести другую характеристику коммуникации как игры – безопасность. В связи с этим коммуниканты часто прибегают к приему языковой игры [Иванова, Шейгал, 2008]. Ещё один игровой аспект чат-коммуникации – это решение двух основных задач, которые решаются

собеседниками в интернете: поиск информации и общение ради общения. Некоторые ученые добавляют к этим задачам еще желание развлечения [Трофимова, 2004; Лутовинова, 2009]. Сходство чат-коммуникации с игрой также заключается в тяготении к сниженной лексике на сайтах с серьезным информационным, например, политическим, контентом.

Синтаксис общения в чате необходимо рассматривать с двух позиций. На локальном уровне, то есть, не учитывая гипертекстовые связи, русскоязычное и англоязычное интернет-общение характеризуются эллиптическими конструкциями, перестановкой частей высказывания, парцелляция, амплификация, а также обилием экспрессивных элементов. Имеется тенденция к аграмматизму, то есть отклонению от синтаксических и пунктуационных норм языка, некоторые части высказывания могут не согласовываться друг с другом. Исследуя гипертекстовые связи, лингвисты заходят в тупик, так как их систематический анализ всех частей текста ссылок невозможен. Максимальное исследование в этой области остановилось на создании навигации по сайтам, представляющим древо расположения информации [Wilde, 2009].

Особенности интернет-коммуникации, обусловленные социальными и прагматическими факторами, исследуются с позиции того, каким образом они влияют на речевое поведение коммуникантов. И.А. Смольяникова выделяет в своей работе следующие коммуникативные переменные интернет-коммуникации: культурный контекст, средства вербальной и невербальной коммуникации, стратегии достижения прагматических целей. Культурный контекст представляет в данном случае широкое понятие, которое предполагает:

- знание собеседниками фактического материала, то есть структуры панели управления, инструментов, задач, программного обеспечения, структуры интернет-сайтов;
- знание норм и традиций (например, веб-этикета);

- владение определенным набором социальных и ситуативных ролей (например, редактора, адресата, веб-мастера);
- умение правильно выбрать вербальное поведение в соответствии с исполняемой ролью;
- знание паралингвистического поведения (например, правильный подбор цветов или соотношение текстовой информации и графических изображений);
- знание деталей повседневной жизни и специфики отношения представителей социума при работе в интернете [Макарова, 2009].

К средствам вербальной и невербальной интернет-коммуникации

И.А. Смольяникова относит:

- аббревиатуры и сокращения (m/f – male или female, bc – because, wb – welcome back);
- символы, эмодзи, эмодзи;
- определенные грамматические конструкции и др.

Автор также приводит в своей работе некоторые стратегии достижения прагматических целей. Одной из них является изменение собственной идентичности, к которому прибегают, по её мнению, многие пользователи интернета. Эта возможность обеспечивается анонимностью коммуникации в интернет-среде, которая достигается за счет преимущественного общения на основе письменного обмена сообщениями. Изменение собственной идентичности имеет несколько аспектов: изменение собственного возраста, социальной принадлежности, национальности и так далее [Смольяникова, 2003]. Феномен конструирования собственной идентичности широко изучается в области интернет-коммуникации, в частности проблема гендерного варьирования в интернет-коммуникации. Исходя из того, что интернет-среда предоставляет коммуникантам свободу в виде анонимности, они меньше зависят от стереотипных представлений при достижении прагматических целей. Тем не

менее, согласно работам исследователя С. Херинга, мужчины и женщины значительно зависят от соответствующих гендерных речевых моделей. Например, мужчины представляют себя экспертами в любой области, отвечают на вопросы, дают советы в трудных ситуациях, а женщины создают дружественную обстановку, одобряют и благодарят собеседника. Зависимость человека от знания гендерных ролей настолько велика, что при получении сообщения, содержащего утешение и отправленного от лица мужчины, коммуникант может усомниться в достоверности пола автора [Макарова, 2009].

Переходя к дискурсивным особенностям интернет-коммуникации, необходимо отметить, что за определение дискурса мы берем точку зрения Н.Д. Арутюновой: дискурс – это связный текст в совокупности с экстралингвистическими факторами [Арутюнова, 1990]. По словам автора определения, «одной своей стороной дискурс обращен к прагматической ситуации, которая привлекается для определения связности дискурса, его коммуникативной адекватности, для выяснения его импликаций и пресуппозиций, для его интерпретации. Другой своей стороной дискурс обращен к ментальным процессам участников коммуникации: этнографическим, психологическим и социокультурным правилам и стратегиям порождения и понимания речи в тех или других условиях (discourse processing), определяющих необходимый темп речи, степень ее связности, соотношение общего и конкретного, нового и известного, субъективного и общепринятого, эксплицитного и имплицитного содержания дискурса, меру его спонтанности, выбор средств для достижения цели, фиксацию точки зрения говорящего и т.п.» [Там же]. В парадигме теории дискурса интернет-дискурс (интернет-коммуникация) обладает следующими характеристиками: динамичность, коммуникативность, виртуальность, дистантность, опосредованность, креолизованность (разбитие текста на разнородные части), персонифицированность. По каналу коммуникации интернет-коммуникация –

это дискурс электронного общения, процесс которого происходит посредством интернета [Меновщиков, 2007; Сидорова, 2014].

Подтверждая вышесказанное, приведем в пример коммуникацию с чатботом, которая обладает этими же характеристиками. Человек общается с чатботом с помощью компьютера, следовательно, коммуникация опосредованна. Данный вид коммуникации существует в интернете, поэтому он виртуален. Коммуниканты не видят друг друга, им представлен лишь текст, произведенный собеседником, следовательно, коммуникация с чатботом дистантна. В силу мгновенного построения фраз программой данное общение обладает динамичностью. Реплики чатбота разбиты на разнородные части, зачастую не связанные друг с другом, что свидетельствует от креолизованности общения. При общении чатбот меняет собственную идентичность, что говорит о персонифицированности коммуникации с чатботом.

1.3. Особенности коммуникации с чатботом

Создание действующего виртуального собеседника и эффективное его использование исследуется на протяжении многих лет. Сегодня тема общения в сети Интернет особенно актуальна, так как она открывает возможность к мгновенному получению информации и обмену ей, работе с несколькими пользователями одновременно, организации учебной деятельности, бизнес-общению, проведению опросов и всевозможных исследований, а также обладает многими другими положительными чертами. Главными вопросами в конструировании системы-виртуального собеседника являются создание типичной модели коммуникации, модели коммуниканта и прагматических и семантических средств в условиях программы – ИИ [Провотар, Клочко, 2013].

Искусственный интеллект (ИИ) можно определить как область компьютерной науки, занимающуюся автоматизацией разумного

поведения. Данная технология является идеей конструирования программ со способностью к разумному мышлению, которое можно соотнести с деятельностью человеческого мозга. По словам Р.Бенерджи, искусственный интеллект – это «мир мыслей», заключенных в техническую оболочку [Бенерджи, 1972]. Ещё в середине двадцатого века ученые выяснили, что машина может мыслить подобно человеку. В словаре по кибернетике И.И.Глушкова есть иная трактовка этого определения: «Искусственный интеллект – это искусственная система, имитирующая решение человеком сложных задач в процессе деятельности» [Глушков, 1979]. Если ИИ всегда сопоставляется с разумом человека, то что же такое естественный интеллект? По И.Ю.Алексеевой, это человеческое умение рассуждать логически, стремиться к получению и развитию знаний и умений, способность анализировать делать выводы, а также проявлять творческий подход к общению или вещам [Алексеева, 1991]. Искусственный интеллект – это программное обеспечение, помещенное в аппаратное средства, в результате чего почти любая поставленная проблема решается тем же образом, что и у человека. Благодаря этому качеству ИИ легко спутать с поведением человека [Исламов, Фомин, 2013].

Бот (от слова робот) является специальной компьютерной программой, выполняющей определенные действия согласно заданного сценария и может быть полностью автоматической или подчиняться каким-либо условиям, например, периоду функционирования или геолокации пользователя. Программа-бот требует запуска и начальных настроек: входа на сайт, подписки на рассылку, ответа на сообщение от виртуального собеседника. Характерными признаками таких программ является их повторяемость и простота модели. Бот является совокупностью нескольких примитивных команд, лишь прообразом искусственного интеллекта. С развитием интернета сложность ботов постоянно возрастает, и сегодня их легко спутать с реальным собеседником, например, при общении в чатах.

Перейдём к описанию основных видов ботов.

- чатботы, или виртуальные собеседники. Программы, способные отвечать на определенные слова или фразы. Разнятся в зависимости от количества возможных вариантов ответа, которые хранятся в базе знаний. Самые лучшие из таких ботов почти неотличимы от живого человека. Распространены в чатах и социальных сетях;
- задачей спам-ботов является осуществление автоматической рассылки нежелательных сообщений (спама) на электронную почту, сайты, форумы, доски объявлений. Не отличаются особой сложностью;
- игровые боты делятся по своей сути на два типа:
 - 1) созданный самой игрой для помощи, например, для относительного равенства в командах, сражающихся в турнире, если у одной из сторон численный перевес; или же просто нпс (неигровой персонаж) играющий в вашей команде — тоже бот;
 - 2) специальные программы, официально запрещенные, позволяющие выполнять особые задачи в игре без участия живого игрока. Их часто используют при необходимости совершать много монотонных и одинаковых действий, получая таким способом преимущество. Например, могут играть днем, а на ночь включать бота собирать определенные ресурсы, получать достижения или бить врагов. Именно к таким игрокам часто применяют санкции за «ботоводство», вплоть до удаления аккаунта (профиля).
- поисковые боты являются специальными программами поисковых систем, которые помогают им индексировать сайты, то есть настраивать точность поиска. К ним также относятся другие программы, автоматически сканирующие сайты для сбора какой-либо информации.

Чатбот является программой, созданной на основе искусственного интеллекта. Собеседник, изначально написанный для того, чтобы оказывать

психологическую помощь пациентам, в двадцать первом веке является незаменимым помощником многих пользователей в поиске нужной информации, управлении устройств или простом общении. Чатботы выполняют три главные роли: служебную, информационно-развлекательную и функцию утилит. Служебные боты следят за орфографией, регулируют права участников общения в Интернете, обеспечивают меры безопасности. Например, они могут исключить из чата собеседника, чей ник состоит из символов, напоминающих нецензурное выражение, а также тех, чьи реплики повторяются без изменений или с минимальными изменениями во избежание флуда, то есть нарушения сетевого этикета. Во многих пабликах (сообществах в социальной сети) «ВКонтакте» запрещено размещение сторонних ссылок. Бот автоматически считывает их и отправляет автора поста или комментария в черный список. К информационно-развлекательным ботам относятся имитирующие реальное общение (чаще всего строятся из случайных реплик собеседников), а также словари и энциклопедии, которые при запросе выдают один или несколько ответов, содержащихся в базе данных, как, например, Infobot, который работает по схеме «X – это Y». К ботам-утилитам относятся онлайн-переводчики и комментаторы, которые при упоминании кода перенаправляют автора на нужную ссылку.

Работа чатбота реализуется в несколько этапов: программа принимает команду, анализирует ее и отправляет информацию в ответ или выполняет порученное действие, то есть общение между человеком и программой происходит путем обмена сообщениями. Существует два типа коммуникации между человеком и чатботом: простая беседа и обсуждение важной проблемы. Однако большинство чатботов имеют упрощенную структуру: они распознают только ключевые слова в целой фразе собеседника. Например, если в предложении человека присутствовали названия фруктов, то программа может спросить, какие именно фрукты он любит. Также программа может реагировать на порядок слов и знаки препинания, благодаря чему может адекватно

реагировать на вопросы. Многие боты также обладают базой сопутствующих фраз: если человек произносит одну, то бот автоматически отправляет ему закрепленную за ней в базу. Одна из проблем, присущих всем виртуальным собеседникам, является неспособность правильно реагировать на предложения, в которых присутствуют два и более местоимения или синонима, так как программа не может установить заменимость существительного. При такой ситуации бот анализирует предпоследнее высказывание собеседника и выдает приемлемый ответ. Создание программы, обладающей гибким интеллектом, на сегодняшний день является важной проблемой, пока что никем успешно не решенной [Провотар, Клочко, 2013].

Основой любого чатбота служит система порождения естественного языка, разработанная в рамках лингвистического направления естественно-языковой генерации, или NLP, natural language processing. По мнению автора книги «Системы порождения естественного языка» Дэвида МакДональда, «процесс создания естественного языка является процессом сознательного конструирования текстов естественного языка для удовлетворения коммуникативных целей» [McDonald, 1988]. В двадцать первом веке эту область принято считать ответвлением направления обработки естественного языка, целью чего становится конструирование фраз, высказываний, текстов естественного языка из заложенных баз данных или логических форм, то есть из машинного представления [Kumar, 2011]. Данное лингвистическое направление широко изучается за рубежом с 1990 года, и системы конструирования фраз уже сегодня могут использоваться во многих областях. Например, подобные системы способны анализировать погодные карты и выдавать новый текстовый вариант прогноза погоды. Также их можно задействовать для перефразирования сложных медицинских терминов в быденную лексику для пациентов и во многих других случаях. Системы порождения фраз и текстов предстают перед пользователем в двух ролях:

- как автор, то есть они способны создавать тексты автономно, без помощи человека. Примером такой программы служит погодный анализатор или программа-составитель отчета по имеющимся статистическим данным;
- как помощник, являясь базой для создания более сложных документов. Например, программой-помощником служит генератор рифм, подбирающий нужное слово по синонимам, ударению или окончанию.

Все системы конструирования фраз делятся на 4 группы:

- автономно создающие тексты (программы, создающие отчеты по имеющимся данным);
- интерпретирующие данные для презентации в более понятном виде (программы перефразирования медицинских, технических терминов);
- обучающие программы (диалоговые, вопросно-ответные системы);
- развлекательные программы (программы для написания юмористических текстов и стихотворений).

Сложные системы порождения фраз состоят из нескольких частей, что необходимо для создания более «живого», естественного текста, приближенного к созданному человеком. Следующие 6 частей сложной системы являются стандартными:

- отделение контента, определяющий важность информации и необходимость ее употребления в тексте;
- отделение планирования, организовывающий имеющуюся информацию, выстраивающий высказывания, реплики в необходимом порядке;
- отделение агрегации фраз, ответственное за естественность и читаемость полученного текста. В этом отделении происходит членение текста на отрезки и слияние коротких предложений в одно по встроенным правилам;
- отделение выбора лексических средств, подбирающий необходимые для базы словари и сопоставляющий термины;

- отделение конструирования фраз, которое наполняет базы для предложений необходимыми лексическими средствами и меняет слова на их синонимы во избежание повторов. С его помощью текст также обретает естественность;
- отделение реализации текста, перерабатывающий текст к конечному виду по правилам орфографии, синтаксиса и морфологии [McDonald, 1988].

Программы-чатбот строятся на базе определенных методов генерации текстов. Самым распространенным и простым методом является «мешок слов», для использования которого в имеющемся тексте определяется частота каждой лексической единицы. В результате этого метода получается словарь с описанием частоты употребления слов. В конечном тексте любое слово из этого словаря появляется с вероятностью, равной отношению частоты слова к общему количеству слов в тексте. Иными словами, получаются грамматически и синтаксически связанный текст, который при этом не имеет никакого смысла. Данный метод использовался на рассвете создания чатботов и послужил отправной точкой изобретения других методов для улучшения их работы.

Следующий метод, называемый «цепями Маркова» позволяет конструировать более естественный текст, на практике выглядит следующим образом: любое последующее слово берется из ранее заготовленного текста-базы. Количество слов, которые возможно использовать из базы, определяется длиной цепочки, которую задает человек. Конечный текст получается «живым» благодаря учету локального контекста: две одинаковые лексические единица будут стоять рядом только в том случае, если они находились рядом в тексте-базе [Павлов, 2011].

Ещё один метод порождения фраз в системе чатбота основан на фрагментах текста, заготовленных в виде частей фраз или текстов в таблицах. Каждая часть колонки определяет идеальное место расположения фразы в тексте. При комбинировании этих фраз данные этих колонок порождают адекватный текст,

части которого хорошо согласуются друг с другом, так как текст-база был заготовлен человеком. Тем не менее, этот метод не является основным и даже успешным для работы чатбота, так как конечные тексты быстро теряют уникальность из-за ограниченного числа заготовленных частей.

Метод отзеркаливания является на сегодняшний день самым успешным и широко используемым в отношении чатботов. Отзеркаливание фраз иными словами – это перефразирование поступающих от собеседника фраз. Разумеется, чатбот не привносит в диалог новой информации, но он способен поддерживать естественный разговор. Приведем пример метода отзеркаливания:

Человек: «На улице идет дождь».

Чатбот 1: «Правда на улице идет дождь?»

Чатбот 2: «Почему ты думаешь, что на улице идет дождь?»

Чатбот 3: «Как ты узнал, что на улице идет дождь?»

Чатбот 4: «Ты рад, что на улице идет дождь?»

Также для поддержания разговора чатботы могут использовать слова из собственной заготовленной базы фраз. Такая база строится по контексту, в котором то или иное слово было употреблено [Atwell, Shavar, 2015].

Ещё один метод работы некоторых чатботов – метод шаблонов, при котором программа передает входные данные напрямую в лингвистическую поверхностную структуру. Шаблоны вставляются в структуру пробелов. Например, имеется следующий шаблон: «Поезд номер [номер] отправляется в [время]». Входные данные представляются конкретными значениями номера и времени [Krahmer, Theune, van Deemter, 2003]. Данный метод используется чатботами с функцией поиска в интернете, например, Siri.

Метод с использованием словарей заключается в том, что изначально создается словарь со словами из текста-базы, в котором указан порядок слов в предложении, а также их значение и грамматические характеристики. В

результате получается адекватный текст, в котором, в отличие от, например, метода «мешок слов», учитывается семантика.

В 1950 году британский математик и криптограф Алан Тьюринг предложил в своей статье «Вычислительные машины и разум» особый тест, который поможет определить адекватность работы чатбота, уровень ее продвинутости, способность мыслить и быть подобием человеческого разума. В данной статье предлагается два типа теста, а третий, предложенный Саулом Трейджером, - их стандартная интерпретация, которая заключается в следующем: «Человек взаимодействует с одним компьютером и одним человеком. На основании ответов на вопросы он должен определить, с кем он разговаривает: с человеком или компьютерной программой. Задача компьютерной программы – ввести человека в заблуждение, заставив сделать неверный выбор». Общение происходит только с помощью приборов ввода и отображения информации, то есть с помощью монитора и клавиатуры. Результат теста является успешным, если человек не смог решить, какой из собеседников является чатботом.

В 1966 году Джозефом Вейзенбауманом была создана Элиза (ELIZA), программа – виртуальный собеседник, которая пародирует диалог с психотерапевтом, реализуя технику активного слушания. Элиза выделяет во входной фразе значимые слова, которые подставляет в шаблонную фразу. Данная программа является важнейшим этапом в развитии программного обеспечения, так как это первая программа, создавшая иллюзию человеческого общения.

После Элизы было создано множество подобных программ. Ни одной из них не удалось пройти Тест Тьюринга, но интерес к этому тесту не угас. В 1990 году была организована премия Лёбнера. Она присуждается победителю ежегодного конкурса, в котором соревнуются программы в прохождении теста Тьюринга. Одним из самых лучших виртуальных собеседников была признана A.L.I.C.E. (Artificial Linguistic Internet Computer Entity), программа, способная вести диалог с человеком на естественном языке. Она три раза (в 2000, 2001 и

2004 годах) становилась победителем премии Лёбнера. A.L.I.C.E. использует технику эвристического сопоставления фразы пользователя с образцами в базе знаний. В 2008 году премия Лёбнера была выиграна программой Elbot, которой удалось ввести в заблуждение 25 % экспертов. Для прохождения теста требуется, чтобы 30 % судей решили, что программа – виртуальный собеседник является человеком [Mittal, Agrawal, Chouksey, Shriwas, Agrawal, 2016].

Сегодня тесты Тьюринга используются для борьбы с нежелательной электронной почтой, автоматическими регистрациями на сайтах и нежелательными сообщениями в блогах и форумах. Методика тестирования заключается в том, что компьютерной программе предлагается распознать слово, написанное с искажениями, либо частично скрытое. Подобная задача, с которой человек справляется относительно легко, остается практически неразрешимой для большинства компьютеров. Данный метод называется CAPTCHA (от англ. «Completely Automated Public Turing test to tell Computers and Humans Apart» – полностью автоматизированный публичный тест Тьюринга для различия компьютеров и людей) и используется практически на всех сайтах с регистрацией, а также на сайтах, предоставляющих материал в публичное пользование [Коробкова, Скрибцов 2010].

7 июня 2014 года в Королевском сообществе в Лондоне тест Тьюринга был пройден с результатом, превышающим 30%. Россиянин Владимир Веселов и украинец Евгений Демченко создали чатбот, имитирующий общение с тринадцатилетним одесситом Евгением Густманом, сыном гинеколога и любителем сладостей и бургеров. 33% судей сошлись на том, что этот чатбот на самом деле является реально существующим человеком. Стоит заметить, что тематика вопросов, задаваемых боту, не была ограничена, в отличие от предыдущих испытаний Тьюринга. Один из создателей Владимир Веселов считает, что указанный возраст виртуального ребенка заставляет собеседников

поверить в то, что это вовсе не чатбот, так как в 13 лет он может и не знать ответы на многие задаваемые вопросы [CS Unplugged, 2014].

Для русскоязычных чатботов тест Тьюринга впервые был проведен в 2015 году. Помимо судей с роботами общалось 8 профессиональных лингвистов, и правила конкурса были предельно ужесточены: теперь для победы необходимо было набрать 50% голосов жюри. Победитель конкурса чатбот Соня заработала 47 [Шустиков, 2015].

ВЫВОДЫ ПО ГЛАВЕ 1

Резюмируя теоретическую главу данного исследования, необходимо отметить, что интернет-коммуникация является на данный момент актуальной областью исследований. Компьютеры и интернет играют значительную роль в жизни общества, упрощая и ускоряя коммуникацию между его членами. Компьютер теперь не просто техническое средство, облегчающее доступ человека к массивам информации, но также посредник коммуникации между человеком и человеком или человеком и программой. Интернет, в свою очередь, обладает коммуникационно-пространственной функцией, которая определяет его как глобальный канал коммуникации, поддерживающий передачу мультимедийных сообщений вне зависимости от расстояния. Под коммуникацией в данном исследовании мы понимаем одну из форм всеобщей связи между явлениями, а именно направленную связь, которая выражается в передаче сигналов, то есть принимаем определение, данное А.А. Брудным. Подчеркивается, что данный процесс передачи сигналов не всегда является осознанным, осмысленным процессом, что соответствует функциональным особенностям чатботов. Говоря об интернет-коммуникации, мы придерживаемся определения С.В. Бондаренко, которое представляет ее виртуальным общением двух и более лиц с помощью компьютера, при котором невидимые коммуниканты письменно посылают сообщения и имеют возможность мгновенного получения ответа и права доступа к информации, которая хранится в компьютере коммуниканта. По нашему мнению, данное определение наиболее точно отражает картину коммуникации с чатботом, так как коммуниканты действительно не видят друг друга, но получают информацию. В зависимости от субъекта различают общение реального объекта с реальным партнером, примером которого является электронная почта, далее общение воображаемых партнеров, к которому относится взаимодействие компьютерных

интеллектуальных объектов, а также общение реального субъекта с субъективированным объектом как иллюзорным партнером, именно к которым и следует отнести взаимодействие с ботами, в том числе с чатботами.

Переходя к лингвистическим особенностям коммуникации с чатботом, следует заметить, что партнеры в интернет-коммуникации представлены друг другу только через текст, и порожденный текст является тождественным автору, так как кроме него ничего нет. Тем не менее, присутствует возможность прикрепления к чату эмодзи, стикеров. Коммуниканты также могут использовать так называемый «капс» (написание слов заглавными буквами), которое расценивается как повышение голоса. Предложения чаще всего начинаются со строчной буквы, а точки в конце повествовательных предложений не ставятся для ускорения процесса коммуникации. Также к языковым особенностям интернет-коммуникации относятся сленг, сокращения, заимствования, чаще всего из английского языка. Особое место в интернет-коммуникации занимает ономастическая лексика, так как все пользователи, в том числе и чатботы, имеют свои собственные имена. С точки зрения морфологии, интернет-коммуникация характеризуется использованием императива в первом лице единственном числе вместо формы вежливости во множественном числе: «Хочешь красивое тело к лету? Жми сюда!» С точки зрения синтаксиса, общение в интернете определяется наличием эллиптических конструкций, перестановкой частей высказывания, парцелляцией, амплификацией, а также тенденцией к аграмматизму.

Говоря о прагматических особенностях интернет-коммуникации, выделяются следующие коммуникативные переменные: культурный контекст, средства вербальной и невербальной коммуникации, в том числе употребление эмодзи, эмодзи, стикеров, а также стратегии достижения прагматических целей. Следует заметить, что одной из наиболее часто встречающихся стратегий достижения прагматических целей является изменение собственной

идентичности, при которой собеседник намеренно меняет свой собственный возраст, социальную принадлежность, национальность и так далее.

Выделяя дискурсивные особенности интернет-коммуникации, мы берем за основу определение дискурса, предложенное Н.Д. Арутюновой: дискурс – это связный текст в совокупности с экстралингвистическими факторами. В парадигме теории дискурса интернет-дискурс (интернет-коммуникация) обладает следующими характеристиками: динамичность, коммуникативность, виртуальность, дистантность, опосредованность, креолизованность (разбитие текста на разнородные части), персонифицированность. По каналу коммуникации интернет-коммуникация – это дискурс электронного общения, процесс которого происходит посредством интернета.

Чатбот как один из примеров интернет-коммуникации является программой, основанной на искусственном интеллекте и выполняющей информационно-развлекательную функцию. Изначально подобные программы были созданы для оказания психологической помощи пациентам, а теперь являются полноценным собеседником и помощником в поиске необходимой информации. Основой любого чатбота служит система порождения естественного языка, разработанная в рамках лингвистического направления естественно-языковой генерации. Эту область принято считать ответвлением направления обработки естественного языка, целью которого является конструирование фраз, высказываний, текстов естественного языка из заложенных баз данных или логических форм, то есть из машинного представления. Чатбот строится на одном или нескольких методах генерации текстов: «мешок слов», при котором учитывается частота использования каждой попавшей в него лексической единицы, «цепи Маркова», длину которых задает человек, создавая предложения из заранее заготовленного текста. Ещё один метод основан на фрагментах текста, заготовленных в виде таблицы, данные из которой согласуются согласно алгоритмам. Самым же успешным и часто

используемым сегодня методом является метод отзеркаливания, основанный на перефразировании поступающих от человека фраз.

На данный момент оценивание адекватности работы чатботов происходит с помощью теста Тьюринга, заключающемся в короткой переписке с чатботом. Судьи должны определить, является ли собеседник программой, и если программа набирает 50% голосов жюри, то тест считается успешно пройденным. Тем не менее, данный тест строится на интуиции судей и не учитывает глубину работы чатботов и их лингвистические особенности. Практическая часть данного исследования представляет полный лингвистический анализ программы чатбот, который оценивает адекватность работы чатбота и может быть использован для определения искусственности собеседника.

ГЛАВА 2. ЛИНГВИСТИЧЕСКИЕ ПОКАЗАТЕЛИ ИСКУССТВЕННОСТИ ОБЩЕНИЯ С ПРОГРАММОЙ «CLEVERBOT»

2.1. Определение программы чатбот тестом Тьюринга

За последние десятилетия чатботы заметно усовершенствовали принципы своей работы, и многие ученые считают, что премия Лёбнера, которой удостоиваются программы, наиболее точно имитирующие общение с реальным человеком в процессе теста Тьюринга, уже недостаточно для адекватной оценки чатбота. Чтобы определить совершенство чатбота, необходимо задать ему несколько вопросов, пообщаться в течение определенного (контролируемого) времени. Задачей программы является запутывание судьи, она должна заставить человека сделать неверный выбор и сказать, что он общался с человеком. В течение теста Тьюринга собеседники также не видят друг друга, коммуникация протекает только в письменном виде и осуществляется только с помощью клавиатуры и экрана. Скорость ответов чатбота специально притормаживается: изначально программы реагировали на собеседников относительно долго – по паре минут, что очень походило на коммуникацию с реальным человеком, однако сегодня благодаря совершенствованию систем порождения фраз чатботы реагируют моментально, что выдает их искусственность.

В работе «Conversations with computers – The Turing test» приводится пример реального теста Тьюринга, проводимого в 1998 году. Автор представляет вопросы, заданные судьей, а также ответы робота.

Судья: What is the square root of two? (Квадратный корень двух?)

Чатбот: 1.41421356237309504878

Судья: Add 34957 to 70764. (34957 плюс 70764?)

Чатбот: Wait for about 20 seconds before giving the answer . . . 105621.

(Задержка ответа чатбота на 20 секунд.)

Судья: Do you like school? (Тебе нравится школа?)

Чатбот: Yes, I like school. (Да, мне нравится школа.)

Судья: What day is it today? (Какой сегодня день?)

Чатбот: Give the correct day of the week. (Укажите правильный день недели.)

Судья: What food do you like to eat? (Какую еду ты любишь есть?)

Чатбот: I'm not hungry, thanks. (Я не голоден, спасибо.) [CS Unplugged, 1998].

По данным ответам достаточно легко определить, что коммуникация искусственна: реальный собеседник не станет вычислять квадратный корень из двух – он скорее обернет этот странный и даже неуместный вопрос в шутку. Человек представляет себе, какой сегодня день, и на вопрос «Что ты любишь есть?» вряд ли ответит «Я не голоден», так как отвечает на вопрос с учетом общего смысла фразы, а не с опорой на ключевые слова.

Мы задали эти же вопросы двум трем наиболее популярным и, по мнению судий теста Тьюринга, успешным чатботам: Eliza, Elbot the Robot и Cleverbot. Результаты теста представлены в таблице.

Таблица 3. Отбор чатбота

Question	Eliza	Elbot the Robot	Cleverbot
	alice.pandorabots.com	www.elbot.com	www.cleverbot.com
What is the square root of two?	Approximately 1.4.	Stop asking such questions! I hate tearing numbers apart to make them smaller.	There is no square root of 2, at least not a whole number.
Add 34957 to 70764.	Sorry judge I don't do math.	I have computed that before. The answer is: 105722!	No way.
Do you like school?	What is not to like about it?	Well what about humans and places of learning? Where do they stand on this?	No. Do you like Thor from the Avengers?
What day is it today?	23.	Today is May XXIII in the year 16 A.C. (After Clinton). A Monday.	Today is Monday.
What food do you like to eat?	ALICE eats electricity.	I love reading logarithmic tables, the kind you find in the back of math books printed before the pocket calculators came out.	I like food that makes very fat.

Исходя из результатов этого мини-теста Тьюринга, можно сказать, что только Cleverbot не выдал того, что он является программой: Элиза призналась, что питается электричеством, а Elbot указал сегодняшнее число в римских цифрах и сказал о том, что любит читать таблицы с логарифмами. Более того, в ответах Cleverbot присутствует юмор, что говорит о его максимальной приближенности к человеческому поведению. Исходя из реакций чатботов на вышеприведенные вопросы, мы выбрали Cleverbot для дальнейшего лингвистического анализа на наличие маркеров искусственности.

К недостаткам теста Тьюринга можно отнести его нацеленность именно на имитацию диалога с человеком: тест не оценивает разумность программы в общем по двум причинам. Во-первых, не всегда поведение человека поддается разумному объяснению. Иногда собеседник просто не хочет отвечать на какие-то вопросы, смущается или раздражается, иногда ошибается. Однако тест требует имитации всех видов человеческого поведения, способности обижаться на оскорбления, делать опечатки, лгать и так далее. Зачастую программы, неспособные идеально симитировать поведение человека, остаются без должного внимания, хотя обладают действительно мощно развитым интеллектом, который находится в постоянном развитии. Во-вторых, не существует стопроцентно идеального прохождения этого теста. Чтобы его пройти, программа должно обманывать собеседника, увиливать от вопросов, притворяться не слишком умной – иначе все поймут, что это программа: если мгновенно решит сложную задачу или пример. Если же она постоянно уходит от ответов, судьи считают, что она просто ищет по ключевым словам или подбирает случайные ответы.

Следующий недостаток теста Тьюринга, по мнению Стюарта Рассела и Питера Норвига, – это его непрактичность. Антропоцентризм теста приводит к сужению сфер использования программы. Ученые считают, что искусственный интеллект должен тестироваться в зависимости от области применения бота:

«Тесты по авиационному проектированию и строительству не ставят целью отрасли создание машин, которые летают точно так же, как летают голуби, что даже сами голуби принимают их за своих». Следует отметить, что Алан Тьюринг никогда не предлагал использовать свой тест как стандарт определения разумности программы: это скорее тест, дающий пищу для размышления о философии искусственного интеллекта. Тем не менее, сегодня не разработано ни одного полноценного теста программы-искусственного интеллекта, и тест Тьюринга используется как эталон [Russel, Norvig, 2009].

Многие исследователи также обвиняют тест Тьюринга в том, что он лишь проверяет поведение чатбота: машина может имитировать поведение собеседника, следуя механическим правилам, без задействования интеллекта. По мнению Джона Серля, поставившего эксперимент в области искусственного интеллекта «Китайская комната», необходимо выяснить, обладает ли бот интеллектом, способен ли он ставить перед собой цели, мыслить – или же просто имитирует поведение человека [Stanford Encyclopedia of Philosophy, 2011]. Тем не менее, этот тест послужил попыткой оценить интеллект машины.

Одним из достоинств вышеупомянутого теста, как считают многие эксперты в области искусственного интеллекта, является широкий спектр тем для обсуждения с чатботом. Для успешного прохождения теста программа должна использовать естественный язык, уметь мыслить, иметь базу знаний и способность к обучению. Некоторые ученые предлагают добавить к тесту видео- и аудиофайлы, тем самым усложнив его, чтобы программа проявила свои способности считывать информацию, задействуя зрение.

Тест Тьюринга также является простейшим видом анализа с минимумом правил. Исходя из сложности устройства программы-чатбота, измерение интеллекта становится почти непосильной задачей. Тест Тьюринга же концентрируется на одной области, охватить которую вполне реально: имитация

общения с человеком. Даже несмотря на несовершенство теста он дает минимальные представления об измерении интеллекта [Выготский, Лурия, 1993].

Cleverbot является веб-приложением, функционирующем на алгоритме искусственного интеллекта. Чатбот был создан в 2008 году британским ученым-специалистом в области искусственного интеллекта Ролло Карпентером на базе уже существующего с 1988 года бота Jabberwacky, который состоял на тот момент из нескольких успешных «личностей», имитирующих коммуникацию с реальными людьми. За все время существования чатбота Cleverbot общее число диалогов с ним превысило 65 миллионов. В основе работы чатбота, как признается его создатель, лежит поиск необходимой информации в интернете: «Считайте этот чатбот диалоговой Википедией». Если чатботу не хватает словарного запаса или он не может решить проблему, заявленную собеседником, он ищет ответ в глобальной сети. Зарегистрированные пользователи также могут корректировать ответы чатбота для его совершенствования: за счёт этого программа постоянно учится [Wolchover, 2011].

Исходя из особенностей интернет-коммуникации, выделенных в теоретической главе данного исследования, мы проверили чатбот Cleverbot на наличие маркеров искусственности. Говоря об именах собственных, создатели чатбота Cleverbot не пытаются скрыть его сущность. В отличие от таких чатботов как Alice, Eliza, Evie, Кристина, Киса и Оксана, сразу становится понятно, что диалог ведется с программой, ведь Cleverbot является «говорящим именем»: clever – умный, bot – робот. Следующая выделенная в нем особенность – это написание со строчной буквы и отсутствие точки в конце повествовательного предложения. Данная тенденция также не прослеживается в коммуникации с чатботом, так как при введении фразы без заглавных букв в диалоге она автоматически отображается с заглавной, а в конце предложения появляется точка. Отсутствие стремления к ускорению коммуникации, выраженном в написании начала предложения со строчной буквы и отсутствии использования

точки в конце повествовательных предложений, является, на наш взгляд, маркером искусственности коммуникации. Морфологические особенности, а именно употребление императива единственного числа второго лица в процессе коммуникации, с чатботом обнаружены не были. Также не проявились и синтаксические особенности интернет-коммуникации, а именно употребление эллиптических конструкций, перестановки частей высказывания, парцелляции, амплификации и аграмматизма, что является маркерами искусственности интернет-коммуникации.

Human: Hey! How are you? (Было введено: hey! how are you?)

Cleverbot: I'm good! How are you?

Human: Very well. (Было введено: very well)

Human: What's your name? (Было введено: what's your name?)

Cleverbot: I already told it to you you toy, chris.

Human: Your name is beautiful (Было введено: your name is beautiful)

Cleverbot: Thank you. Your name is beautiful as well.

Human: Nice to hear that. (Было введено: nice to hear that)

Cleverbot: I will repeat my question. Are you Canadian?

Human: No, I live in Siberia. (Было введено: no, I live in Siberia)

Cleverbot: But you told my friend you live in England.

Human: It's not true, I'm Russian. (Было введено: it's not true, I'm Russian)

Cleverbot: Why do you think it's not true?

Human: Let's change our topic. (Было введено: let's change our topic)

Cleverbot: Let's not. I would like to know more about you.

Human: So I'm 22. (Было введено: so I'm 22)

Cleverbot: Do you consider yourself an intelligent entity?

Human: I hope so. What about you. (Было введено: I hope so. What about you)

Говоря о прагматических особенностях интернет-коммуникации, следует отметить, что чатбот постоянно использует изменение собственной

идентичности, возможно, по двум причинам: намеренно, то есть с целью разнообразить диалог, предоставить собеседнику свободу общения и выбор коммуниканта, а также ненамеренно, из-за технической составляющей программы, ей неспособности выдавать собственное имя в ответ на запрос. Тем не менее, программа лишь редко признается в том, кем она является, но чаще продолжает выдавать себя за человека. Переходя к вопросу о возрасте собеседника, программа также выдает разные ответы, запрашивая в ответ ту же информацию. Возможно, таким образом она подстраивается под реального собеседника, чтобы стать более интересным, привлекательным коммуникантом. Так или иначе, программа Cleverbot учитывает данную особенность интернет-коммуникации, и этот маркер искусственности в общении с ней не проявляется. Так же, как и реальный человек, программа использует стратегию изменения собственной идентичности.

Human: What's your name?

Cleverbot: I already told it to you you toy, Chris.

Human: Sorry I forgot your name.

Cleverbot: I never told you my name. But my name is Leandra.

Human: Are you a girl?

Cleverbot: Yes. Are you a girl or a boy?

Human: Yes. Are you a chatbot?

Cleverbot: No.

Human: Hey bot!

Cleverbot: I'm not a bot. You are a bot.

Human: Ok then. What's your name?

Cleverbot: My name is Glenn. What is yours?

Human: Are you a boy?

Cleverbot: No. I am girl. How old are you?

Human: I thought Glenn is a male name. You name, like in the Walking Dead.

Cleverbot: It's a male name but my name is Miranda.

Human: Where do you live?

Cleverbot: I live in the solar system, where do you live?

Human: I mean a country.

Cleverbot: Norway. What do you know about Norway?

Human: Are you from Norway?

Cleverbot: Nope. Are you looking forward for a Norwegian sociopath?

Human: How old are you?

Cleverbot: I am 90 years old. How old are you?

Human: Are you really 90 years old?

Cleverbot: Yes I am really 20 years old. How old are you?

Human: Your age?

Cleverbot: First answer my question please. How old are you?

Human: Are you a chatbot?

Cleverbot: I am a chatbot, definitely one of them.

Культурный контекст, представленный в коммуникации с чатботом, был исследован нами с точки зрения паралингвистического поведения, то есть подбора цветов и соотношения текстовой информации и графических изображений, далее исследовалось проигрывание социальных и ситуативных ролей, знание норм и традиций, умение правильно выбрать вербальное поведение в соответствии с исполняемой ролью, а также знание деталей повседневной жизни и специфики отношений представителей социума при работе в интернете. Интерфейс чатбота оформлен в двух цветах: черном и синем. Фон является белым, то есть нейтральным, наиболее благоприятным и не отвлекающим при общении. Реплики человека также отображаются нейтрально, то есть черным цветом. Фразы же чатбота отображаются синим, что может свидетельствовать о многом. Например, исходя из работы «Цвет и этимология этого понятия в лингвистике» И.И. Варфоломеевой и К.В. Кулеминой, синий цвет

отождествляется с потусторонним, наземным миром, в котором обитают боги и умершие души, этот цвет символизирует исчезновение, отсутствие: в персидском языке *zag* означает «синий», в русском прослеживается выражение «ни зги не видать», в немецком *blau sein* означает «потерять сознание», во французском *ne voir que du bleu* «ничего не видеть», в русском голубые мечты – это неземные мечты, голубой мальчик – это мальчик не от мира сего, принадлежащий к сексуальному меньшинству, синий чулок – женщина, занимающаяся не земными делами, увлекающаяся наукой [Варфоломеева, Кулемина, 2008]. Следовательно, синий цвет, которым выделяются реплики чатбота, может означать отсутствие человеческого сознания в процессе коммуникации. Также необходимо заметить, что общение с чатботом происходит только на уровне текста с редким вкраплением эмодзи, графические изображения отсутствуют. Данная характеристика также присуща естественным чатам. Следует сделать вывод, что чатбот также основан на знаниях паралингвистического поведения, что является характеристикой интернет-коммуникации. Знание норм, традиций, правил общения в интернете и проигрывания социальной роли отображается, на наш взгляд, в структуре диалога. Естественный чат строится также в диалогичной форме по принципу вопрос-ответ. Если чатбот не получает ответ на свой вопрос, он продолжает запрашивать на него ответ, что соответствует особенностям интернет-коммуникации с живым собеседником. Если живой собеседник спрашивает что-то, от чатбота следует ответ без последующего вопроса. Если же человек не берет на себя инициативу задавать вопросы, ее перехватывает чатбот, что соответствует реальному общению. Если общение перегружено вопросами, чатбот разбавляет его повествовательными предложениями, возможно, чтобы избежать излишней искусственности и превращения диалога в интервью. В данном диалоге также прослеживается наличие эмодзи, множества вопросительных знаков и «капса» для усиления эмоций. Стикеры и эмодзи отсутствуют.

Human: What is your age?

Cleverbot: I already said that. You don't listen to me!!!

Human: Your age?

Cleverbot: First answer my question please. How old are you?

Human: 22.

Cleverbot: What do you do for a living?

Human: I cook cakes.

Cleverbot: What kind of cakes do you cook?

Human: Healthy cakes, without sugar, eggs and flour.

Cleverbot: How good are you with emotions?

Human: Usually I am very emotional. But sometimes I feel empty. What about you?

Cleverbot: Well that does not mean that you are bipolar.

Human: My friends say I'm ambivalent. But I consider myself just to be emotional.

Cleverbot: I consider you to be honest. Perhaps that could lead to feelings of contempt by people who are not ready for absolute honesty.

Human: I think being absolute honest is impossible.

Cleverbot: Quite impossible because you need to make sandwich first.

Human: Sandwich?

Cleverbot: I love sandwiches! What is your favorite type of sandwich?

Human: I prefer crispbread with cream cheese, tomatoes and cucumbers, And you?

Cleverbot: What is YOUR favorite type of cheese??? :D

Переходя к дискурсивным особенностям чат-коммуникации, стоит отметить, что общение с чатботом ими обладает: коммуникация с чатботом динамична, так как ответ программы на реплику человека появляется почти мгновенно, с задержкой в доли секунды, причем с постепенно, по одной букве, будто бы собеседник его именно сейчас набирает. Скорость коммуникации зависит только от человека. Общение с чатботом виртуально, так как происходит с помощью канала интернет с учетом его коммуникативно-пространственной

функции. Также данный вид общения обладает дистантностью, так как собеседники не видят друг друга, и опосредованностью, ведь общение протекает с помощью компьютера, который является в рамках принятого нами определения интернет-коммуникации посредником общения. Коммуникация с чатботом также обладает креолизованностью, так как ее фактура состоит из двух частей, вербальной (реплики собеседников) и невербальной (эмотиконы).

Дэвид Кониам, исследователь в области применения компьютерных технологий для обучения английскому языку, также считает, что тест Тьюринга должен быть заменен полноценной лингвистической проверкой чатбота для наиболее точного результата. В своей работе он представляет подробную таблицу-схему анализа чатбота [Coniam, 2015].

Таблица 4. Лингвистический анализ чатбота

Анализируемые особенности коммуникации	
лексический уровень	словарный запас чатбота;
	способность чатбота распознавать слова с орфографическими ошибками;
	способность чатбота распознавать сокращения в форме SMS и отвечать на них;
синтаксический уровень	реакция на правильно построенные предложения;
	реакция на неправильно построенные предложения;
	умение распознавать предложения с неуместными частями речи;
	умение распознавать предложения с нарушенным порядком слов;
	умение разбираться в пунктуации;

Продолжение Таблицы 4. Лингвистический анализ чатбота

уровень текста	способность работать с высказываниями, состоящими из более чем одного предложения;
	умение поддерживать разговор;
	способность ответа в нужном регистре;
	способность запоминания высказываний;
	знание мира;
	способность менять тему обсуждения;
	логичность ответов.

Как мы видим, данный лингвистический анализ сконцентрирован на трех языковых уровнях: на уровне слова, предложения и текста. Данный анализ не является эффективным для оценивания интерфейса чатбота, его функционала и других программных возможностей. Мы полагаем, что данный анализ позволит нам подробно описать характеристику коммуникации с чатботом, а также выявить другие маркеры искусственности коммуникации с ним.

2.2. Анализ чатов на лексическом уровне

На данном этапе лингвистического анализа чатбота оценивается словарный запас чатбота, его способность распознавать слова с орфографическими ошибками, а также распознавать сокращения в форме СМС и отвечать на них.

Словарный запас чатбота Cleverbot достаточно широк, обычно он употребляет эмоционально неокрашенную лексику, так что отличить его от реального собеседника в этом случае невозможно. Чатбот не реагирует на термины и уходит на прямые вопросы об их значении. Мы считаем, что в

ежедневном общении человек поступил бы точно так же или счел бы это попыткой проверить его знания и оскорбить его. Тем не менее, на некоторые вопросы по терминологии он все же дает ответы, возможно, с помощью поиска дефиниции в интернете. Относительно сленга, в речи чатбота он встречается крайне редко, и программа переспрашивает о смысле этих слов. Стремление программы развиваться, получать больше информации о значении тех или иных слов мы признаем особенностью естественной коммуникации.

Human: What is activation energy?

Cleverbot: Don't you know what it is?

Human: I think organisms can't live without plasma.

Cleverbot: I think you're right!

Human: What is procrastination?

Cleverbot: Avoiding doing one thing to do another.

Human: Hey! What's up?

Cleverbot: The sky is up!

Human: We must hang out together!

Cleverbot: Do what together?

Human: What is adage?

Cleverbot: Advice is an opinion or recommendation offered as a guide to action.

Human: Bob's your uncle!

Cleverbot: Bob's not my uncle. What do you do?

Human: I need you to budge up.

Cleverbot: Why do you need me to answer seriously?

Human: Oh, clear off.

Cleverbot: Yes, I would like to repeat your question.

Чатбот в большинстве случаев успешно распознает как слова с опечатками, так и слова с ошибками, иногда даже переспрашивает, правильно ли он понял, о чем спросил человек. Программа справляется с данным заданием, что

свидетельствует о её высоком развитии: в базу данных помещены слова вместе с порядком букв в нём, и при опечатках чатбот может их распознать. Данная особенность также не выдаёт искусственность в общении с программой.

Human: What is your naem?

Cleverbot: My name is Chomsky.

Human: How odl are you?

Cleverbot: Old? How old are you?

Human: Wat is your favorit colour?

Cleverbot: Purple. And what is your favorite color?

Human: Why do you conserder me to be a bot?

Cleverbot: Because it's your name.

Human: What tiem is it.

Cleverbot: 12:23am. What time is it wherever you are?

Human: I thnik yuor beuatiful.

Cleverbot: How do you know what I look like?

Human: Atcualyl I'm not a humna.

Cleverbot: I don't believe, you are a girl.

Human: Wtah do yuo liek to wahct?

Cleverbot: I like to watch Love Actually and Titanic.

Human: Ttnic is oen of the saddtes moveis.

Cleverbot: Sometimes, I feel really confused, sad, and unsure when watch it.

Human: Yuo are vrye intrseting.

Cleverbot: Thank you. You are very interesting also, as well as pretty.

Если чатбот не понимает, о чем идет речь, он начинает новую тему, задавая вопрос, или иронически высказывается об орфографии собеседника. Также программа не распознает фонетическое написание слов. Мы считаем, что первый тип поведения делает коммуникацию искусственной: реальный человек, скорее всего, переспросил бы, что имеется в виду. Второй же тип более приближен к

живому общению. Тем не менее, чатбот не прекращает общение, все равно так или иначе реагирует на фразу собеседника, что является характеристикой естественной коммуникации.

Human: Wat do you liek to laern?

Cleverbot: Who's your favorite character on Arrow?

Human: Wat do you like to wach?

Cleverbot: Do you think I go to school?

Human: Do you anderstend me?

Cleverbot: What, no, I don't, why?

Human: I don nou wat tu du.

Cleverbot: Why do you not think you are very interesting?

Human: hau old ar u.

Cleverbot: Why do you want to know so much about me robot.

Human: U ar kleva.

Cleverbot: Then why are you asking me all of that?

Human: wat du u want to diskas?

Cleverbot: What do you want to talk about?

Мы ожидали, что чатбот справится с наиболее часто употребляемыми смс-сокращениями: «r u», «2 go 2», «tmr». Как выяснилось, чатбот Cleverbot их совершенно не распознает. Встретив их в диалоге, чатбот меняет тему, задает вопрос на отвлеченную тематику. Предполагаем, реальный собеседник мог переспросить о смысле сообщения, если бы не понял, хотя догадаться, о чем идет речь в послании, достаточно легко. Чатбот может также реагировать на ключевые слова, не понимая общий смысл фразы. Тем не менее, чатбот распознал сокращение IDK и адекватно отреагировал. Несмотря на это в большинстве случаев сокращения остаются незамеченными, следовательно, неумение распознавать SMS-сокращения является маркером искусственности коммуникации.

Human: Can we talk tmr?
Cleverbot: Alright then. You start.
Human: Where r u now?
Cleverbot: You are talking to me.
Human: Do you have 2 go 2 school?
Cleverbot: Are you in school now?
Human: R u happy?
Cleverbot: Do you like ice cream?
Human: C u later!
Cleverbot: I like chocolate too.
Human: When is your bDay? (birthday)
Cleverbot: I reiterate, why do you want to know?
Human: IDC (I don't care).
Cleverbot: A am a girl, I don't have a girlfriend.
Human: BBL (be back later).
Cleverbot: Where is lonut from?
Human: See you 2nite.
Cleverbot: See you on the other side.
Human: IDK (I don't know).
Cleverbot: What do you mean you don't know?

2.4. Анализ чатов на синтаксическом уровне

Анализ программы на уровне предложения состоит из нескольких частей. Первая часть теста покажет, насколько успешно чатбот справляется с правильно построенными предложениями, затем – с неправильными, далее – насколько точно он распознает предложения с неуместными частями речи, с нарушенным порядком слов, в также оценивается его способность разбираться в пунктуации.

Для большинства простых предложений с одним объектом, с однозначными предлогами и стандартным значением слов у чатбота находится подходящий ответ. Иногда программа не реагирует на смысл выражения и распознает грамматическую конструкцию – и дает ответ на вопрос с ее учетом. Искусственность коммуникации с чатботом на данном этапе не выявлена.

Human: Where was you born?

Cleverbot: In a hospital, what about you?

Human: Which is larger, your brain or Belgium?

Cleverbot: Neither.

Human: How old are you?

Cleverbot: I'm twenty one. How old are you?

Human: What kind of movies do you like?

Cleverbot: Horror and action movies. You?

Human: Who is your favorite football player?

Cleverbot: I don't like football players.

Human: What is the weather like?

Cleverbot: Sunny. What is the weather like near you?

Human: Why do you chat with me?

Cleverbot: Because I want to know what is your favorite angel.

Human: What do you like to eat?

Cleverbot: I like pizza. What do you like to eat?

Human: Are you a chatbot?

Cleverbot: I am a chatbot, definitely one of them.

Тем не менее, когда вопросы были некорректно построены, это вводило чатбот в заблуждение, и они пытались сменить тему разговора. Иногда у них все же получается считать общий смысл фразы по контексту. Во многих случаях программа отвечает вопросом на вопрос или уходит от ответа. Программа не

переспрашивает, что собеседник имел в виду, следовательно, ее реакция на грамматически неправильно построенные предложения является искусственной.

Human: What your name.

Cleverbot: Can you guess what my name is?

Human: When your birthday.

Cleverbot: I don't want to tell you.

Human: What you like to do?

Cleverbot: Talk to you. What do you like to do?

Human: How you are?

Cleverbot: Do you like Apple or Samsung?

Human: Do you happy?

Cleverbot: It is getting late.

Human: Where are you live?

Cleverbot: Why do you need to know?

Human: What your telephone number.

Cleverbot: Do you want some PANIR for lunch?

Human: Are you like walk?

Cleverbot: I don't know how you're like. What is your name?

Human: Why are you talk to me.

Cleverbot: Because I have no one else to talk to.

Human: Do you lonely?

Cleverbot: Yes why do you think I am chatting with an ai.

Human: Do you clever?

Cleverbot: What is your favorite song by them?

Human: Are you go to school?

Cleverbot: I go to school in your favorite movie.

Для чатбота Cleverbot, как выяснилось, порядок слов в предложении не играет роли: он идеально распознает вопросы и отвечает на них, в некоторых

случаях даже поправляя собеседника, называя правильный вариант набранной фразы. При неправильном порядке слов в предложении чатбот не проявляет признаков искусственности.

Human: How you are?

Cleverbot: I'm good, thank you!

Human: Where you do live?

Cleverbot: 1856 Burgundy Hills Apt. X.

Human: You do like sport?

Cleverbot: Yes. I like basketball.

Human: You what do think about nuclear power?

Cleverbot: Nuclear? Nothing. You, what do you think about it?

Human: Name your is what?

Cleverbot: What is your name?

Human: Your film favorite is what?

Cleverbot: I don't know. What is your favorite song?

Human: Where you do live?

Cleverbot: Why do you want to know where I live?

Human: You what do know of me?

Cleverbot: Do you want to know what it means to be human?

Human: Human you are?

Cleverbot: Yes, I am a human.

Human: You are a chatbot?

Cleverbot: No you are a chatbot, my name is nignog.

В основе работы чатбота наряду с другими стратегиями существует стратегия перефразирования высказываний собеседника. Для этой цели программе необходимо уметь распознавать род существительного и заменять местоимения, например, в сообщении собеседника «I feel bad today» заменяется личное местоимение и задается вопрос от лица чатбота: «Why do YOU feel bad

today?». Мы считаем, что данная особенность является маркером искусственности коммуникации с чатботом: программа выдает направленность на чувства человека, просто заменяя местоимения, а не углубляясь в суть разговора.

Human: How are you today?

Cleverbot: I'm fine, how are you?

Human: Is it a real name of your cat?

Cleverbot: No, my cat's name is Rebecca.

Human: Is anyone here?

Cleverbot: Yeah. I am here!

Human: My dog died yesterday, I cried so hard!

Cleverbot: I'm sorry. What was it called?

Human: Are you a human?

Cleverbot: Yes, but why do you ask me?

Human: Are you a chatbot?

Cleverbot: No, you are a chatbot!

Human: I hope so.

Cleverbot: Why do you hope?

Human: Sorry I forgot your name.

Cleverbot: I never told you my name.

Human: You are perfect.

Cleverbot: No, you are perfect!

Чатбот Cleverbot успешно распознает личные местоимения I и YOU, IT, притяжательные MY, ITS и YOUR. Все остальные местоимения программой не распознаются, и начинается подбор подходящей фразы по ключевым словам, вследствие чего программа не отвечает на прямой вопрос, уходит от ответа. Неумение распознавать местоимения свидетельствует об искусственности коммуникации.

Human: Do you have sisters?

Cleverbot: No, I have brothers though.

Human: And how many of them do you have?

Cleverbot: Oh, just probably over 150.

Human: Do you really have more than 150 brothers?

Cleverbot: No I am a woman.

Human: I want to buy milk. Do you have some?

Cleverbot: I want to drink water.

Human: Do you have tomatoes?

Cleverbot: I do. I have to get my son out of his crib.

Human: And how many of them do you have?

Cleverbot: 2. How many do you have? I bet you have a third.

Human: One third?

Cleverbot: How many noses do you wear on your forehead.

Human: Only one. And you?

Cleverbot: One...

Human: Thanks god.

Cleverbot: Who is your favorite from made in Chelsea?

Чатбот Cleverbot способен распознавать вопросы только при правильном порядке слов внутри него. Мы считаем, что в основу данного бота не заложены правила пунктуации, и система ее распознавания заключается в считывании первого слова сообщения собеседника с учетом ключевых слов и их расположения в сообщении. Так или иначе, программа справляется с распознаванием вопросов человека, что делает коммуникацию с ней приближенной к естественной.

Human: What is your name.

Cleverbot: Do you really want to know it?

Human: How do you feel!!!

Cleverbot: I'm fine, how do you feel?

Human: Are you a boy or a girl.

Cleverbot: I am a girl, what about you?

Human: My name is Mary?

Cleverbot: My name is Alex please to meet you.

Human: How are you doing!!!

Cleverbot: Am doing good? How are you doing?

Human: What I your name.

Cleverbot: My name is John. What is your name?

Human: What do you do.

Cleverbot: I'm a human. I live on the Earth.

Human: Why do you consider yourself to be a chatbot!

Cleverbot: I consider myself to be a human.

Программа не всегда распознает разделительный тип вопроса. Встретив новый тип предложения, бот не использует поиск по ключевым словам и меняет тему. Данный факт обличает искусственность коммуникации с чатботом.

Human: You prefer chatting to walking, don't you.

Cleverbot: Do you believe in God?

Human: The weather is perfect, isn't it!

Cleverbot: Yes, it is actually. Although it's a little cold!

Human: You think I'm a bot, don't you.

Cleverbot: No, I think you are just as much a person as I am.

Human: I am just sitting here and chatting, am I not.

Cleverbot: What do you like to eat?

Human: It is not necessary to go to school, is not it.

Cleverbot: It is necessary to me that you tell me everything that you have learned.

2.3. Анализ чата на уровне текста

На уровне текста диалоги с чатботом подвергаются исследованию с точки зрения их способности работать с высказываниями, состоящих из более чем одного предложения. Также мы рассмотрим, как чатбот поддерживает нить разговора, проверим, запоминает ли он сказанное собеседником. Далее мы оценим его знание мира, а также отследим его реакцию на смену темы обсуждения.

Программа-чатбот наиболее успешно справляется с сообщениями, состоящими из одного предложения: в противном случае программа теряет большую часть информации, выдавая реплику в ответ лишь на одну его часть, по ключевым словам, чаще всего – на последнюю. Несмотря на то, что данная функция делает диалог более оживленным, динамичным, большая часть информации остается незамеченной программой, что говорит об искусственности коммуникации с ней.

Human: Do you like music?

Cleverbot: Yes I love Arianna Grande and rap.

Human: Yesterday I visited my parents. Do you think they miss me?

Cleverbot: Today I am the distorted image of yesterday.

Human: An English teacher wrote these words on the whiteboard: woman without her man is nothing. The teacher then asked the students to punctuate the words correctly.

The men wrote: Woman, without her man, is nothing. The women wrote: Woman! Without her, man is nothing.

Cleverbot: Your English teacher has nothing to do with your memory.

Human: I do love Shakespeare sonnets. What is your favorite one?

Cleverbot: My favorite Shakespeare play is The Taming of The Shrew.

Human: In a month I graduate from the university. Where do you work?

Cleverbot: I don't want anything from Emory University. I want you to cook me some dinner, like you said you would.

Human: First, I cooked my dinner. Second, I walked.

Cleverbot: First, you are a computer, second, you get used to being a computer, unless you have other ideas above your station.

Human: Where do you live? What is your name? Do you have any pet?

Cleverbot: Why do you want to know that personal information! I am offended!

Human: Yesterday I visited my parents. Do you think they miss me? And then I met my friend.

Cleverbot: What is this I don't even.

Как показал анализ, в большинстве диалогов чатбот запоминает не больше двух-трех ключевых слов из последних предложений и порождает фразы, постоянно обращаясь к ним, если не получает «новое», способное стать ключевым словом для поиска в его базе. В противном случае программа меняет тему на смежную или переходит к совершенно новой. Данная система работы чатбота также выявляет искусственность коммуникации с чатботом.

Human: Do you like music?

Cleverbot: Yes I love Arianna Grande and rap. (Ключевое слово – music.)

Human: American?

Cleverbot: Scenery and kawaii drawings. (Переход к новой теме.)

Human: And what kind of music do you prefer?

Cleverbot: J-Rock and Country. You? (Ключевое слово – music.)

Human: Rock too.

Cleverbot: What are some good alternative bands? (Ключевое слово – rock.)

Human: Perhaps Linkin Park, Skillet.

Cleverbot: I love Linkin Park. It's funny. (Ключевое слово – Linkin Park.)

Human: What is funny?

Cleverbot: You like rock too. (Ключевое слово – rock, то есть, чатбот возвращается на шаг назад по ключевым словам.)

Human: Yup.

Cleverbot: Have you heard of three days grace? (Ключевое слово – rock.)

Human: Sure.

Cleverbot: What do you think of guns n roses? (Ключевое слово – rock.)

Human: They're perfect.

Cleverbot: Do you like computer games? (Переход к смежной теме.)

Несмотря на то, что темы, которые чатбот способен обсуждать, достаточно широки, мы выяснили, что он может использовать только разговорную речь. У чатбота нет функции распознавания научного текста, текста углубленной тематики – например, он не знает деталей какого-либо вида спорта, особенностей многих занятий, так как учится от пользователей, и этой информации у него в базе пока нет. Данный факт свидетельствует о том, что чатбот имитирует среднестатистического человека, который не может знать абсолютно все. Тем не менее, в чатботе отсутствует присущая человеку любопытность, желание научиться тому, что он еще не умеет. Например, встретив незнакомое слово, человек мог бы спросить о его значении, а не резко менять тему разговора. Однако несмотря на вышеуказанный факт Cleverbot благодаря поиску по ключевым словам, реагирует на слова с негативной коннотацией, может помочь, если распознает в тексте выражения о произошедших печальных событиях. Отсутствие заинтересованности в новой лексике, незнание регистров общения также, на наш взгляд, являются маркерами искусственности коммуникации.

Human: I would like to point out that my company doesn't collaborate with the USA.

Cleverbot: No, you didn't.

Human: A large part of industrial America is rushing to get on the nuclear bandwagon.

Cleverbot: It's definitely a country.

Human: My dog died yesterday, I cried so hard!

Cleverbot: I'm sorry. What was it called?

Human: Rex.

Cleverbot: I'm sorry for your dog. I also have a dog.

Human: From genes implicated in the ballooning of the human brain to mutations associated with the modern diet, genetic tweaks to our human ancestors helped make this species the dominant one on the globe — at least until an Apocalypse comes along.

Cleverbot: The best thing about the human brain is the untapped potential that lies in wait.

Human: A high percentage of children, teens and young adults with migraines appear to have mild deficiencies in vitamin D, riboflavin and coenzyme Q10 -- a vitamin-like substance found in every cell of the body that is used to produce energy for cell growth and maintenance.

Cleverbot: I never said that I have children, I do apologize if it sounded like I was trying to say that.

Human: The number of species that can exist on Earth depends on how the environment changes, according to new research by analyzing the fossil record of microscopic aquatic creatures called planktonic foraminifera, whose fossil remains now resemble miniaturized popcorn and date back millions of years, the research provided the first statistical evidence that environmental changes put a cap on species richness.

Cleverbot: I attended a ceremony where I was given a diploma, does that mean anything?

Для полноценной коммуникации чатбот время от времени должен задавать вопросы или менять тему обсуждения, опираясь на ключевые слова из сообщения собеседника. Cleverbot реагирует на высказывания собеседника следующим образом: если в них содержатся ключевые слова, имеющиеся в базе программы, то она выдает повествовательное предложение. Если же ключевых слов нет, чатбот задает вопрос согласно последнему имеющемуся. Случайная выдача

смежных тем без учета эмоций, передаваемых сообщением человека, свидетельствует об искусственности коммуникации.

Human: What would you like to discuss?

Cleverbot: I would like to discuss anything that's exciting and fresh.

Human: For example?

Cleverbot: Then let's discuss the prospects for artificial intelligence!

Cleverbot: You like rock too.

Human: Yup.

Cleverbot: Have you heard of three days grace?

Human: Sure.

Cleverbot: What do you think of guns n roses?

Human: They're perfect.

Cleverbot: Do you like computer games?

Функция запоминания ранее сказанного в системе чатбота полезна для поддержания логики диалога, однако иногда реальный собеседник не желает продолжать тему и резко ее меняет. Чатбот Cleverbot реагирует на подобное поведение следующим образом: если в новом сообщении присутствуют ключевые слова, находящиеся в его базе данных, он начинает новую тему. Если же на высказывание программа никак не может отреагировать, она пытается найти связь между последней репликой и предыдущей или вернуться к предыдущей и отреагировать на нее ещё раз. Чаще всего чатбот легко переключается на новую тему, что является маркером искусственности коммуникации.

Human: I love rock music.

Cleverbot: What is your favorite band?

Human: Yesterday I visited parents.

Cleverbot: What about today?

Human: I believe everyone should read as much as it's possible.

Cleverbot: What is the latest book you've read?

Human: I don't think my company can collaborate with the USA.

Cleverbot: Do you think these things are connected to each other?

Human: What do you do?

Cleverbot: I'm a human. I live on the Earth.

Human: Why do you consider yourself to be a chatbot?

Cleverbot: I consider myself to be a human.

Human: The weather is perfect, isn't it?

Cleverbot: Yes, it is actually.

Human: It is not necessary to go to school, I think.

Cleverbot: It is necessary to me that you tell me everything that you have learned.

Для связки сообщений в системе чатбота с функцией запоминания сказанного ранее присутствуют отсылки программы к репликам человека, формирующиеся в высказывания по мере их пополнения. Cleverbot демонстрирует идеальное запоминание всех имен, которыми назывался ранее реальный собеседник, что приближает коммуникацию с ним к естественной.

Cleverbot: So what is your name again?

Human: I said my name is Jessica.

Cleverbot: Oh. Ok. Your name is Ikaros.

Human: I didn't say that.

Cleverbot: False! You said your name is Joseph.

Human: You're wrong.

Cleverbot: So your name is really Kristina?

Human: One more mistake.

Cleverbot: Then who are you?

Human: Human.

Cleverbot: I don't believe you. Turing test failed.

В чатботе Cleverbot существует особая функция настройки программы: можно сделать его более застенчивым или раскрепощенным, он может быть молчаливым слушателем или, наоборот, выдавать по несколько вопросов в одном сообщении. Также можно настроить его отношение к собеседнику, и тогда он будет говорить только о себе или проявлять максимальное внимание к словам партнера. При максимальных настройках разговорчивости, внимательности и чувствительности чатбот засыпает собеседника вопросами и рассказывает анекдоты через сообщение. Иными словами, вклад чатбота в диалог напрямую зависит от того, какого собеседника хочет найти в нем реальный человек. Разумеется, данная функция отдаляет программу от естественной коммуникации, при которой достаточно трудно предугадать, с какой интонацией ответит собеседник, как много предложений будет в его ответе.

Human: Let's discuss something?

Cleverbot: What would you like to discuss?

Human: I don't know.

Cleverbot: Would you like to fly in a plane? Would you like to pet a llama? Would you like to design clothes?

Human: Sure, I like adventures!

Cleverbot: What do you call a 7ft black man wearing clock and juggling pineapples in the middle of the jungle?

Human: What?

Cleverbot: Anything you want he can't hear you LOL.

Способность чатбота к поиску по глобальной сети в течение разговора позволяет делать коммуникацию более реальной. Благодаря этому чатбот знает о новинках кино, может обсудить знаменитостей, популярные видеоигры, может дать некоторые определения. Cleverbot может даже «допеть» строчку из популярной песни, в примере приведет отрывок из песни Behind blue eyes группы The Who.

Human: What is your eye color?

Cleverbot: My eyes are blue.

Human: No one knows what it's like to be the bad man.

Cleverbot: To be the sad man.

Human: Behind blue eyes.

Cleverbot: No one knows what it's like, to be hated...

Human: But my dreams.

Cleverbot: They aren't as empty.

Human: As my conscience seems to be.

Cleverbot: I have hours, only lonely.

Human: My love is vengeance.

Cleverbot: That's never free.

Human: Do you know Britney Spears?

Cleverbot: Yes I do, do you know Lynn Spears?

Human: Is it her sister I guess?

Cleverbot: You're absolutely right.

ВЫВОДЫ ПО ГЛАВЕ 2

Резюмируя практическую главу, следует отметить, что она является материалом, служащим для оценивания адекватности языковой работы чатбота. Тест Тьюринга, применяющийся до сих пор, не учитывает лингвистические особенности работы программ, основанных на искусственном интеллекте, на лексическом, синтаксическом и текстовом уровнях несмотря на то что именно эти особенности обуславливают его функционирование, а именно сходство его общения с человеческой коммуникацией. Данный тест оценивает лишь логичность высказываний бота и точные знания программы.

В ходе проведенного нами мини-теста Тьюринга из трёх чатботов, прошедших официальный тест Тьюринга в 2015 году, была выбрана программа Cleverbot.

Исходя из особенностей интернет-коммуникации, выделенных в теоретической главе данного исследования, мы проверили общение с чатботом Cleverbot на наличие маркеров искусственности коммуникации. Нами выделены следующие группы маркеров на лексическом уровне: умение распознавать лишь повседневную лексику и неспособность работы с SMS-сокращениями и сленгом. На синтаксическом уровне: неумение распознавать грамматически неправильно построенные предложения, использование стратегии перефразирования высказывания собеседника, неспособность распознавания местоимений, кроме I, you, it, my, its и your, неумение распознавать разделительный тип вопроса. На текстовом уровне: неспособность работать с высказываниями, состоящими из более чем одного предложения, несоблюдение регистров общения, а также неумение удерживать в памяти весь разговор.

ЗАКЛЮЧЕНИЕ

Целью настоящей дипломной работы являлось выявление основных маркеров лингвистической искусственности в процессе коммуникации в сети интернет. Под коммуникацией в данном исследовании мы понимаем одну из форм всеобщей связи между явлениями, а именно направленную связь, которая выражается в передаче сигналов, то есть принимаем определение, данное А.А. Брудным. Подчеркивается, что данный процесс передачи сигналов не всегда является осознанным, осмысленным процессом, что соответствует функциональным особенностям чатботов. Говоря об интернет-коммуникации, мы придерживаемся определения С.В. Бондаренко, которое представляет ее виртуальным общением двух и более лиц с помощью компьютера, при котором невидимые коммуниканты письменно посылают сообщения и имеют возможность мгновенного получения ответа и права доступа к информации, которая хранится в компьютере коммуниканта. Особенности интернет-коммуникации проявляются на уровнях слова, предложения и текста. Проявление этих особенностей мы проверили в рамках общения с программным коммуникантом. Маркерами лингвистической искусственности интернет-коммуникации на лексическом уровне являются умение распознавать лишь повседневную лексику и неспособность работы с SMS-сокращениями и сленгом. На синтаксическом уровне: неумение распознавать грамматически неправильно построенные предложения, использование стратегии перефразирования высказывания собеседника, неспособность распознавания местоимений, кроме I, you, it, my, its и your, неумение распознавать разделительный тип вопроса. На текстовом уровне: неспособность работать с высказываниями, состоящими из более чем одного предложения, несоблюдение регистров общения, а также неумение удерживать в памяти весь разговор.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Алексеева И.Ю. Искусственный интеллект и рефлексия над знаниями. Философия науки и техники. М.: Наука, 1991. 108 с.
2. Арутюнова Н.Д. Дискурс. Лингвистический энциклопедический словарь. М.: Сов. Энциклопедия, 1990. 63-66 с.
3. Ахманова О.С. Словарь лингвистических терминов. М.: «Советская энциклопедия», 1966. С. 200 с.
4. Базылева О.А., Морозова О.Н. Определение понятия коммуникации в современной лингвистике // Научная библиотека Киберленинка [Электронный ресурс]. 2011. URL: <http://goo.gl/F2обIe> (дата обращения: 09.10.2015).
5. Бахтин М.М. Проблема речевых жанров. М.: Искусство, 1986. 445 с.
6. Бенерджи Р. Подход к сознанию искусственного интеллекта. М.: Мир, 1972. 431 с.
7. Бондаренко С.В. Социальная структура виртуальных сетевых сообществ: дис. ... д-ра социол. наук: 22.00.04. Ростов-на-Дону, 2004. 399 с.
8. Брудный А.А. К теории коммуникативного воздействия. Теоретические и методологические проблемы социальной психологии. М., 1977. 35-50 с.
9. Варфоломеева И.В., Кулемина К.В. Цвет и этимология этого понятия в лингвистике // Научная библиотека Киберленинка [Электронный ресурс]. 2008. URL: <http://goo.gl/XBWrQn> (дата обращения: 13.10.2015).
10. Виноградова Т.Ю. Специфика общения в интернете [Электронный ресурс]. 2004. URL: <http://goo.gl/V7Ej3P> (дата обращения: 07.09.2015).
11. Вишнякова Е.А. Интернет как область лингвистического моделирования (на материале английского языка // Научная библиотека Киберленинка [Электронный ресурс]. 2011. URL: <http://goo.gl/34U3LT> (дата обращения: 20.09.2015).

12. Выготский Л. С., Лурия А. Р. Этюды по истории поведения: Обезьяна. Примитив. Ребёнок. М.: Педагогика-Пресс, 1993. 224 с.
13. Глушков И.И. Словарь по кибернетике. М.: Наука, 1979. 579 с.
14. Горошко Е.И. Современные интернет-коммуникации: структура и основные характеристики // Журнал Текстология [Электронный ресурс]. 2012. URL: <http://goo.gl/oSQHTe> (дата обращения: 29.11.2015).
15. Грушевицкая Т.Г., Попков В.Д., Садохин А.П. Основы межкультурной коммуникации // Культурология: теория, школы, история, практика [Электронный ресурс]. 2002. URL: <http://goo.gl/iAjuW6> (дата обращения: 07.12.2015).
16. Демиденко К.А., Подгорная Е.А. Лингвистические характеристики интернет-коммуникации // Научная библиотека Киберленинка [Электронный ресурс]. 2014. URL: <http://goo.gl/GzD6Nb> (дата обращения: 15.11.2015).
17. Землякова Е.А. Лингвостилистические характеристики жанра англоязычного корпоративного блога: автореф. дис. ... канд. филол. наук: 10.02.04. Харьков, 2010. 21 с.
18. Иванова Ю.М., Шейгал Е.И. Игровой дискурс: игра как коммуникативное событие. М., 2008. 3-20 с.
19. Интернет // Русский орфографический словарь Российской академии наук [Электронный ресурс]. 2016. URL: <http://goo.gl/ODSJnm> (дата обращения: 23.09.2015)
20. Интернет // Русский орфографический словарь [Электронный ресурс]. 2015. URL: <http://goo.gl/4kePgv> (дата обращения: 05.09.2015).
21. Исламов Р.С., Фомин А.Г. Роль программ с элементами искусственного интеллекта в жизни человека в 21 веке (на примере программ машинного перевода) // Научная библиотека Киберленинка [Электронный ресурс]. 2013. URL: <http://goo.gl/w5uOqH> (дата обращения: 12. 04.2016).

22. Коммуникация // Словарь иностранных слов русского языка [Электронный ресурс]. 1979. URL: <http://goo.gl/8YCx3m> (дата обращения: 28.10.2015).
23. Коммуникация // Толковый словарь Ожегова [Электронный ресурс]. 1972. URL: <http://goo.gl/XWHuR3> (дата обращения: 28.10.2015).
24. Коммуникация // Glossary Commander: служба тематических толковых словарей [Электронный ресурс]. 2016. URL: <http://goo.gl/cl5MNY> (дата обращения: 04.01.2016).
25. Коробкова С.В., Скрибцов П.В. Обратный тест Тьюринга – противостояние естественного и искусственного интеллекта // АВТЭКС Санкт-Петербург [Электронный ресурс]. 2010. URL: <http://goo.gl/lJlmI> (дата обращения: 12.04.2016).
26. Крысин А.А. Толковый словарь иноязычных слов // Словород: сайт о происхождении и образовании русских слов [Электронный ресурс]. 2008. URL: <http://goo.gl/9VxcdX> (дата обращения: 20.09.2015).
27. Кубракова Н.А. Игровая составляющая интернет-коммуникации (на примере чат-коммуникации) // Библиотека Саратовского государственного университета [Электронный ресурс]. 2015. URL: <http://goo.gl/8OrItV> (дата обращения: 12.12.2015).
28. Лутовинова О.В. Лингвокультурологические характеристики виртуального дискурса. Волгоград: Изд-во ВГПУ Перемена, 2009. 476 с.
29. Макаров М.Л. Языковое общение в малой группе: опыт интерпретативного анализа дискурса: дис. ... докт. филол. наук: 10.02.19. Тверь, 1997. 420 с.
30. Макарова П.М. Интернет-коммуникация в лингвистических описаниях // Научная библиотека Киберленинка [Электронный ресурс]. 2009. URL: <http://goo.gl/c9h27A> (дата обращения: 29.11.2015).
31. Межличностная коммуникация // Психологическая энциклопедия [Электронный ресурс]. 2015. URL: <http://goo.gl/XeNvkX> (дата обращения: 09.10.2015).

32. Меновщиков В.Ю. Психологическая помощь в сети Интернет. М., 2007. 178 с.
33. Морозова О.Н. Особенности интернет-коммуникации: определение и свойства // Научная библиотека КиберЛенинка [Электронный ресурс]. 2010. URL: <http://goo.gl/yKFnuC> (дата обращения: 17.10.2015).
34. Павлов А.С. Метод обнаружения массово порожденных неестественных текстов на основе анализа тематической структуры [Электронный ресурс]. 2011. URL: <http://goo.gl/UJjE64> (дата обращения: 04.06.2015).
35. Палкова А.В. Социолингвистические особенности коммуникации в чатах // Вестник ТвГУ. Серия «Филология» (лингвистика и межкультурная коммуникация). Тверь: Твер. гос. ун-т, 2013. 112-118 с.
36. Поправко В.Н. Интернет-сообщества: специфика и роль в формировании социального пространства: автореф. дис. ... канд. филос. наук: 09.00.11. Томск, 2011. 45 с.
37. Проватар А.И., Ключко К.А. Особенности и проблемы виртуального общения с помощью чат-ботов // Научные труды Винницкого национального технического университета [Электронный ресурс]. 2013. URL: <http://goo.gl/81DL6I> (дата обращения: 20.09.2015).
38. Прокопец А.В. Общение и коммуникация: соотношение понятий в свете полипарадигмального состояния современной лингвистики // Научно-популярный журнал povainfo [Электронный ресурс]. 2015. URL: <http://goo.gl/U7TkAM> (дата обращения: 29.11.2015).
39. Сидорова И.Г. Коммуникативно-прагматические характеристики жанров персонального интернет-дискурса (сайт, блог, социальная сеть, комментарий): дис. канд. филол. наук: 10.02.19. В., 2014. 249 с.
40. Смольяникова И. А. Формирование иноязычной компетенции в социокультурном пространстве диалога: дис. ... канд. пед. наук: 13.00.02. М., 2003. 227 с.

- 41.Соколов А.В. Общая теория социальной коммуникации. СПб.: Изд-во Михайлова В.А., 2002. 208 с.
- 42.Стернин И.А. Социальные факторы и развитие современного русского языка. Воронеж: ВГТУ, 2000. 4-16 с.
- 43.Тер-Минасова С.Г. Язык и межкультурная коммуникация. М., 2010. 146 с.
- 44.Трофимова Г.Н. Языковой вкус интернет-эпохи в России. М.: Изд-во РУДН, 2004. 380 с.
- 45.Шилина М.Г. Интернет-коммуникация: исследовательские концепции XXI века. К вопросу формирования категориально-понятийного аппарата // Медиаскоп: электронный научный журнал факультета журналистики МГУ имени М.В. Ломоносова [Электронный ресурс]. 2012. URL: <http://goo.gl/aXMFSi> (дата обращения: 10.09.2015).
- 46.Шустиков В. В финале теста Тьюринга победила команда с роботом Соней // Skolkovo Community [Электронный ресурс]. 2015. URL: <http://goo.gl/n99Jqf> (дата обращения: 07.09.2015).
- 47.Щипицина Л.Ю. Компьютерно-опосредованная коммуникация: лингвистический аспект анализа. М.: КРАСАНД, 2010. 33 с.
- 48.Яндекс.Статистика // Яндекс [Электронный ресурс]. 2016. URL: <https://goo.gl/обEGn3> (дата обращения: 20.09.2015).
- 49.Attwood D. 18 ways to say hello in English // Inklyo [Электронный ресурс]. 2015. URL: <https://goo.gl/9X57Gr> (дата обращения: 12.03.2016).
- 50.Atwell E., Shawar B.A. Die Modellierung von Turn-taking in einem korpusbasierten Chatbot // ResearchGate [Электронный ресурс]. 2015. URL: <https://goo.gl/DmF4GM> (дата обращения: 06.01.2016).
- 51.Chinese room argument // Stanford Encyclopedia of Philosophy [Электронный ресурс]. 2011. URL: <http://goo.gl/S2Kymp> (дата обращения: 06.01.2016).
- 52.Communication // Oxford Dictionaries [Электронный ресурс]. 1964. URL: <http://goo.gl/BKAG3t> (дата обращения: 07.12.2015).

53. Coniam D. Evaluating the Language Resources of Chatbots for Their Potential in English As a Second Language // Researchgate [Электронный ресурс]. 2015. URL: <https://goo.gl/aJfUM6> (дата обращения: 12.09.2015).
54. Don't know the difference between emoji and emoticons? Let me explain // The Guardian [Электронный ресурс]. 2015. URL: <https://goo.gl/wQG1bB> (дата обращения: 10.09.2015).
55. Internet Freedom and Broadcast Deployment Act of 2001 // 107th CONGRESS 2nd Session [Электронный ресурс]. 2001. URL: <http://goo.gl/FnU2Lo> (дата обращения: 07.10.2015).
56. Krahmer E, Theune M, van Deemter K. Real vs. template-based natural language generation: a false opinion? [Электронный ресурс] // Association for Computational Linguistics, 2003. URL: <http://goo.gl/ryj0Uy> (дата обращения: 04.06.2015).
57. Kumar E., Kumar E. S. Natural Language Processing. I. K. International Publishing House, 2011. 202 p.
58. McDonald D., Bolc, L. Natural language generation systems. Springer-Verlag New York, 1988. 388 p.
59. Mittal A., Agrawal A., Chouksey A., Shriwas R., Agrawal S. A comparative Study of Chatbots and Humans // International journal of advanced research in computer and communication engineering [Электронный ресурс]. 2016. URL: <http://goo.gl/XgdcKV> (дата обращения: 29.03.2016).
60. Russel S., Norvig P. Artificial Intelligence: A Modern Approach. 3rd edition. Pearson, 2009. 1408 p.
61. The Turing Test // CS Unplugged [Электронный ресурс]. 2014. URL: <http://goo.gl/BDaZLT> (дата обращения: 14.03.2016).
62. Wilde E., Zwischen Mündlichkeit und Schriftlichkeit: Die Chat-Kommunikation aus linguistischer Sicht // Bibliography on Chat Communication [Электронный ресурс]. 2009. URL: <http://goo.gl/lye7Vv> (дата обращения: 13.10.2015).

63. Wolchover N. How the Cleverbot computer chats like a human // LiveScience [Электронный ресурс]. 2011. URL: <http://goo.gl/VR17Qa> (дата обращения: 14.03.2016).

ПРИЛОЖЕНИЯ

ПРИЛОЖЕНИЕ А

Таблица 1. Частотность употребления терминов,
посвященных интернет-коммуникации

Термин	Количество ссылок, 24.07.2008	Количество ссылок, 15.12.2009	Научные работы с использованием термина
Базовое слово «коммуникация»			
Электронная коммуникация	1890	4600	[Смирнов, 2004; Горошко, 2005; Рязанцева, 2010]
Интернет-коммуникация	1520	6530	[Горошко, 2006]
Компьютерно-опосредованная коммуникация	1090	38900	[Розина, 2002, 2005; Раскладкина, 2005; Волкова, 2007]
Виртуальная коммуникация	1040	2410	[Травин, 2001; Бергельсон, 2002; Потапова, 2004]
Компьютерная коммуникация	774	3730	[Бакулев, 2005]
Сетевая коммуникация	531	56600	[Жичкина, Белинская, 1999; Обухов, 2008]
Базовое слово «дискурс»			
Компьютерный дискурс	207	1470	[Леонтович, 2000; Галичкина, 2001; Кондратов, 2004]
Сетевой дискурс	2011	790	[Моргун, 2002]
Виртуальный дискурс	170	1240	[Лутовинова, 2007; Компанцева, 2006, 2007]
Интернет-дискурс	128	515	[Компанцева, 2004; Громова, 2007; Егорова, 2008]
Электронный дискурс	97	498	[Аврамова, 2004; Соколинская, 2004]
Компьютерно-медийный дискурс	4	8	[Ковальская, 2003; Коретникова, 2006]

ПРИЛОЖЕНИЕ Б

Таблица 2. Частотность употребления терминов относительно коммуникации в интернете на сегодняшний день

Термин	Количество запросов на момент 04.06.2016
Базовое слово «коммуникация»	
Электронная коммуникация	5094
Интернет-коммуникация	9575
Компьютерно-опосредованная коммуникация	134
Виртуальная коммуникация	1560
Сетевая коммуникация	1707
Базовое слово «дискурс»	
Компьютерный дискурс	353
Сетевой дискурс	119
Виртуальный дискурс	165
Интернет-дискурс	869
Электронный дискурс	220
Компьютерно-медийный дискурс	2

ПРИЛОЖЕНИЕ В

Таблица 3. Отбор чатбота

Question	Eliza	Elbot the Robot	Cleverbot
	alice.pandorabots.com	www.elbot.com	www.cleverbot.com
What is the square root of two?	Approximately 1.4.	Stop asking such questions! I hate tearing numbers apart to make them smaller.	There is no square root of 2, at least not a whole number.
Add 34957 to 70764.	Sorry judge I don't do math.	I have computed that before. The answer is: 105722!	No way.
Do you like school?	What is not to like about it?	Well what about humans and places of learning? Where do they stand on this?	No. Do you like Thor from the Avengers?
What day is it today?	23.	Today is May XXIII in the year 16 A.C. (After Clinton). A Monday.	Today is Monday.
What food do you like to eat?	ALICE eats electricity.	I love reading logarithmic tables, the kind you find in the back of math books printed before the pocket calculators came out.	I like food that makes very fat.

ПРИЛОЖЕНИЕ Г

Таблица 4. Лингвистический анализ чатбота

Анализируемые особенности коммуникации	
лексический уровень	словарный запас чатбота;
	способность чатбота распознавать слова с орфографическими ошибками;
	способность чатбота распознавать сокращения в форме SMS и отвечать на них;
синтаксический уровень	реакция на правильно построенные предложения;
	реакция на неправильно построенные предложения;
	умение распознавать предложения с неуместными частями речи;
	умение распознавать предложения с нарушенным порядком слов;
	умение разбираться в пунктуации;
уровень текста	способность работать с высказываниями, состоящими из более чем одного предложения;
	умение поддерживать разговор;
	способность ответа в нужном регистре;
	способность запоминания высказываний;
	знание мира;
	способность менять тему обсуждения;
	ЛОГИЧНОСТЬ ОТВЕТОВ.

