

Министерство образования и науки Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Пермский национальный исследовательский политехнический  
университет»

Факультет Гуманитарный  
Направление 45.03.02 «Лингвистика»  
Кафедра «Иностранные языки, лингвистика, перевод»

Зав. кафедрой ИЯЛП  
\_\_\_\_\_ (\_\_\_\_\_)  
« \_\_\_ » \_\_\_\_\_ 20\_\_  
Г

## **Выпускная квалификационная работа**

**На тему**

**Формирование денотативной модели  
предметной области работа-эколога**

**Студент**

Вяткин Денис Сергеевич (\_\_\_\_\_)

**Состав ВКР:**

Основная часть: стр.

Научный руководитель –  
канд.филол.н., доцент  
Соболева О.В. \_\_\_\_\_

Пермь 2018

## **Оглавление**

ВВЕДЕНИЕ.....	3
1 ПРИМЕНЕНИЕ МЕТОДА ДЕНОТАТИВНОГО АНАЛИЗА ДЛЯ ВЕРИФИКАЦИИ ОПИСАНИЯ ПРЕДМЕТНОЙ ОБЛАСТИ.....	7
1.1 Внутрязыковой перевод как отдельный вид перевода....	7
1.2 Роль понимания в процессе перевода.....	10
1.3 Понятие денотата и методика построения денотатного графа.....	14
1.4 Автоматизированные системы для работы с текстом....	19
Выводы по главе 1:.....	25
2 ДЕНОТАТНЫЙ ГРАФ КАК ЛИНГВИСТИЧЕСКИЙ ИНСТРУМЕНТ ОБЕСПЕЧЕНИЯ РОБОТОТЕХНИЧЕСКИХ ИССЛЕДОВАНИЙ.....	26
2.1 Построение и верификация денотатного графа.....	26
2.2 Автоматизированное создание денотатного графа.....	34
Выводы по главе 2:.....	39
ЗАКЛЮЧЕНИЕ.....	40
СПИСОК ЛИТЕРАТУРЫ.....	42

## ВВЕДЕНИЕ

Данная выпускная квалификационная работа посвящена особенностям внутриязыкового перевода, в основе которого лежит денотативная модель отображения содержания текста, которая в дальнейшем может быть использована для лингвистического обеспечения робототехнических исследований.

Ни для кого не секрет, что в последнее время исследования в области робототехники привлекают все больший интерес. Наряду с программистами и инженерами лингвисты также играют одну из ключевых ролей в создании систем искусственного интеллекта.

Хотелось бы отметить, что идеи о создании робототехники встречаются и в художественной литературе. Для нас интересной представляется мысль, высказанная в научно-фантастической пьесе Карла Чапека «R.U.R.»: чтобы роботы начали функционировать, их необходимо обучить: *«Понимаете, нам приходится оставлять немного места для естественного развития. <...> Это приблизительно то же самое, что у людей школа. Они учатся говорить, писать и считать. <...> Вы можете прочитать им двадцать томов Научного словаря, и они повторят вам все подряд, наизусть. Но ничего нового они никогда не выдумают»* [Чапек].

Мысль об обучении роботов достаточно проста и понятна, однако в действительности ее реализация требует больших усилий. Так, для того чтобы система смогла прочитать текст, его нужно формализовать, то есть перевести текст с естественного языка на язык формализованный, понятный системе.

Таким образом, считаем очевидной **актуальность** нашего исследования, в котором рассматривается один из способов перевода текстов с естественного языка на язык формализованный, который способна понять система. Кроме того, актуальность обусловлена и связью с общелингвистическими теоретическими проблемами, такими как текст и закономерности его восприятия и понимания, семантика языковых единиц текста, выделение основного содержания и его последующая экспликация.

**Объектом** исследования является модель денотативного анализа текста как один из способов формализации текста. В качестве **предмета** исследования выступает механизм построения денотатного графа.

**Цель** нашего исследования заключается в оценке применимости денотативного анализа текста для описания предметной области.

Поставленная цель предполагает решение ряда конкретных *задач*:

- изучить процессы восприятия и понимания текста;
- изучить понятие «денотат», «денотатный граф»;
- изучить модель построения денотатного графа;
- сформировать корпус текстов по теме «Твердые бытовые отходы»;

- формализовать отобранные тексты с помощью методики денотативного анализа;
- провести верификацию построенного денотатного графа.

Для исследования были использованы следующие **методы**:

1. анализ теоретических работ по семантике текста и способам ее формализации;
2. метод сплошной выборки для нахождения текстов по теме «Твердые бытовые отходы»;
3. моделирование денотатного графа на основе отобранных текстов;
4. эксперимент для верификации построенных графов.

**Материалом исследования** послужила лекция о твердых бытовых отходах и способах их утилизации, а также научные статьи по данной тематике.

**Научная новизна** работы состоит в том, что в ней метод денотативного анализа текста рассматривается как один из этапов на пути создания систем искусственного интеллекта.

**Методологическую базу** составляют фундаментальные исследования по проблеме понимания и смысла текста (Н.И. Жинкин, А.И. Новиков, А.А. Смирнов, Л.П. Добраев), а также труды, посвященные семантике текста и способам ее формализации (А.И. Новиков, Н.М. Нестерова).

**Гипотеза** исследования состоит в том, что текст, представленный в виде денотатного графа, может быть использован в процессе обучения системы.

**Теоретическая значимость** работы заключается в исследовании процесса отображения семантического

содержания текста с помощью денотатного графа, позволяющего создать адекватный вторичный текст.

**Практическая значимость** обусловлена тем, что модель денотативного анализа текста может быть использована в процессе обучения системы, то есть способствовать развитию исследований по проблемам искусственного интеллекта.

**Апробация работы.** Результаты исследования были представлены на научно-практической конференции «Перевод, реклама, PR в современной межкультурной коммуникации» (г. Пермь, 2017), а также в рамках Первого межвузовского научного семинара «Актуальные проблемы лингвистического обеспечения робототехнических исследований». Практическая часть исследования выполнялась при поддержке кафедры ИТАС ПНИПУ в рамках гранта РФФИ «Исследование и разработка модели управления автономного роботизированного комплекса, предназначенного для эффективного и безопасного мониторинга техногенных образований» (номер проекта 17-47-590128). В ходе работы над проектом по теме исследования опубликовано в соавторстве 3 научные статьи (2 в рецензируемом научном издании, включенном в реестр ВАК РФ, 1 в издании, входящем в РИНЦ; на данный момент все статьи находятся на стадии выхода из печати).

# **1 ПРИМЕНЕНИЕ МЕТОДА ДЕНОТАТИВНОГО АНАЛИЗА ДЛЯ ВЕРИФИКАЦИИ ОПИСАНИЯ ПРЕДМЕТНОЙ ОБЛАСТИ**

## **1.1 Внутрязыковой перевод как отдельный вид перевода**

Как известно, перевод – достаточно широкое понятие, которое по-разному трактуется исследователями. В нашей работе мы обращаемся к так называемому внутриязыковому переводу. В статье «О лингвистических аспектах перевода» Р. О. Jakobson выделяет три способа интерпретации вербального знака: он может быть переведен в другие знаки того же языка, на другой язык, или же в другую, невербальную систему символов [Jakobson 1978, с. 16-24]. Этим трем видам перевода в переводоведении соответствуют следующие определения:

1) внутриязыковой перевод, или переименование – интерпретация вербальных знаков с помощью других знаков того же языка;

2) межъязыковой перевод, или собственно перевод, – интерпретация вербальных знаков посредством какого-либо другого языка;

3) межсемиотический перевод, или трансмутация, – интерпретация вербальных знаков посредством невербальных знаковых систем.

Таким образом, под внутриязыковым переводом традиционно понимается истолкование словесных знаков посредством знаков того же языка [Jakobson 1978, с. 19]. Вслед за Р. О. Jakobsonом мы трактуем понятие «перевод» в широком смысле: **перевод** – мыслительная операция, направленная на порождение вторичного текста.

Представляется интересным, что о внутриязыковом переводе упоминает еще Ф. Шлейермахер в работе «О разных методах перевода» (F. Schleiermacher „Über die verschiedenen Methoden des Übersetzens“): „Die Thatsache, daß eine Rede aus

einer Sprache in die andere übertragen wird, kommt uns unter den mannigfaltigsten Gestalten überall entgegen. Wenn auf der einen Seite dadurch Menschen in Berührung kommen können, welche ursprünglich vielleicht um den Durchmesser der Erde von einander entfernt sind; wenn in eine Sprache aufgenommen werden können die Erzeugnisse einer andern schon seit vielen Jahrhunderten erstorbenen: so dürfen wir auf der andern Seite nicht einmal über das Gebiet Einer Sprache hinausgehen, um dieselbe Erscheinung anzutreffen. Denn nicht nur daß die Mundarten verschiedener Stämme eines Volkes und die verschiedenen Entwicklungen derselben Sprache oder Mundart in verschiedenen Jahrhunderten schon in einem engeren Sinne verschiedene Sprachen sind, und nicht selten einer vollständigen Dolmetschung unter einander bedürfen; selbst Zeitgenossen, nicht durch die Mundart getrennte, nur aus verschiedenen Volksklassen, welche durch den Umgang wenig verbunden in ihrer Bildung weit auseinander gehen, können sich öfters nur durch eine ähnliche Vermittlung verstehen“ [Schleiermacher 1838, с. 207]. Приведем собственный перевод основных положений данной мысли: «С одной стороны, перевод необходим для общения людей, живущих в разных концах Земли <...>; с другой стороны, с таким же положением вещей мы сталкиваемся в пределах одного языка. Ведь в посредничестве переводчика нуждаются не только племена одного народа, использующие разные изводы одного языка, <...> но и современники, не разделенные диалектами, а лишь происходящие из разных классов, получившие разное образование и не связанные с постоянным общением». Действительно, внутриязыковой перевод окружает нас

повсюду, хотя иногда мы этого даже не замечаем. Так, объясняя какой-нибудь физический закон, учитель превращает язык научного стиля в более доступный язык, понятный школьникам. Очевидно, если происходит общение среди ученых, обладающих данным категориальным аппаратом, такой перевод не понадобится. Таким образом, перевод внутри своего языка точно так же, как и перевод на иностранный язык, имеет свою коммуникативную задачу и средства ее достижения.

Хотелось бы отметить, что в теории перевода долгое время существовала лингвистическая парадигма, то есть в центре внимания исследователей были два текста. Однако на сегодняшний день интересной представляется коммуникативная теория перевода, согласно которой два субъекта (автор и читатель) становятся определяющим фактором для понимания перевода. В связи с этим обратимся к определению перевода, данному В. И. Провоторовым. Ученый трактует перевод следующим образом:

- языковая интерпретация (анализ слов, предложений, грамматических форм);
- стилистическая интерпретация (метафоры, рифмы);
- историческая интерпретация (знание эпохи, культуры автора);
- субъективно-психологическая интерпретация (воображение переводчика, перевоплощение в автора) [Провоторов 2001, с. 92].

Конечно, при внутриязыковом переводе в основном будет преобладать языковая интерпретация.

На рис. 1 представлен предложенный нами процесс внутриязыкового перевода:

Рис. 1. Процесс внутриязыкового перевода



Рассмотрим подробнее участников данного *процесса*:

1. **Автор:** лицо, имеющее определенную интенцию, которая реализуется в исходном тексте.
2. **Реципиент:** в нашем случае это робот, проходящий процесс обучения, направленный на формирование знаний об определенной предметной области. Материалом для обучения служит текст перевода – конечный результат деятельности переводчика, представляющий собой денотатный граф. Текст перевода создается для определенной цели. В нашем случае цель перевода – обучение робота.
3. **Переводчик:** посредник в человеко-машинной коммуникации. Переводчик осмысляет исходный текст и подвергает его денотативному анализу с целью создания

денотатного графа – формы представления текста, понятной для работа.

Таким образом, представляется интересным рассмотреть более подробно деятельность переводчика, а именно процесс понимания текста и непосредственно процесс внутриязыкового перевода, в основе которого лежит создание денотатного графа. Именно этим вопросам будут посвящены следующие параграфы главы 1.

## **1.2 Роль понимания в процессе перевода**

Как известно, А.Н. Крюков, создавая герменевтическую модель перевода, назвал понимание основным законом перевода. С мнением ученого невозможно не согласиться. Действительно, чтобы передать текст на другой язык или выразить ту же мысль средствами исходного языка (внутриязыковой перевод) необходимо осознать содержание исходного материала. Недаром специалист в области теории перевода В.Н. Комиссаров подчеркивал, что «переводчик вынужден понимать переводимый текст более глубоко, чем это обычно делает “нормальный читатель”» [Комиссаров 2001, с. 156]. Ведь в этом и заключается миссия переводчика – предотвращение «коммуникативного страдания», которое возникает вследствие полного или частичного непонимания [Chesterman]. Рассмотрим процесс понимания более детально.

Прежде всего, хотелось бы отметить ученых, занимающихся данной проблемой. К их числу, несомненно, относятся А.А. Смирнов [1948], А.Н. Соколов [1947], Л.П.

Доблаев [1982], Н.И. Жинкин [1976, 1982], А.А. Леонтьев [1976], А.И. Новиков [1983, 1999, 2000, 2003, 2007], В.В. Красных [1988], Н.П. Пешкова [2002], А.А. Залевская [1988, 2001] и другие. Среди зарубежных исследователей можно отметить Chomsky, Van Dijk, Kintsch [1978, 1983], F. Smith [1979], Mayer [1984], M. Chun [1997].

Так что же такое понимание? В научных трудах понимание рассматривается как особый аналитико-синтетический процесс, включающий в себя выделение смысловых вех – «пунктов понимания» – и объединение их в единое целое. Следовательно, понимание как «процесс мышления, направленный на разрешение стоящих перед личностью задач», подразумевает как анализ информации, так и синтез, причем последнее является основным условием для формирования целостного образа содержания первоисточника [Костюк 1950].

Проблемой понимания занимался и Н.И. Жинкин, который считал понимание письменного сообщения специфическим процессом, заключающимся в выделении ключевой мысли из текстового сообщения, которое в дальнейшем компрессируется во внутренней речи читателя до «комплекса смыслов». В результате у реципиента формируется своеобразная модель текста, представляющая собой сжатое до «тематических смысловых точек» тематическое образование [Жинкин, 1976].

А.И. Новиков трактует понимание как «опосредованный аналитико-синтетический процесс, базирующийся на активной интеллектуальной переработке воспринимаемого текста» [Новиков 1983, с. 36]. Согласно ученому, данная

переработка заключается в «членении текста на смысловые отрезки, выделение различного рода контекстных объединений, а также объединение их в общих смысл» [Новиков 1983, с. 36]. Таким образом, А. И. Новиков приходит к выводу о том, что основным механизмом понимания служит механизм внутренней речи, а основная закономерность понимания заключается в свертывании информации [Новиков 1983, с. 36].

Н.А. Герте, изучая вопрос реферирования текста, определяет понимание следующим образом: это «сложный аналитико-синтетический процесс смыслового свертывания, который происходит за счет способности ключевых элементов отображать основное содержание текста оригинала» [Герте 2014, с. 191].

А.А. Смирнов в процессе изучения данной проблемы разработал 7 уровней понимания текста, которые различаются глубиной:

1. Отнесение познаваемого предмета к самой общей категории.
2. Отнесение предмета или явления к общей категории хорошо известных предметов.
3. Вычленение специфических особенностей в познаваемом предмете на основе сравнения.
4. Переход на основе анализа, синтеза, абстрагирования и конкретизации, обобщения и индивидуализации от восприятия в целом к осмыслению отдельных частей предмета или явления и к пониманию их взаимосвязи.

5. Установление причинно-следственных связей с помощью индукции и дедукции.

6. Установление логических связей между осмысливаемыми явлениями.

7. Степень действий людей и выяснение мотивов этих действий [Смирнов 1948].

Интересной представляется теория понимания текста, предложенная Kintsch and Van Dijk. Согласно ученым, реципиент воспринимает текст на трех уровнях: дословный, семантический и ситуативный: «readers of a text build three different mental representations of the text: a verbatim representation of the text, a semantic representation that describes the meaning of the text and a situational representation of the situation to which the text refers» [Kintsch & Van Dijk, с. 365]. Хотелось бы отметить, что высказанная исследователями идея о ситуативном восприятии текста тесно связана с отображением денотативного содержания текстового сообщения, что в дальнейшем будет рассмотрено в данной работе.

Следует отметить, что в процессе понимания мы значительно перестраиваем воспринимаемую нами информацию. Так, согласно А.А. Смирнову, существуют следующие формы отхождения от подлинника:

- Обобщение или сгущение того, что дано в подлиннике в конкретной развернутой форме.
- Конкретизация и детализация того, что дано в тексте в более общем и сжатом виде.

- Замена одного содержания другим, равнозначным по смыслу, а также по степени общности и детализации [Смирнов 1966].

Таким образом, в процессе понимания целого текста могут быть использованы все три формы отхождения от подлинника.

В.В. Красных соотносит этапы понимания с уровнями его значения:

1) поверхностное значение – понимание в этом случае ограничивается лишь восприятием внешней формы текста;

2) глубинное значение – в процессе восприятия текста учитывается не прямое значение, заложенное в тексте;

3) смысл – понимание достигается за счет осознания концепта, общего смысла всего речевого сообщения [Красных 1998].

Таким образом, можно заметить, что проблема понимания рассматривается во взаимосвязи с понятием «смысл». Согласно А.И. Новикову, эти два понятия «оказываются комплементарными, то есть одно предполагает другое» [Новиков 2003].

Н.И. Жинкин подчеркивал, что «смысл – это то, что отражает наличную действительность» [Жинкин 1982, с. 131].

Таким образом, можно сказать, что в процессе внутриязыкового перевода при порождении вторичного текста смысл определяет выбор и задает отбор языковых средств. При этом языковые средства отбираются таким образом, чтобы их «декодирование приводило к формированию конкретного и целостного образа содержания» [Новиков 2007, с. 29].

В заключение хотелось бы отметить, что до сих пор ведутся дискуссии по поводу того, каким образом представлен воспринимаемый нами текст в сознании. Мы вслед за Н.И. Жинкиным и А.И. Новиковым считаем, что продукт понимания существует в интеллекте человека в виде денотатов. Рассмотрению данного понятия будет посвящен следующий параграф.

### **1.3 Понятие денотата и методика построения денотатного графа**

Что же такое денотат? Н.И. Жинкин под денотатом понимает «определенную вещь, явление, предмет», отмечая при этом, что это не понятие и не отношение, потому что последнее у него ничто иное, как принадлежность смысла [Жинкин 1956, с. 183]. А.П. Бобырева рассматривает денотат как «предмет объективного мира, обозначаемый словом» [Бобырева 1985, с. 31]. Однако Г.М. Стрелковский и Л.К. Латышев подчеркивают, что денотат – «это не просто реальный предмет действительности, а его отражение в мозгу человека» [Стрелковский 1980, с. 20].

Данный термин встречается и в работе немецкого логика, математика и философа Готлоба Фреге «О смысле и значении» [Über Sinn und Bedeutung]. Ученый предложил свою модель функционирования знака, получившую название «треугольник Фреге», вершинами которого является знак, значение и концепт (die Form, die Bedeutung, der Sinn). Согласно Frege, „es liegt nun nahe, mit einem Zeichen (Namen, Wortverbindung, Schriftzeichen) außer dem Bezeichneten, was

die Bedeutung des Zeichens heißen möge, noch das verbunden zu denken, was ich den Sinn des Zeichens nennen möchte, worin die Art des Gegebenseins enthalten ist“ [Frege 1892, с. 26] – Знак (будь то слово, словосочетание или графический символ) состоит из двух сторон, между которыми практически не существует границы. Первая сторона – обозначаемое, то есть денотат, вторая – смысл (*перевод наш*).

В толковом словаре немецкого языка DUDEN можно найти следующее определение: 1) vom Sprecher bezeichneter Gegenstand oder Sachverhalt in der außersprachlichen Wirklichkeit 2) begrifflicher Inhalt eines sprachlichen Zeichens im Gegensatz zu den emotionalen Nebenbedeutungen [Duden] – 1) обозначаемый говорящим предмет или явление внеязыковой действительности 2) абстрактное содержание языкового знака, противопоставленное дополнительному эмоциональному значению (*перевод наш*). Webster Dictionary определяет денотат как реально существующий объект, определяемый словом, знаком или лингвистическим выражением, противопоставляется десигнату (означаемому) – an actually existing object referred to by a word, sign, or linguistic expression —contrasted with *designatum* [Webster Dictionary].

Вопрос денотата и денотативного значения встречается в трудах многих ученых, в частности, английский философ, экономист Дж.С. Милль в своей работе «Система логики силлогистической и индуктивной», выдвигая идею «соозначения», употребляет термин денотат (denotation – означение), которому соответствует понятие «объем содержания» в логике [Милль, 1914].

Рудольф Карнап, немецко-американский философ, разработал свою систему терминов, предлагая понятия «экстенционал» и «смысл» или «интенционал» [Карнап, 2007].

Проблема феномена денотата рассматривается также в работах известного американского логика и математика Алонзо Чёрча. Ученый утверждал, что «денотат есть функция смысла имени, то есть если дан смысл, то этим определяется существование и единственность денотата» [Чёрч, 1960].

На наш взгляд, достаточно полное определение денотата сформулировал А.И. Новиков. На эту точку зрения мы и опираемся в нашей работе. Согласно ученому, денотат – «отраженные в мышлении и выраженные в тексте соответствующими языковыми средствами объекты и явления реальной действительности» [Новиков 1983, с. 26]. Важной является мысль Новикова о том, что «отдельное слово не имеет своего денотата и приобретает его в момент употребления и обязательно во взаимодействии с другими элементами языка при активном участие интеллекта и мыслительных операций» [Новиков 1983, с. 26].

Таким образом, можно сделать вывод, что содержание текста отображается в интеллекте человека в виде денотатов, связанных между собой предметными отношениями. Однако следует отметить, что ученые выдвигали разные мнение о том, как структурировать данное содержание материальными средствами.

Так, Н.И. Жинкин считает, «всякая речь может быть сведена к системе предикатов, которые, последовательно

дополняя друг друга, раскрывают состав и соотношение признаков неизвестного ранее предмета действительности» [Жинкин 1956, с. 148]. При этом предикаты не нанизываются друг за другом в одну однородную цепь, а образуют некоторую иерархию, где одни предметные признаки являются главными, другие – дополнительными, а третьи – дополнительные к этим дополнительным [Жинкин 1956]. В.Д. Тункель строит систему разпорядковых предикатов, где одни предикаты являются главными – предикаты первого порядка, другие дополнительные – предикаты третьего порядка, третьи дополнительные ко вторым и т.д. [Тункель 1964]. Л.П. Доблаев считает, что наиболее общим видом смысловой связи в тексте является связь между субъектом и предикатом. Он вводит понятия текстового субъекта и предиката. Эти элементы отличаются от логического содержания тем, что логические субъект и предикат представляют собой понятия, а текстовый предмет и предикат – логические суждения. Построение структуры текста, по мнению ученого, заключается в выделении этих текстовых субъектов и предикатов и установление их соотношения по всему тексту [Доблаев 1969]. Согласно В.Б. Апухтину, в качестве основной единицы содержательного анализа текста выступает предикатема, которую составляет тема и рема. Считается, что эта минимальная предикативная структура порождает смысл, под которым понимается нечто третье, возникающее из взаимодействия этих двух элементов. Предикатема является промежуточным этапом в построении смысловой структуры, поскольку последняя обычно представляет собой несколько взаимосвязанных предикатем.

В этом процессе реализуется принцип влияния и слияния смыслов взаимосвязанных предикатом, в результате чего смысл исходной предикатемы может трансформироваться как в тему, так и рему предикатемы следующего ранга [Апухтин 1978].

На наш взгляд, наиболее подходящим способом экспликации денотатной структуры является **денотатный граф** – «свернутое эксплицитное отображение структуры содержания текста, которой могут соответствовать различные языковые формы» [Новиков 1991, с. 61]. При этом вершинами графа являются имена денотатов, полученные в результате содержательного анализа текста и применения необходимых знаний о данном фрагменте действительности, а ребрами – предметные отношения между этими денотатами [Новиков 1983, с. 131].

Рассмотрим методику построения денотатного графа, предложенную А.И. Новиковым в 80-е годы XX века:

1. Выделение «ключевых» элементов текста, то есть наиболее важных, существенных элементов для понимания, причем не на уровне отдельных слов, а на уровне денотатов, являющихся единицами содержания.
2. Выделение подтем. На данном этапе задача заключается в том, чтобы определить элементы, принадлежащие к верхним уровням текста, т.е. выделить главный предмет описания и его подтемы. В тексте, как правило, содержится несколько подтем.
3. Определение субподтем, которые раскрывают содержание подтем, и их связи, как правило, не выходят за уровень этих подтем. На данном этапе построения денотатной структуры

удобно считать, что все субпотемы непосредственно подчинены своим подтемам и образуют один уровень иерархии.

4. Графическое представление иерархии подтем и субподтем. На данном этапе таблица связей преобразуется в граф, имеющий вид иерархического дерева, где вершине первого уровня соответствует имя главного предмета, вершинам второго уровня – имена подтем, а третий уровень соответствует субподтемам. Ребрам в таком графе соответствует наличие определенной связи между вершинами, причем конкретный вид связи здесь может быть и не выражен.
5. Определение соотношений денотатов. Задачей данного этапа является приведение полученной иерархической структуры функциональных элементов в соответствие с моделью ситуации, формировавшейся в интеллекте в результате понимания текста. Это достигается за счет экспликации предметных отношений, существующих между денотатами, что приводит к уточнению соотношения элементов полученной структуры, детализации ее уровней, к общему ее преобразованию [Новиков 1983, с. 147-151].

Таким образом, модель построения денотатного графа можно представить следующим образом (см. рис. 2):

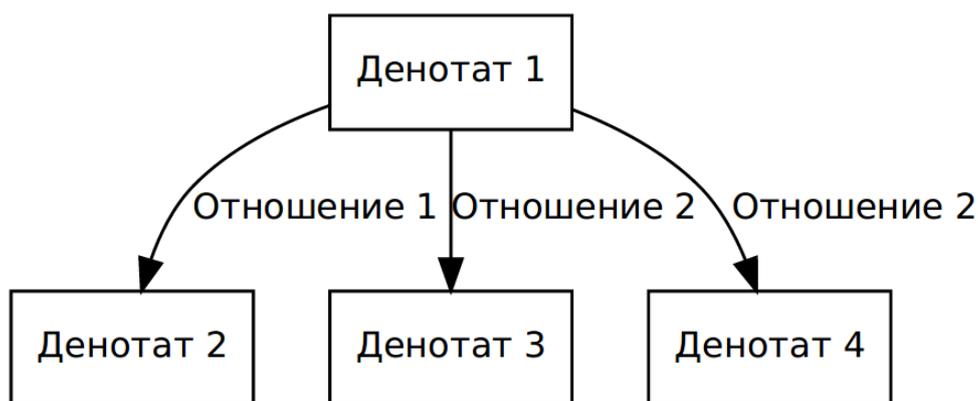


Рис. 2. Модель построения денотатного графа

Таким образом, нельзя не согласиться с Н.А. Герте, которая считает, что графовое представление материала вызывает полную перестройку текста: линейный текст преобразуется в иерархическую структуру, которая является моделью ситуации и отражает логику предметов и их отношений. Денотатный граф позволяет визуальнo продемонстрировать тематические уровни исходного текста – основную тему, подтемы и субподтемы. При этом устраняется языковая избыточность, свойственная первичному тексту [Герте 2016, с. 8-9].

#### **1.4 Автоматизированные системы для работы с текстом**

В последнее время стремительное развитие получил новый раздел прикладной лингвистики – компьютерная лингвистика. Компьютерная лингвистика трактуется как «область использования компьютерных инструментов – программ, технологий организации и обработки данных – для моделирования функционирования языка в тех или иных условиях, а также сферу применения компьютерных моделей

языка в лингвистике и смежных с ней дисциплинах» [Баранов 2007, с. 13]. В поле зрения исследователей данной отрасли попадают разные проблемы, к числу которых можно отнести машинный перевод, автоматический анализ и синтез текстов, создание и поддержку автоматических словарей, создание лингвистических баз данных, создание автоматических систем обучения языку и т.д.

В огромном объеме сегодняшней информации мы не можем обходиться без использования новых технологий. Так, для обработки огромного массива информации разрабатываются системы автоматического аннотирования и реферирования.

В работе Jie Tang, Duo Zhang, Limin Yao, Yi Li «Automatic Semantic Annotation Using Machine Learning» [Автоматическое семантическое аннотирование с использованием методов машинного обучения] говорится о том, что “semantic annotation has become an important sub-discipline of artificial intelligence, language processing, text mining, and Semantic Web” [Jie Tang 2012, с. 2] – «Семантическое аннотирование представляет собой отдельное направление в развитии искусственного интеллекта, лингвистической обработки, интеллектуального анализа текстов, а также семантической паутины» (*перевод наш*).

Нельзя не согласиться с исследователями, отмечающими, что “manual annotation is expensive and cannot be used to deal with the large volume of the existing documents on the Web. Automatic semantic annotation is an ideal solution to the problem” [Jie Tang 2012, с. 4] – «ручное аннотирование

- дорогостоящая и трудоемкая задача. Оно не предназначено для работы с большим объемом информации в Интернете. Именно поэтому семантическое автоматическое аннотирование является идеальным решением проблемы» *(перевод наш)*.

Ученые рассматривают различные методы автоматического аннотирования, включающие supervised machine learning based method, unsupervised machine learning based method, and ontology based method (обучение с учителем, обучение без учителя и онтологический подход). Особое внимание в работе уделяется такому важному аспекту в семантическом аннотировании, как извлечению отношений. Целью данного метода является “finding semantic relations between entities. With this technology, we can integrate large databases of relational information and generate new information for data mining, question answering systems, and information retrieval” [Jie Tang 2012, с. 19] – «Данный метод направлен на поиск смысловых отношений между сущностями. Благодаря этому мы можем интегрировать реляционные базы данных и создавать новую информацию для интеллектуального анализа данных, вопросно-ответных систем и систем по извлечению информации» *(перевод наш)*.

Создан ряд доступных систем, предназначенных для семантического аннотирования: Protégé -2000 (Эрикссон, Фергюсон, Шахар, и Мусен, 1999), WebKB (Мартин и Эклунд, 1999), SHOE (Хефлин и Хендлер, 2000), Annotea (Каган и Койвунен, 2001), Ontobroker (Фенсел, Декер, Эрдман, и Штудер, 1998) и SEAN (Мукерджи, Ян, а Рамакришнан, 2003). Следует отметить, что большинство систем автоматического

аннотирования рассчитаны на работу с определенным стилем текстов.

Хотелось бы также отметить, что ученые подчеркивают, “the fully automatic creation of semantic annotations is also an unsolved problem” [Jie Tang 2012, с. 3]- «цель полного автоматического аннотирования еще не удалось достичь» (*перевод наш*).

Наряду с автоматическим аннотированием активно исследуются вопросы, связанные с автоматическим созданием реферативного перевода.

Первые исследования по автоматическому реферированию появились около 1960 года. К числу первых работ, несомненно, можно отнести научные труды Ханс Петер Луна [Luhn, 1959] и Гербет Эдмундсона [Edmundson, 1969]. Ученые создали первую систему реферирования и предложили простые методы для составления автоматического реферата. Они использовали метод разбиения текста на предложения, которое оценивается по определенным признакам: Например, KEYPHRAZE, «расположение предложения», «корреляция с названием», «позитивные/негативные ключевые слова», «частотность слова»

Следует отметить, что в основе данного автоматического реферирования лежит «механический» алгоритм, не учитывающий структурные особенности текста.

Следующим шагом в развитии автоматического реферирования стало применение синтаксического анализа (парсинга), представляющего собой процесс сопоставления линейной последовательности лексем естественного или

формального языка с его формальной грамматикой [Глухих 2011]. Результатом такого анализа является RST-дерево (или дерево разбора), применение которого при автоматическом реферировании было предложено Д. Марку [Marcu, 1997]. Синтаксическое дерево отражает структуру данных, в которую преобразуется текст в процессе анализа.

Проблема автоматического реферирования нашла свое отражение и в диссертационной работе Н.А. Герте «Денотативная модель реферативного специализированного перевода». Несмотря на то, что большая часть исследования посвящена изучению возможностей создания реферативного перевода при помощи денотативной модели анализа текстов и скопос-теории, в одном из параграфов рассматривается проблема автоматизированного реферативного перевода: «В рамках выполнения работ по созданию системы автоматизированного реферативного перевода участниками проекта были разработаны метод, модель и алгоритм построения денотативной модели текста на основе его внешней формы и модели предметной области». [Герте 2017, с. 127].

Авторы выдвинули идею о том, что «для процесса компьютерного понимания текста достаточно разработать систему формальных признаков, которые позволяют распознавать ключевые, семантически активные слова, неоднословные номинации и высказывания рассматриваемого текста. <...> Исходя из такого подхода, мы можем выделить два способа, подходящих для создания программы автоматического реферирования: формальный и денотативный» [Герте 2015а, с. 81]. Для нас больший интерес

представляет денотативный подход, поэтому рассмотрим его более детально.

Данный подход основана на модели понимания текста А.И. Новикова и на методе денотативного анализа текста, разработанного научной школой Н.И. Жинкина и А.И. Новикова: «Реферативный перевод в рамках такого подхода можно представить как поэтапный процесс  $T_1 - ДС - T_2$ . Первый этап состоит в переходе от внешней языковой структуры текста к структуре его содержания, которая представлена в виде денотатной структуры. Второй этап заключается в переходе от выделенной денотатной структуры к новой внешней форме вторичного текста» [Герте, 2017, с. 132].

Как справедливо отмечает исследователь, «денотативная модель, используемая для создания системы автоматического реферирования, предполагает формализованное представление конкретной предметной области, но не отдельного текста. Предметная область получается в процессе семантического анализа корпуса текстов, которые по тематике относятся к выбранной области. Следовательно, целью обработки данного корпуса является выявление базовых денотатов и их отношений, которые лягут в основу модели предметной области, и последующее их введение в машину» [Герте 2017, с. 133].

Как можно заметить, ключевым шагом на пути решения проблемы автоматического реферирования является создание качественной предметной области. При этом, как отмечает Н.А. Герте, «главным этапом станет создание

самообучающейся системы и обработка корпуса текстов для ее обучения» [Герте, 2015, с. 83].

В заключение хотелось бы отметить, что исследования в области компьютерной лингвистики помогают и работе переводчика (например, посредством создания систем по извлечению терминов). К числу бесплатных ресурсов можно отнести Tilde [42] и Translated LABS [43]. Разумеется, существуют и коммерческие версии. В качестве примера можно рассмотреть систему Synchronterm, в основу работы которой положены статистический алгоритм для нахождения терминов в исходном тексте, а также статистический, синтаксический и морфологический алгоритмы для перевода данного термина в случае межъязыкового извлечения (*перевод наш*)- statistical algorithms for identifying candidate terms in a source text, and then applies statistical, syntactical and morphological algorithms to find possible equivalents for the source terms, in case of bilingual extraction. [Terminology Coordination]. Однако хотелось бы отметить, что необходимо значительное усовершенствование данных систем для качественного извлечения терминов. В настоящий момент к числу терминов системы зачастую приписывают и вспомогательные глаголы, а также имена собственные. Но, несмотря на эти погрешности, работа с большим объемом текста, не используя системы автоматического извлечения терминологии, была бы значительна усложнена.

Безусловно, проводимое нами исследование, направленное на обучение системы, также может способствовать развитию и усовершенствованию систем

автоматического аннотирования и реферирования, а также автоматического извлечения терминов.

В главе 2 будет продемонстрирован процесс создания и верификации денотатного графа, а также предложен универсальный алгоритм для описания любой предметной области.

### **Выводы по главе 1:**

1. Перевод – достаточно широкое понятие, под которым понимается не только межъязыковая, но и внутриязыковая коммуникация. В рамках нашей работы мы рассматриваем перевод внутри одного языка, так называемый внутриязыковой перевод.
2. В процессе перевода ключевую роль играет механизм понимания. Рассмотрев данную дефиницию у ряда зарубежных и отечественных ученых, в качестве рабочего мы выбрали определение, сформулированное А.И. Новиковым: «опосредованный аналитико-синтетический процесс, базирующийся на активной интеллектуальной переработке воспринимаемого текста».
3. Продукт понимания материализуется в форме денотатного графа, представляющего, в свою очередь, инструмент внутриязыкового перевода. В след за научной школой А.И. Новикова мы определяем денотатный граф как «свернутое эксплицитное

отображение структуры содержания текста, которой могут соответствовать различные языковые формы».

4. Нашу жизнь невозможно представить без новых технологий. На сегодняшний день существуют различные автоматизированные системы, помогающие осуществлять аннотирование и реферирование текстов, извлекать термины и т.д. Безусловно, это значительно упрощает работу лингвистов и переводчиков. Совершенствованию данных автоматизированных систем может способствовать исследование, проводимое в рамках нашей работы.

## **2 ДЕНОТАТНЫЙ ГРАФ КАК ЛИНГВИСТИЧЕСКИЙ ИНСТРУМЕНТ ОБЕСПЕЧЕНИЯ РОБОТОТЕХНИЧЕСКИХ ИССЛЕДОВАНИЙ**

### **2.1 Построение и верификация денотатного графа**

Как отмечал А.А. Ухтомский, «мы можем воспринимать лишь то, к чему подготовлены наши доминанты» [Ухтомский 1973]. Действительно, если дать ребенку прочитать технический текст о процессе производства карбамида, вряд ли у него возникнет какой-то образ о том, как это происходит. Об этом не стоит забывать и при создании роботов. Их «мозг»

представляет собой белый лист, который необходимо заполнить.

На наш взгляд, в процессе обучения робота необходимо представлять содержание текста в форме денотатной структуры. После того как робот получит первоначальную информацию о предметной области (его доминанты будут подготовлены), он сможет сам понимать тексты, то есть представлять их содержание в форме денотатного графа.

В первой главе была подробно описана методика построения денотатного графа. Рассмотрим ее на материале нашего исследования.

В качестве материала мы выбрали лекцию по твердым бытовым отходам и способам их утилизации (количество знаков: 12 100, количество слов: 1542) [23], а также две научные статьи, посвященные данной проблеме.

Статья №1. Снижение экологической нагрузки при обращении с твердыми бытовыми отходами за счет использования горючих компонентов.

Количество печатных знаков – 12 857.

Количество слов – 1816 [3].

Статья №2. Чистая территория как инновация. Раздельный сбор твердых бытовых отходов в Перми – опыт и перспективы.

Количество печатных знаков – 21 418.

Количество слов – 2 743 [2].

Хотелось бы отметить, что выбор данной тематики неслучаен. Во-первых, 2017 год был объявлен в России годом экологии. Во-вторых, проблема утилизации твердых бытовых отходов является одной из насущных экологических проблем,

требующих незамедлительного решения. В связи с этим создание робота-эколога, способного разделять твердые бытовые отходы, могло бы стать одним из шагов на пути решения данной проблемы. Родившаяся на электротехническом факультете ПНИПУ идея создания робота-эколога в 2017 была поддержана Российским фондом фундаментальных исследований. Работа над проектом потребовала тесного сотрудничества специалистов по информационным технологиям и лингвистов. Наша выпускная квалификационная работа в своей практической части представляет собой обобщение некоторых предварительных итогов, полученных нами во время решения задач лингвистического сопровождения этого робототехнического проекта.

В ходе нашего исследования мы составили универсальный алгоритм для обучения системы, состоящий из ряда этапов:

1. Подбор материала.
2. Построение денотатного графа.
3. Верификация денотатного графа посредством теста, предложенного для решения человеку.
4. Устранение недочетов в построенном графе.
5. Представления графа в форме денотатных пар для ввода в систему.
6. Верификация денотатного графа посредством теста, который решает система.
7. Дополнение предметной области для полного обучения системы.

В данном алгоритме пункт 4 является факультативным, поскольку граф может быть построен и без ошибок. Что

касается пункта 7, предметная область не может быть просто так дополнена, необходимо разрабатывать методы самообучения системы, что напрямую связано с проблемой создания искусственного интеллекта и, как следствие, не может быть решено в полной степени в рамках данной работы.

Хотелось бы отметить, что для визуализации графа была использована система Graphviz - пакет утилит по автоматической визуализации графов, заданных в виде описания на языке DOT [34]. Одним из достоинств данного инструмента является возможность вывода графов в разные форматы. Ниже представлен фрагмент, построенного нами графа (см. рис. 3):

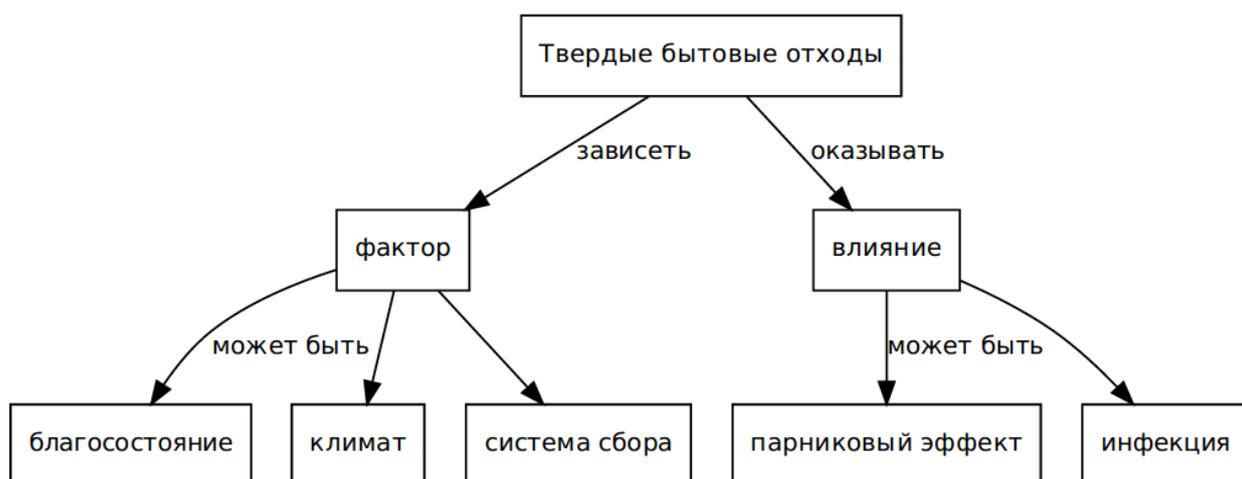


Рис. 3 Фрагмент денотатного графа

Как видно из представленного фрагмента графа (рис. 3), в иерархической структуре денотаты могут быть одновременно как главными, так и зависимыми. Например, денотат «влияние» является подчиненным по отношению к денотату «твердые бытовые отходы» и в то же время главным для таких денотатов, как «парниковый эффект» и «инфекция». Кроме того, денотат может вступать в разные

отношения с другими денотатами. Так, денотат «твердые бытовые отходы» связан с денотатом «фактор» посредством связи «зависеть», а с денотатом «влияние» – с помощью связи «оказывать».

На рис. 4 представлен другой фрагмент графа, который посвящен способам утилизации твердых бытовых отходов:

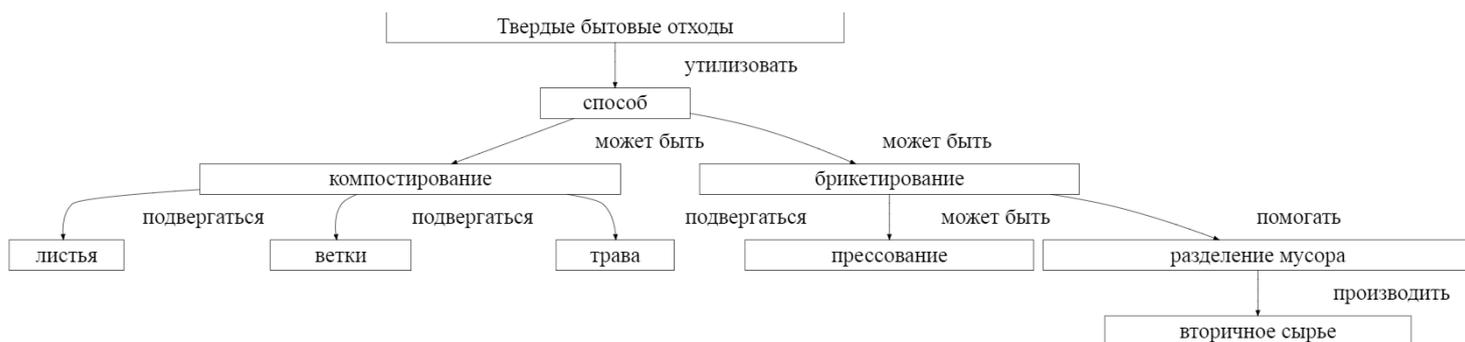


Рис. 4 Фрагмент денотатного графа

Данный фрагмент в очередной раз свидетельствует о том, что один и тот же денотат может быть как главным, так и зависимым. Например, денотат «компостирование» является зависимым по отношению к денотату «способ» и в то же время - главным для таких денотатов, как «листья», «ветки» и «трава».

Для того чтобы проверить, является ли денотатная структура понятной реципиентам и главным образом системе, необходимо верифицировать денотатный граф. Под верификацией понимается «проверка истинности теоретических положений, установление достоверности опытным путем» [40]. Для этого нами был разработан тест из 10 вопросов с 4 вариантами ответа, один из которых являлся правильным. Испытуемым предлагалось решить данный тест, опираясь на наш граф. В верификации графа приняло участие 19 студентов 4 курса кафедры иностранных языков,

лингвистики и перевода. В целом тест был решен успешно. Наибольшую сложность вызвал последний вопрос, поскольку ответ не следовал явно из денотатного графа. На рис. 5 представлен график, показывающий отклонение ответов от правильного варианта:

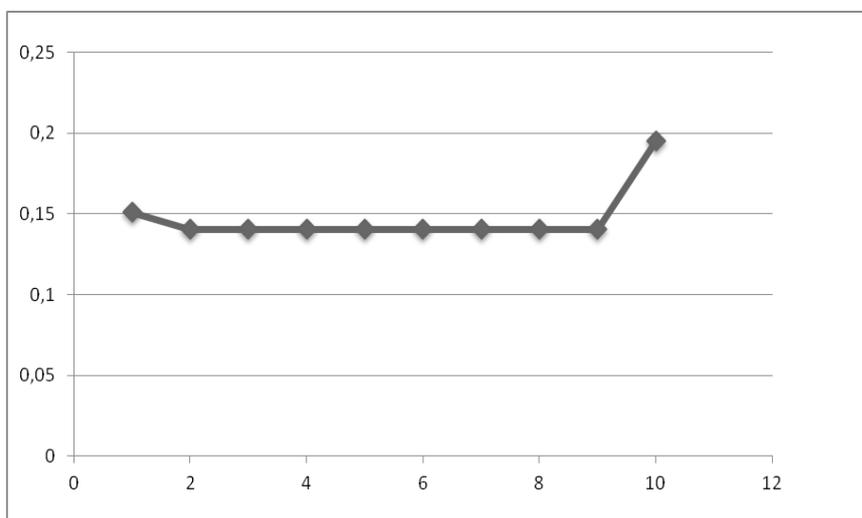


Рис. 5 Отклонение ответов от правильного варианта

На горизонтальной оси графика отображены номера вопросов, а на вертикальной – среднее абсолютное отклонение чисел от среднего значения, вычисленного с использования функции СРОТКЛ в Microsoft Excel.

В ходе решения теста была выявлена некорректность при составлении графа, а именно в следующих денотатных парах, представленных в таблице 2:

Таблица 2

Непонятные для испытуемых денотатные пары

Денотат	Отношение	Денотат
твердые бытовые отходы	Оказывать	Влияние
влияние	может быть	Ловушка

Как выяснилось, испытуемым было непонятно, что стоит за денотатом «ловушка». В связи с этим мы эксплицировали данный денотат следующим образом (см. табл. 3):

Таблица 3

Дополнение к предметной области

<b>Денотат</b>	<b>Отношение</b>	<b>Денотат</b>
ловушка	Являться	опасность
опасность	Для	животных

Следующим шагом нашего исследования было представление содержания лекции в виде денотатных пар: денотат – отношение – денотат и их ввод в систему для ее обучения. В таблице 4 представлены денотатные пары приведенных выше фрагментов графа:

Таблица 4

Описание предметной области в форме денотатных пар

<b>Денотат</b>	<b>Отношение</b>	<b>Денотат</b>
твердые бытовые отходы	Зависеть	фактор
твердые бытовые отходы	Оказывать	влияние
фактор	может быть	благосостояние
фактор	может быть	климат
фактор	может быть	система сбора
влияние	может быть	парниковый эффект
влияние	может быть	инфекция
твердые бытовые отходы	Утилизировать	способ
способ	может быть	компостирование
способ	может быть	брикетирование
компостирование	подвергаться	листья
компостирование	подвергаться	ветки
компостирование	подвергаться	трава
брикетирование	может быть	прессование
брикетирование	помогать	разделение мусора
разделение мусора	Производить	вторичное сырье

Далее приведена форма представления денотатных пар для ввода в систему:

```
{
  "denotat1" : "Твердые бытовые отходы",
  "relation" : "зависеть",
  "denotat2" : "фактор"
},
{
  "denotat1" : "фактор",
  "relation" : "может быть",
  "denotat2" : "благополучие"
},
{
  "denotat1" : "фактор",
  "relation" : "может быть",
  "denotat2" : "климат"
},
{
  "denotat1" : "фактор",
  "relation" : "может быть",
  "denotat2" : "благоустройство"
},
{
  "denotat1" : "фактор",
  "relation" : "может быть",
  "denotat2" : "система сбора"
},
{
  "denotat1" : "Твердые бытовые отходы",
  "relation" : "оказывать",
  "denotat2" : "влияние"
```

```

},
{
"denotat1" : "влияние",
"relation" : "может быть",
"denotat2" : "парниковый эффект"
},
{
"denotat1" : "влияние",
"relation" : "может быть",
"denotat2" : "инфекция"
},

```

После того как мы внесли соответствующие корректировки в денотатный граф и представили его в форме денотатных пар, система [29] должна была решить тест, аналогичный тому, который был предложен студентам.

Ниже представлен пример ввода теста в систему:

```

{
"К твердым бытовым отходам не относится:": {
  "а": "упаковочные материалы",
  "б": "пищевые отходы",
  "в": "железо",
  "г": "нефть"
},

```

Выяснилось, что система не может правильно ответить на два вопроса, одним из которых является вопрос о том, что не относится к твердым бытовым отходам. Из четырех вариантов ответа (железо, нефть, пищевые отходы, упаковочные материалы) система выбирает наугад либо железо, либо нефть. Это происходит вследствие того, что в

графе отсутствует денотат «железо», представлен только денотат «металл». Человек, имея фоновые знания, понимает, что железо относится к металлам и, следовательно, является твердым бытовым отходом. Однако система не знает этого, что и приводит к ошибке.

Ниже (см. рис. 6) приведен график, показывающий отклонение ответов испытуемых и системы от правильного варианта:

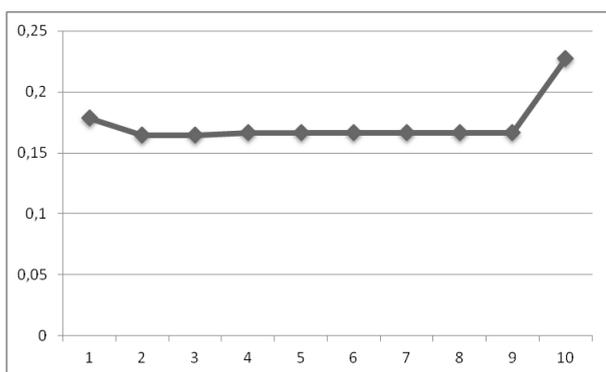


Рис. 6 Отклонение ответов испытуемых и системы от правильного варианта

Как видно из представленного графика, отклонение ответов от правильного значения больше по сравнению с графиком, представленном на рис. 4. Из этого следует, что количество правильных ответов системы меньше по сравнению с количеством правильных ответов, которые дали испытуемые.

Таким образом, можно сделать вывод, что для качественного обучения системы необходимо представлять ей полную информацию о предметной области.

Хотелось бы отметить, что при верификации третьего денотатного графа в тесте наряду с закрытыми мы использовали и открытые вопросы. Например, «назовите участников сбора твердых бытовых отходов: \_\_\_\_\_». Как и

предполагалось, человек без труда смог ответить на данные вопросы. Система, в свою очередь, пока не умеет качественно отвечать на открытые вопросы. Необходима определенная доработка со стороны программистов.

Таким образом, можно сделать вывод, что при предоставлении системе полной информации о предметной области она будет справляться с решением тестов с закрытыми вопросами и обучаться, что в дальнейшем может способствовать созданию метода самообучения.

## **2.2 Автоматизированное создание денотатного графа**

После того как мы представили системе предметную область, мы решили проверить, насколько она сможет самостоятельно построить денотатный граф по тексту, относящемуся к данной тематике.

Для эксперимента мы составили текст на основе трех денотатных графов, предложенных системе. Рассмотрим составленный нами текст:

«Твердые бытовые отходы являются результатом бытовой деятельности человека и представляют собой гетерогенную смесь, состоящую из разных компонентов. К числу твердых бытовых отходов можно отнести пищевые отходы, бумагу, пластмассу, текстиль, дерево, стекло, металлы, смет, кожу и резину. Хотелось бы отметить, что период разложения того или иного отхода различается. Так, дерево разлагается в течение 50 лет, а для разложения резины потребуется целых 1000 лет.

Если говорить о химическом составе твердых бытовых отходов, то необходимо отметить, что они состоят из таких

углеродосодержащих компонентов, как углерод, кислород, водород, азот и сера.

Ни для кого не секрет, что твердые бытовые отходы оказывают негативное влияние на окружающую среду. Например, они приводят к усилению парникового эффекта или загрязнению подводных вод. В связи с этим необходимо разрабатывать различные альтернативные способы утилизации отходов, а не просто складировать или сжигать мусор.

Одним из путей решения данной проблемы может стать практика раздельного сбора мусора, успешно практикующаяся в Европе. Несомненно, Россия также пытается внедрить раздельный сбор мусора. Однако приходится сталкиваться с рядом проблем, к числу которых можно отнести несовершенную законодательную базу, низкую культуру граждан и незначительную плату за вывоз мусора. Первым шагом в решении данной проблемы может стать создание аппаратов по приему тары, а также мобильных точек для сбора опасных отходов.

Твердые бытовые отходы можно использовать и в качестве альтернативных источников энергии. Горючую фракцию, входящую в состав твердых бытовых отходов, необходимо извлекать, тем самым экономить такие природные топливные ресурсы, как нефть, газ и уголь».

На наш взгляд, в приведенном тексте некоторые средства выражения могут стать камнем преткновения для системы (см. таблицу 5).

Таблица 5 Лингвистические трудности для системы

<b>Название проблемы</b>	<b>Пример</b>
--------------------------	---------------

Причастный оборот	<u>состоящую из разных</u> <u>компонентов;</u> <u>Горючую фракцию, входящую</u> <u>в состав твердых бытовых</u> <u>отходов</u> <u>Одним из путей решения</u> <u>данной проблемы может стать</u> <u>практика отдельного сбора</u> <u>мусора, успешно</u> <u>практикующаяся в Европе</u>
Сложные предложения	<u>Если говорить о химическом</u> <u>составе твердых бытовых</u> <u>отходов, ...</u>
Средства связи	<u>Хотелось бы отметить, что</u> <u>Например</u> <u>Несомненно</u> <u>Тем самым</u> <u>Ни для кого не секрет</u> <u>Первым шагом в решении</u> <u>данной проблемы</u> <u>Также предпринимаются</u> <u>шаги</u>

Поэтому мы упростили данный текст:

«Твердые бытовые отходы являются результатом бытовой деятельности человека и представляют собой гетерогенную смесь. Гетерогенная смесь состоит из разных компонентов. Компонентами являются пищевые отходы, бумага, пластмасса, текстиль, дерево, стекло, металлы, смет,

кожа и резина. Период разложения того или иного компонента различается. Так, дерево разлагается в течение 50 лет, а для разложения резины потребуется целых 1000 лет.

Твердые бытовые отходы состоят из углеродосодержащих компонентов. Углеродосодержащие компоненты - это углерод, кислород, водород, азот и сера.

Твердые бытовые отходы оказывают негативное влияние на окружающую среду. Например, они приводят к усилению парникового эффекта или загрязнению подводных вод. В связи с этим необходимо разрабатывать различные альтернативные способы утилизации отходов, а не просто складировать или сжигать мусор.

Необходимо отдельно собирать мусор, что делают в Европе. Россия также пытается внедрить отдельный сбор мусора. Однако приходится сталкиваться с рядом проблем. Проблемами являются несовершенная законодательная база, низкая культура граждан и незначительная плата за вывоз мусора. Первым шагом в решении данной проблемы может стать создание аппаратов по приему тары, а также мобильных точек для сбора опасных отходов.

Твердые бытовые отходы можно использовать и в качестве альтернативных источников энергии. Твердые бытовые отходы состоят из горючей фракции. Горючую фракцию необходимо извлекать. Это поможет сэкономить такие природные топливные ресурсы, как нефть, газ и уголь».

К сожалению, хотелось бы отметить, что построение системой денотатного графа еще требует технических доработок. Тем не менее, на наш взгляд, составленные нами

тексты идеально подходят для проведения эксперимента, направленного на выявление способности системы строить денотатные графы на основе изученной предметной области, что, безусловно, является одним из шагов на пути создания самообучающейся системы и, как следствие, искусственного интеллекта.

**Выводы по главе 2:**

1. Рассмотренную методику построение денотатного графа мы реализовали на текстах по теме «Твердые бытовые отходы». Выбор данной тематики обусловлен необходимостью более рационального и экологически безопасного обращения с твердыми бытовыми отходами в нашей стране.
2. Построенные денотатные графы были верифицированы. В процессе верификации сначала приняли участия студенты, а затем и сама система. Для верификации нами был составлен тест из 10 вопросов. В ходе исследования выяснилось, что на данном этапе система может решать тесты с закрытыми вопросами и правильно отвечать на все вопросы при предоставлении полной информации о предметной области.
3. На основе материала нашего исследования мы составили текст для проведения эксперимента, призванного установить, насколько качественно система сможет понять данный текст и построить по нему денотатный граф, руководствуясь знаниями о предметной области, полученными в процессе обучения.

## ЗАКЛЮЧЕНИЕ

Данная работа была посвящена изучению возможностей применения метода денотативного анализа текста для описания и верификации предметной области.

Ключевым теоретическим понятием, на которое мы опирались в нашем исследовании, было понятие денотата. Изучив его дефиниции у разных ученых, в качестве рабочего мы выбрали определение, сформулированное А.И. Новиковым. Под денотатом исследователь понимает «отраженные в мышлении и выраженные в тексте соответствующими языковыми средствами объекты и явления реальной действительности».

Цель работы заключалась в оценке применимости денотативного анализа текста для описания предметной области. Мы предложили универсальный алгоритм для описания любой предметной области. Для анализа была взята лекция по твердым бытовым отходам и способам их утилизации, а также две научные статьи, посвященные данной тематике. На основе материала исследования были построены денотатные графы, которые впоследствии были верифицированы и представлены в виде денотатных пар для ввода в систему. В ходе верификации выяснилось, что для правильного решения теста система нуждается в больших сведениях о предметной области по сравнению с человеком, имеющим фоновые знания. Кроме того, система пока не умеет решать тесты с открытыми вопросами, что требует определенных технических доработок.

В ходе проведения исследования наша гипотеза о том, что текст, представленный в виде денотатного графа, может быть использован в процессе обучения системы, несомненно, подтвердилась.

Таким образом, в нашем исследовании мы показали, что метод денотативного анализа текста, предложенный А.И. Новиковым, имеет большое прикладное значение в процессе формализации текста. В конце работы мы предложили созданные нами тексты, которые могут быть использованы для проведения эксперимента, проверяющего способность системы строить денотатные графы и, следовательно, адекватно понимать предложенные ей тексты. В дальнейшем нам представляется перспективным продолжать данное исследование и создать универсальный метод описания предметной области, благодаря которому система сама могла бы обучаться и запоминать сведения о произвольной предметной области.

### **СПИСОК ЛИТЕРАТУРЫ**

1. Апухтин В.Б. Психолингвистический метод анализа смысловой структуры текста. Канд. дис. М.: 1978. 178 с.
2. Андреева О. Ю. Чистая территория как инновация. Раздельный сбор твердых бытовых отходов в Перми - опыт и перспективы / О. Ю. Андреева, З. М. Кашафутдинова // Материалы 3-й Международной научно-практической конференции "Шумпетеровские чтения", г. Пермь, 22 ноября 2013 г. - Пермь: , Изд-во ПНИПУ, 2013. 100-104 с.
3. Армишева Г. Т. Снижение экологической нагрузки при обращении с твердыми бытовыми отходами за счет использования горючих компонентов / Г. Т. Армишева, В. Н. Коротаев, В. Г. Кривошеин // Научные исследования и инновации. - , 2010. - Т. 4, № 3. 3-8 с.
4. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. 3-е изд. М.: ЛКИ, 2007. 137 с.
5. Бобырева А.П. Об использовании денотатных графов при выделении смыслового содержания иноязычных текстов // Науч. тр. / МГПИИЯ им. М. Тореца. 1985. Вып. 207. 26-40 с.

6. Брандес М.П., Провоторов В.И. Предпереводческий анализ текста: учеб. пособие. – М.: 2001. 224 с.
7. Герте Н.А. Денотативная модель реферативного специализированного перевода: автореф. дис. ... кан. филол. наук: 10.02.19 / Герте Наталия Александровна. Пермь, 2016. 19 с.
8. Герте Н.А. Денотативная модель реферативного специализированного перевода : диссертация ... кандидата филологических наук : 10.02.19 / Герте Наталия Александровна. - Пермь, 2016. 254 с.
9. Глухих М.И. Синтаксический анализ, обратная польская запись. [Электронный ресурс]. – URL: [http://kspt.ftk.spbstu.ru/media/files/2011/course/cpp/slides/15\\_SyntaxAnalysis.pdf](http://kspt.ftk.spbstu.ru/media/files/2011/course/cpp/slides/15_SyntaxAnalysis.pdf) (дата обращения 01.05.2018)
10. Добраев Л.П. Логико-психологический анализ текста. Саратов, 1969. 171 с.
11. Жинкин Н.И. Развитие письменной речи учащихся III-VII кл. – Изв. АПН РСФСР, 1956, №78. 141-250 с.
12. Жинкин Н.И. Речь как проводник информации. – М.: Наука, 1982. 159 с.
13. Карнап Р. Значение и необходимость. Исследования по семантике и модальной логике. – М.: ЛКИ, 2007. 384 с.
14. Комиссаров В.Н. Современное переводоведение: курс лекций. – М.: ЭТС, 2001. 424 с.
15. Костюк Г.С. О психологии понимания: научные записки Института психологии УССР. – Киев, 1950. – Т. II. С. 53.
16. Милль Дж.С. Система логики силлогистической и индуктивной. – 2-е изд. – М.: 2014. 880 с.

17. Новиков А.И. Семантика текста и ее формализация. – М.: 1983. 216 с.
18. Новиков А.И., Нестерова Н.М. Реферативный перевод научно-технических текстов / А.И. Новиков, Н.М. Нестерова. М.: Ин-т языкознания АН СССР, 1991. 148 с.
19. Новиков А.И. Текст и «контртекст»: две стороны процесса понимания // Вопр. психолингвистики. – 2003. – № 1. 64–76 с.
20. Новиков А.И. Текст и его смысловые доминанты / под ред. Н.В. Васильевой, Н.М. Нестеровой, Н.П. Пешковой. – М.: Институт языкознания РАН, 2007. – 224 с.
21. Смирнов А.А. Проблемы психологии памяти. М.: 1966. 423 с.
22. Стрелковский Г.М., Латышев Л.К. Научно-технический перевод: Пособие для учителей немецкого языка. М.: 1980. 176 с.
23. Твердые бытовые отходы, их утилизация. [Электронный ресурс] // Studmed.ru URL: <http://www.studmed.ru/docs/document6176?view=1> (дата обращения: 03.11.2017).
24. Тункель В.Д. К вопросу о устной передаче речевого сообщения. Канд. дис. М.: 1964. 194 с.
25. Ухтомский А.А Письма//Новый мир. 1973. №1. 25-38 с.
26. Чапек К. [Электронный ресурс] // РУР URL: <http://lib.ru/SOCFANT/CHAPEK/rur.txt> (дата обращения 03.11.17).
27. Чёрч А. Введение в математическую логику. – М.: 1960. – Т. 1, 478 с.

28. Якобсон Р. О. О лингвистических аспектах перевода  
Вопросы теории перевода в зарубежной лингвистике:  
Сборник статей: Пер. с англ., нем., франц. Вступительная  
статья и общая ред. перевода В. Н. Комиссарова. М.: 1978.
29. Chesterman A. Proposal for a Hieronymic Oath // The  
Translator. - 2001. - №2. P. 139-154.
30. Denotat Test Solver [Электронный ресурс] // Решатель  
тестов на базе денотатной структуры // GitHub URL:  
[https://github.com/daniel-kurushin/test\\_solver](https://github.com/daniel-kurushin/test_solver) (дата обращения:  
27.11.2017).
31. DUDEN [Электронный ресурс] URL:  
<https://www.duden.de/rechtschreibung/Denotat> (дата  
обращения: 08.12.17).
32. Edmundson H.P. New methods in automatic extracting //  
Journal of the Association for Computing Machinery. - 1969. -  
Vol. 16(2). P. 264-285.
33. Frege Gottlob: Über Sinn und Bedeutung. In: Zeitschrift für  
Philosophie und philosophische Kritik, N. F., Bd. 100/1, 1892. S.  
25-50.
34. FriedrichSchleiermacher's sämtliche Werke. Dritte  
Abtheilung. Zur Philosophie. Zweiter Band (Berlin, 1838). P. 207-  
245.
35. Graphviz - Graph Visualization Software. [Электронный  
ресурс] URL: [graphviz.org](http://graphviz.org) (дата обращения: (01.04.2018)).
36. Jie Tang, Duo Zhang, Limin Yao, Yi Li. Automatic Semantic  
Annotation Using Machine Learning. IGI Global, 2012. P.44
37. Kintsch, W. & Van Dijk, T.A. Toward a model of text  
comprehension and production. Psychological Review, 85 (5),  
1978. P. 363-394.

38. Marcu D. The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts. Department of Computer Science. – Toronto: University of Toronto, 1997. P. 8.
39. Merriam-Webster [Электронный ресурс] URL: <https://www.merriam-webster.com/dictionary/denotata> (дата обращения: 08.12.17).
40. Murdock V.G. Aspects of Sentence Retrieval: PhD thesis. – Massachusetts: University of Massachusetts Amherst, 2006.
41. Slovari.ru [Электронный ресурс] URL: <http://www.slovari.ru/search.aspx?s=0&p=3068&di=vsis&wi=3133> (дата обращения: 23.11.2017).
42. Terminology Coordination. [Электронный ресурс] // DG Trad URL: <http://termcoord.eu/discover/free-term-extractors/term-extraction-tools> (дата обращения: 20.04.2018).
43. Tilde. [Электронный ресурс] // Terminology Services URL: <https://www.tilde.com/> (дата обращения: 20.04.2018).
44. Translated LABS. [Электронный ресурс] // Terminology Extraction URL: <https://labs.translated.net/terminology-extraction/> (дата обращения: 20.04.2018).