МИНОБРНАУКИ РОССИИ

**ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ**

**«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

Факультет *романо-германской филологии*

Кафедра *французской филологии*

Направление *«Лингвистика»*

Профиль *«Теория и методика преподавания иностранных языков и культур»*

Курсовая работа на тему:

**«Лингвистика: разделы и направления. ИТ в лингвистике»**

Выполнила: студентка 1 курса 1 группы Рябухина Елизавета Юрьевна

Руководитель: Донина Ольга Валерьевна

**Содержание**

**Введение3**

Основная часть2

1. Лингвистика: разделы и направления4
2. Информационные технологии в лингвистике8
3. Основные лингвистические программы и ресурсы, представленные в сети интернет9

3.1 Программы анализа и лингвистической обработки текстов9

* 1. Программы для автоматической обработки текстов10
  2. Программы преобразования текстов10

3.4 Психологические программы11

1. Использование некоторых лингвистических программ в исследовании русскоязычных и англоязычных искусствоведческих текстов13

**Заключение20**

Список литературы21

**Введение**

Становление современного информационного общества приводит к коренным изменениям во всех сферах жизни и деятельности человека. В сознании людей все больше утверждается мысль о том, что "будущий стратегический потенциал общества будут составлять не вещество и энергия, а информация и научные знания", что "мощь любого государства будет определяться не уровнем развития промышленности, новизной и эффективностью ее технической базы, а уровнем информатизации общества". В недалеком будущем реально защищенным в социальном плане может быть лишь только широкообразованный человек, способный гибко перестраивать направление и содержание своей деятельности в связи со сменой технологий или требований рынка.

В наши дни владение ИТ становится в один ряд с такими качествами, как умение читать и писать. Сегодня специалист с высшим образованием должен "свободно ориентироваться в мировом информационном пространстве, иметь необходимые знания и навыки поиска, обработки и хранения информации с использованием современных ИТ, компьютерных систем и сетей".

С вступлением общества в век компьютерных технологий появилась возможность более эффективной обработки и представления информации. Это позволило эффективно хранить и обрабатывать большие потоки информации. Но на современном этапе развития информационной культуры общества знания устаревают очень быстро, и человек вынужден учиться всю жизнь. Огромный объём знаний, накопленный человечеством, заставляет искать иные подходы к организации процесса обучения.

Как отмечают в своей статье "Филология и современные информационные технологии (К постановке проблемы)" К.В. Вигурский и И.А. Пильщиков, "осознание фундаментальной роли информации в общественном развитии и огромные темпы роста ИТ обусловили необходимость формирования особой информационной культуры личности. Для использования новых компьютерных технологий в жизни требуется новое мышление, которое должно воспитываться у ребёнка с младших классов. Развитие и широкое внедрение информационных технологий воздействует на все сферы современной жизни, включая экономику, политику, науку и образование".

**ИТ** – "это совокупность методов и программно-технических средств, объединенных в технологическую цепочку, обеспечивающею сбор, обработку, хранение, распределение и отображение информации с целью снижения трудоемкости процессов использования информационных ресурсов".

Сегодня использование ИТ значительно увеличивает эффективность функционирования любой сферы человеческой деятельности. Лингвистика не является исключением. Какими бы далекими областями, на первый взгляд, ни казались ИТ и лингвистика, неоспоримая связь между ними подтверждается хотя бы фактом существования **компьютерной лингвистики (КЛ** **)** – "направления в прикладной лингвистике, ориентированного на использование компьютерных инструментов – программ, компьютерных технологий организации и обработки данных – для моделирования функционирования языка в тех или иных условиях, ситуациях, проблемных сферах и т.д., а также всей сферы применения компьютерных моделей языка в лингвистике и смежных дисциплинах". Более того, если учесть, что главным предметом исследования в лингвистике являются тексты и что именно современные ИТ открывают новые возможности для обработки и анализа текстов и предоставляют разнообразные средства создания, распространения, поиска и учета текстовой информации, связь и взаимодействие этих двух сфер становятся еще более очевидными.

Сказанное выше объясняет **актуальность** темы данной работы. Ее **целью** является анализ возможностей и эффективности применения ИТ в лингвистике в целом и в исследовании автора, связанном с изучением искусствоведческих текстов на английском и русском языках, в частности. Для достижения поставленной цели необходимо решить следующие **задачи** .

· Охарактеризовать основные лингвистические программы и ресурсы, представленные в сети Интернет.

· Рассмотреть те программы и ресурсы, которые целесообразно использовать при исследовании искусствоведческих текстов.

Общая концепция работы состоит в том, чтобы, понять, каким образом и насколько эффективно можно использовать ИТ непосредственно при обработке текстов искусствоведческой проблематики, и выяснить, имеют ли полученные с помощью ИТ языковые данные научную ценность для магистерской диссертации.

**Основная часть**

1. ЛИНГВИСТИКА: РАЗДЕЛЫ И НАПРАВЛЕНИЯ

Существует достаточно большое число определений понятия лингвистика. В «Лингвистическом энциклопедическом словаре» (1990) лингвистика (языкознание, языковедение) определяется как «наука о естественном человеческом языке вообще и о всех языках мира как индивидуальных его представителях». Несколько конкретизируя приведенное выше определение, Ю.С.Маслов пишет, что она исследует сущность и природу языка, проблему его происхождения и общие законы его развития и функционирования. Более детально задачи лингвистики рассмотрены в упомянутом энциклопедическом словаре , а также в работе В.В.Звегинцева.

Существуют многочисленные попытки выделения внутри лингвистики отдельных ветвей или направлений четко отделяются друг от друга теоретическая и прикладная лингвистика. Общим для перечисленных и некоторых других работ этого периода (конец 60— начало70 годов ХХ века) является твердое убеждение, что эти два направления лингвистики взаимосвязаны и дополняют друг друга, что успешное функционирование прикладной лингвистики возможно лишь на базе лингвистических теорий, разработанных в рамках теоретической лингвистики.

Некоторые из существовавших в то время и новых зарождающихся подходов к анализу языковых явлений (машинный перевод, автоматическая обработка речевой информации, порождающая грамматика, дескриптивная лингвистика, математическая лингвистика и др.) почти единогласно относились авторами к проблемам прикладной лингвистики.

Несколько позже Р.Г.Пиотровский, четко разделяя теоретическую и прикладную лингвистику (но не признавая последнюю как самостоятельный раздел языкознания), выделяет структурное и математическое языкознание, а также «новую» лингвистику, к которой относит инженерную лингвистику. В последней он выделяет далее вычислительную лингвистику,

экспериментальную фонетику, лингвистическое обеспечение систем научно-технической информации .

Группа ученых Ленинградского университета, опираясь на три основных аспекта любой достаточно развитой области знания — теорию, эксперимент, практику, — и учитывая их диалектическую взаимосвязь в процессе познания, выделила в языкознании три взаимосвязанных направления: теоретическую лингвистику, экспериментальную лингвистику и прикладную лингвистику. При этом структурную лингвистику авторы считают симбиозом таких наук, как теоретическое языкознание, психология, логика, семиотика, математика, математическая лингвистика. С их точки зрения, это совокупность методов теоретической лингвистики и математики [22, 4]. В последние годы в рамках прикладной лингвистики выделяют также компьютерную лингвистику.

Как видно из вышесказанного, «прикладная лингвистика» — понятие, до сих пор не имеющее четкого определения и конкретного конечного перечня решаемых ею задач. Тем не менее попытки разобраться с этими проблемами были и есть.

Одно из первых определений понятия прикладная лингвистика принадлежит В.В.Звегинцеву. Он определил ее как новую область лингвистики, «которая осуществляет реализацию лингвистиче- ских знаний с целью решения всякого рода практических задач». Далее, несколько детализируя сказанное, автор отмечает: «Прикладная лингвистика представляет новый взгляд на задачи изучения языка. Исходя из этого нового взгляда, она производит переоценку достигнутого в науке о языке, направляет по определенному руслу лингвистические исследования и, конечно, комплектует собственную тематику».

В «Лингвистическом энциклопедическом словаре» прикладная лингвистика определяется как «направление в языкознании, занимающееся разработкой методов решения практических задач, связанных с использованием языка». При этом задачи прикладной лингвистики делятся на традиционные, или «вечные», «новые». К числу первых относят создание и совершенствование письменности, разработку систем транскрипции устной речи, систем транслитерации иноязычных слов, унификацию и стандартизацию научно-технической терминологии, создание словарей различных типов, перевод с языка на язык, обучение языку и т.д.. К новым задачам прикладной лингвистики относятся те, появление которых обусловлено современной научно-технической революцией, характеризующейся укреплением взаимосвязи общественных, естественных технических наук. Перечень таких задач достаточно широк. Наиболее удачным представляется определение прикладной лингвистики и перечня решаемых ею задач, сделанное А.Е.Кибриком. По мнению автора, «прикладная лингвистика — раздел языкознания, в котором разрабатываются методы решения практических задач, связанных с оптимизацией использования языка как важнейшего средства человеческой коммуникации». Множество задач, решаемых прикладной лингвистикой, А.Е.Кибрик выделяет с опорой на те функции языка, которые оптимизируются задачами прикладной лингвистики.

В частности, он выделяет следующие четыре функции:

1) оптимизация способов фиксации и хранения речевой информации;

2) оптимизация способов передачи информации;

3) оптимизация интеллектуальных способностей человека, связанных с использованием языка;

4) оптимизация использования языка как средства массовой коммуникации.

В рамках каждого из этих разделов автор выделяет конкретные задачи. Их также можно условно разделить на «вечные» и «новые» (ср. с. 6). К числу задач первого типа, отмечаемых А.Е.Кибриком, можно отнести создание алфавитов и письменностей, создание систем транскрипции и транслитерации, задачи сурдопедагогикизадачи использования языка в медицине, языковое планирование, языковое строительство, нормализацию языка и т.д.

«Новые» задачи — это такие проблемы, решение которых возможно с использованием современных информационных технологий.

1. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЛИНГВИСТИКЕ

Конкретизируя определение понятия «информационные технологии» по отношению к лингвистике, можно сказать, что информационные технологии в лингвистике — это совокупность законов, методов и средств получения, хранения, передачи, распространения, преобразования информации о языке и законах его функционирования с помощью компьютеров. Если соотнести это определение с теми задачами, которые решает современная прикладная лингвистика, то можно отметить, что понятие «информационные технологии» в лингвистике относится в основном к задачам прикладной лингвистики. К их числу можно отнести:

1) создание систем искусственного интеллекта;

2) создание систем автоматического перевода;

3) создание систем автоматического аннотирования и реферирования текстов;

4) создание систем порождения текстов;

5) создание систем обучения языку;

6) создание систем понимания устной речи;

7) создание систем генерации речи;

8) создание автоматизированных информационно-поисковых систем;

9) создание систем атрибуции и дешифровки анонимных и псевдо анонимных текстов;

10) разработка различных баз данных (словарей, карточек, каталогов, реестров и т.п.) для гуманитарных наук;

11) разработка различного типа автоматических словарей;

12) разработка систем передачи информации в сети Интернет и т.д.

Эти комплексные задачи включают целый ряд более мелких проблем. К их числу относится автоматизация следующих процессов:

1) построение словарей текстов;

2) морфологический анализ слова;

3) определение значения многозначного слова;

4) синтаксический анализ предложения;

5) поиск слова в словаре;

6) порождение предложения и т.д.

1. ОСНОВНЫЕ ЛИНГВИСТИЧЕСКИЕ ПРОГРАММЫ И РЕСУРСЫ, ПРЕДСТАВЛЕННЫЕ В СЕТИ ИНТЕРНЕТ

На сайте Русской виртуальной библиотеки представлен **Каталог лингвистических программ и ресурсов в Сети, составленный С.В. Логичевым** . Данный каталог включает в себя описание программ, связанных с анализом текстов и вычислительной лингвистикой, а также соответствующих ресурсов, доступных сегодня в глобальной сети Интернет. Упор при составлении каталога, по словам автора, делался на бесплатные программы, доступные для загрузки или использования в режиме on-line. Также описаны коммерческие версии некоторых наиболее интересных программ. Тематически каталог разбит на следующие разделы:

· программы анализа и лингвистической обработки текстов;

· программы преобразования текстов;

· психолингвистические программы;

· генераторы текстов;

· системы обработки естественного языка и машинного перевода;

· каталоги и коллекции ресурсов;

· словари и тезаурусы;

· поисковые машины и системы полнотекстового поиска;

· системы синтеза и распознавания речи.

Рассмотрим наиболее интересные и полезные, на наш взгляд, программы.

* 1. ПРОГРАММЫ АНАЛИЗА И ЛИНГВИСТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТОВ

**Link Grammar Parser** – это синтаксический парсер (то есть синтаксический анализатор) английского языка. Работает со словарем, включающим около 60000 словарных форм.

**Проекты Cibola** **/** **Oleada** реализуют обширные компьютерные системы лингвистического анализа текстов. Компоненты системы включают средства работы с мультиязыковыми текстами, построения конкорданса для текстов на более чем 16 языках, статистического анализа, автоматического перевода, различные словари и тезаурусы.

**Russian** **Morphological** **Dictionary** – программа С.Сикорского для синтаксического и морфологического анализа русскоязычных текстов. Использует морфологический словарь, включающий 120000 слов.

**Mystem** – компактный, быстрый и бесплатный морфологический парсер русскоязычных текстов, реализованный на основе словаря Зализняка.

**Лингвоанализатор** – on-line версия программы математического анализа структуры текста. Целью анализа является определение близости любого из предлагаемых пользователем текстов к одному из авторских эталонов, определенных заранее. Программа анализирует входной текст и выдает имена трех писателей, которые могли бы быть его наиболее вероятными авторами.

**Система StarLing** позволяет работать с мультиязычными текстами большой длины, с транскрипционными знаками. Среди преимуществ: удобный поиск, возможность анализа и синтеза словоформ по словарю Зализняка, а также перевода по словарю Мюллера. Есть функции для сравнительно-исторических исследований.

**Морфологический анализатор** – on-line версия программы морофлогического анализа слов русского/английского языков. Позволяет получить для вводимого слова базовую форму и морфологическую информацию. Программа реализована на основе словарей Зализняка (рус.яз.) и Мюллера (англ.яз.).

* 1. ПРОГРАММЫ ДЛЯ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТОВ

**АОТ** **(автоматическая обработка текста** ). Среди предлагаемых продуктов представляют интерес: модуль графематического анализа текста, компоненты морфологического анализа для русского, немецкого и английского языков, модуль автоматического уничтожения омонимии, модуль семантического анализа текста, система лингвистического поиска, различные тезаурусы и словники.

**Textarc** – революционная программа для визуализации и исследования текстов, являющаяся настоящим произведением искусства. Текст воспроизводится на экране компьютера в виде галактики, в которой слова играют роль звезд. Часто встречающиеся слова светятся ярко, а редкие – вовсе не видны. Строки текста выводятся в виде закручивающейся спирали по границе "галактики" точечным шрифтом (высотой в один пиксель). Программа позволяет видеть структуру текста, взаимосвязи между словами и контекстом, просматривать частотный и алфавитный индекс слов и конкорданс. При движении курсора по галактике слова вспыхивают и загораются лучами (указателями на контекст), а также звучат в разной тональности.

**LeoBilingua** – программа, позволяющая генерировать билингва-текст (текст из двух синхронных половин на разных языках).

**Инструментальные средства МедиаЛингва** предназначены для создания систем автоматического аннотирования, классифицирования, поиска и морфологической обработки текстовой информации.

* 1. ПРОГРАММЫ ПРЕОБРАЗОВАНИЯ ТЕКСТОВ

**Программы** **преобразования** **текстов** представлены такими разработками, как: **HTML Batch Editor** , **Словогрыз** , **ClearText** **Reader's** **Edition** , **xReplacer** , **xMarkup** **v2.1.1** ., **XML редакторы** , **Xchange Search&Replace** , **KDiff3** , **WinMerge** , **TextTransformer v1.4.1** . Особое внимание обратим лишь на некоторые из них.

Например, **Словогрыз** предназначен для поиска и замены текста по заданным маскам. Программа умеет находить и заменять в тексте не только строго заданные слова и фразы, но и текстовые элементы, определяемые произвольными поисковыми шаблонами. Позволяет сохранять/загружать сценарии преобразований.

**KDiff** **3** и **WinMerge** – программы, позволяющие визуально сравнивать текстовые файлы.

**TextTransformer** **v** **1.4.1** – мощная программа, предназначенная для выполнения сколь угодно сложных преобразований над множеством текстовых файлов. Позволяет анализировать, изменять, создавать и удалять заданные элементы обрабатываемых текстов.

* 1. ПСИХОЛОГИЧЕСКИЕ ПРОГРАММЫ

В разделе **психолингвистических программ** представлены:

**ПСИ-Офис версия 2.1** – система психолингвистического анализа текстов, включающая 3 компоненты: программа поиска вложенных слов в тексте, т.е. слов, "спрятанных" внутри и на переходах между словами; программа поиска повторяющихся фрагментов текста при анализе "автоматического письма" (такие тексты пишутся с целью анализа текущих подсознательных процессов); программа синтеза подсознательного компонента текста.

**ВААЛ-2000** – экспертная система лексического и контент-анализа текстов; прогнозирует эффект неосознаваемого воздействия текста на массовую аудиторию, анализирует тексты с точки зрения такого воздействия, выявляет личностно-психологические качества автора текста.

**Приемы журналистики & Public** **Relations** – программа, представляющая собой консультанта, помогающего пользователю при написании различных текстов. Пользователь выбирает ряд параметров, характеризующих желаемый результат, а программа выдает ему рекомендации по написанию текста и иллюстрирует их примерами.

**Psyberia** **.** **ru** и **Лаборатория ПСИТОН**содержат профессиональные психодиагностические и психолингвистические программы.

Среди многочисленных генераторов текстов отметим, например: **Болтун**(виртуальный собеседник, который обладает зачатками искусственного разума; может реагировать на реплику пользователя своей репликой), **Russian** **Word** **Constructor** **(** **RWC** **)** (программа для генерации русскоязычных стихоподобных текстов ("инструмент поэта"); способна конструировать русские неологизмы на основе заданного словаря с лексико-статистической информацией), **Весна** (генератор псевдофилософских текстов).

**Системы обработки естественного языка и машинного перевода**представлены в каталоге программами: **Natural** **Language** **Projects** **at** **ISI** , **Автоматический словарь Мультитран** , **Translate** **.** **Ru** , **LEO** , **PEREVODOV** **.NET** , **Проекты НИИ ИИ** , **Computer** **Aided** **Translation** , **Google** **Переводчик** .

Возможности использования **Translate** **.** **Ru** и **Google** **Переводчика**будут проиллюстрированы в Главе 3.

Каталоги и коллекции ресурсов включают в себя **Каталог программ по вычислительной лингвистике** , **Ресурсы, связанные с анализом текста** , **Справочно-информационный портал "Русский язык"** , **TextAnalysis** **Info** , **The** **Linguist** **List** **, LTI** **Projects** . **Портал "Русский язык",** например, интересен тем, что содержит массу полезной информации и включает on-line словари русского языка, такие как: Полный электронный орфографический словарь русского языка под ред. В.В. Лопатина; Словарь трудностей произношения и ударения в современном русском языке; Новый словарь русского языка, Толково-словообразовательный словарь, Словарь нарицательных имён и др.

Среди многочисленных **словарей и тезаурусов** стоит отметить представляющие для автора особый интерес (в связи с исследованием англоязычных искусствоведческих текстов) **British** **National** **Corpus** (коллекция более чем 100 миллионов слов современного английского письменного и устного языка), **Acronym Finder** (полный словарь акронимов и сокращений английского языка, содержит более 173 тысячи элементов), **WordNet** (электронный тезаурус, отражающий все возможные толкования слов английского языка и показывающий взаимосвязи между ними; для каждого слова можно найти синонимы, объединенные в смысловые группы и получить ссылки на родовые или производные понятия), **Толковый словарь Merriam** **Webster** (on-line версия знаменитого толкового словаря английского языка).

**Поисковые машины и системы полнотекстового поиска** представлены **Поисковой системой Я** **ndex** , **Поисковой системой Ищейка** , **Verity** **Ultraseek** , **Quintura** **Search** , **ARM** **Engine** **4.0**и многими другими.

Интересны **программы синтеза и распознавания речи,** например: **Sacrament Text-to Speech Engine v2.0**(система синтеза русской речи, созданная минской компанией "Сакрамент"; позволяет произносить тексты мужским и женским голосом, делая при этом интонационные паузы, изменяя по желанию пользователя тон и тембр речи), **Govorilka** (небольшая программа чтения текста голосом для Windows; настраивается на различные языки и голоса; использует редактируемые словари произношений; позволяет записать синтезированную речь в MP3-файл), **CSLU Toolkit** (набор инструментальных средств для создания приложений, обрабатывающих речевую и звуковую информацию; осуществляет синтез и распознавание речи, озвучивание текстов, обработку спектрограмм и звуковых сигналов) и другие.

Таким образом, мы убедились в существовании огромного множества разнообразных программ и ресурсов, которые могут значительно облегчить работу с лингвистическим материалом. Лингвисту остается только выбрать подходящие для исследования программы и воспользоваться их преимуществами.

1. ИСПОЛЬЗОВАНИЕ НЕКОТОРЫХ ЛИНГВИСТИЧЕСКИХ ПРОГРАММ В ИСЛЕДОВАНИИ РУССКОЯЗЫЧНЫХ И АНГЛОЯЗЫЧНЫХ ИСКУССТВОВЕДЧЕСКИХ ТЕКСТОВ

В данной главе мы попытаемся оценить возможности некоторых программ применительно к исследованию русскоязычных и англоязычных искусствоведческих текстов.

Рассмотрим программы анализа и лингвистической обработки текстов **TextAnalyst** **2.0**и**Худломер** и системы обработки естественного языка и машинного перевода **Translate** **.** **Ru** и **Google** **Переводчик** **.**

Сначала обратимся к программе **TextAnalyst** **2.0** (для использования необходимо скачивание) [11] Это смысловой анализатор, который за считанные минуты позволяет ознакомиться с текстами любой тематики. Программа разработана в качестве инструмента для анализа содержания текстов, смыслового поиска информации, формирования электронных архивов и предоставляет пользователю следующие основные возможности:

· анализ содержания текста с автоматическим формированием семантической сети с гиперссылками;

· получение смыслового портрета текста в терминах основных понятий и их смысловых связей;

· анализ содержания текста с автоматическим формированием тематического древа с гиперссылками;

· выявление семантической структуры текста в виде иерархии тем и подтем;

· смысловой поиск с учетом скрытых смысловых связей слов запроса со словами текста;

· автоматическое реферирование текста;

· автоматическая индексация текста с преобразованием в гипертекст;

· ранжирование всех видов информации о семантике текста по «степени значимости» с возможностью варьирования детальности ее исследования.

При исследовании искусствоведческих текстов многие из этих возможностей могут быть полезными. Для начала работы запускаем программу и выбираем текстовый файл для анализа. В нашем случае это будет файл "Поэзия земли.txt", содержащий статью о русском художнике Владимире Копылове. Теперь главное окно TextAnalyst должно выглядеть примерно так:

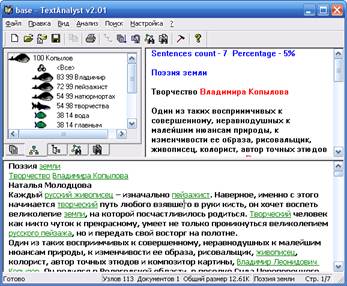


Рисунок 1 – Интерфейс программы TextAnalyst

Вся работа по анализу текста уже сделана, остается лишь ознакомиться с ее результатами. Прежде всего, изучив предложенный материал, TextAnalyst формирует сеть основных (наиболее значимых) понятий, содержащихся в представленных ему текстах (верхнее левое окно на Рис. 1). В нашем случае в сеть понятий входят: "Копылов", "Владимир", "пейзажист", "натюрмортах", "творчества", "вода", "главным", "живописца", "рисунок", "художественный" и др.

Такая сеть служит представлением смысла текста и основой для всех видов дальнейшего анализа. **Сеть понятий** – это множество терминов из текстов – слов и словосочетаний, связанных между собой по смыслу. В сеть включены не все термины текста, а лишь наиболее значимые, несущие основную смысловую нагрузку. Аналогичным образом представлены и смысловые связи между понятиями текстов – отражаются лишь наиболее явно выраженные из них. Поэтому, с одной стороны сеть достаточно полно описывает смысл текстов, а с другой – позволяет отбросить несущественную информацию и представить содержание в сжатом виде, так называемом “смысловом портрете”.

Таким образом, можно сразу увидеть всю информацию по каждому понятию, буквально бросив единственный взгляд на набор его связей в сети. В результате, передвигаясь по смысловым связям от понятия к понятию, можно находить и прицельно исследовать лишь интересующие места текстов, не затрудняя себя просмотром всей попавшейся на пути информации.

Теперь обратим внимание на числа, стоящие в дереве возле понятий. Ближайшее к понятию число представляет его **смысловой вес** (например, "Копылов" – 100, "Владимир" – 99, "творчества" – 98). Его значение (от 1 до 100) показывает, сколь важную роль играет понятие для смысла всего текста – как много информации в тексте касается данного понятия. Максимальное значение, равное 100, говорит о том, что понятие является ключевым и представляет важнейшую тему текста. Маленькое, близкое к единице значение показывает, что соответствующая тема лишь вскользь упомянута в тексте и в нем очень мало информации, относящейся к данному понятию. Второе число, стоящее перед смысловым весом, ближе к раскрытому узлу, представляет **вес связи** от понятия в вершине раскрытого списка к данному. Вес связей также всегда принимает значение от 1 до 100.

Кроме того, программа предоставляет услугу автоматического **реферирования** (в меню "Анализ" выбираем пункт "Реферирование"). Формируемый реферат содержит список наиболее информативных предложений текстов (тезисов) (Рисунок 2, правое верхнее окно). Конечно, это еще не полноценный реферат, так как тезисы в основном не связаны между собой стилистически, а просто выбраны из текста и расположены в порядке их встречаемости. Однако и такой подстрочник реферата оказывается достаточно информативным, чтобы составить общее представление о тексте и уяснить его основные мысли. Более того, все предложения реферата снабжены отсылками к соответствующим местам исходных текстов, что позволяет просмотреть контекст интересующего тезиса. Подробность реферата можно легко настраивать, изменяя количество формирующих его предложений. При этом каждое предложение реферата характеризуется относительной степенью значимости во всем тексте.

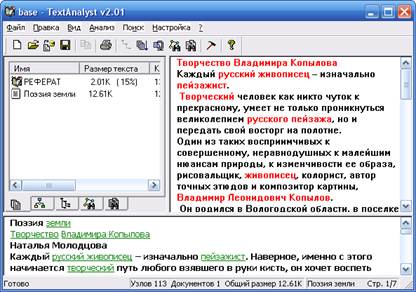


Рисунок 2 – Реферирование

Таким образом, данная программа может быть достаточно полезной в исследовании автора, поскольку позволяет быстро получить смысловой портрет текста, что необходимо в процессе работы с большим количеством разнообразных текстов, когда нелегко удержать в памяти содержание каждого из них. TextAnalyst удобно использовать и при поиске текстов для исследования: не обязательно тратить время на чтение полного текста, с помощью реферирования можно быстро соориентироваться, о чем идет речь (о каких художниках, картинах, направлениях в живописи и т.д.), и сделать вывод о том, подходит ли данный текст для исследования. Однако серьезным недостатком программы применительно к нашему исследованию, связанному с англоязычными текстами, является отсутствие возможности обрабатывать такие тексты.

Теперь рассмотрим автоматический on-line классификатор функционального стиля текста **Худломер** [13]. Данная программа определяет стиль текста: разговорный стиль, стиль художественной литературы, газетно-информационный стиль, научно-деловой стиль. На Рисунке 3 представлен интерфейс Худломера:

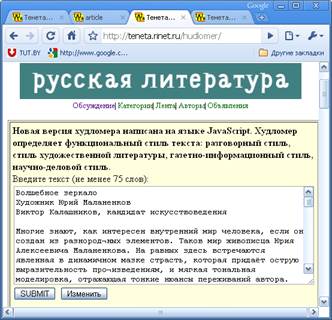


Рисунок 3 – Интерфейс Худломера

В поле под надписью "Введите текст (не менее 75 слов)" вставляем исследуемый текст (в нашем случае, например, статью о русском живописце Ю. Маланенкове "Волшебное зеркало"). Затем нажимаем "SUBMIT" и получаем результат (Рисунок 4):



Рисунок 4 – Ответ Худломера

Итак, видим, что Худломер считает данный текст газетной статьей, то есть относит ее к газетно-информационному стилю. С таким определением вполне можно согласиться, если учесть, что статья была взята из искусствоведческого журнала "Художественный совет" и поэтому является примером публицистики. К тому же, на рисунке видно, что красная полоска под названиями стилей едва пересекает границу между "худло" (стиля художественной литературы) и "газетной статьей", а публицистика как раз и занимает это промежуточное положение.

В нашем случае Худломер определил стиль достаточно точно, однако, как отмечают сами создатели программы, он может давать погрешности, особенно при малых объемах текстов.

Мы также провели подобный эксперимент со статьей на английском языке, взятой из англоязычного искусствоведческого журнала "Art in America", в результате которого Худломер отнес ее к разговорной речи, что, очевидно, не является верным. Возможно, это и есть пример ошибочной работы программы, а возможно, Худломер не анализирует тексты на английском языке, о чем, кстати, в описании программы не упоминается.

Стоит отметить, что построен автоматический классификатор функционального стиля текста на основе спектров длин слов, характерных для каждого из четырех стилей (то есть предполагается, что самые короткие слова встречаются в текстах разговорного стиля, а самые длинные – в научных статьях).

Теперь обратимся к переводчикам **Translate.Ru** [1] и **Google Переводчик** [14]. В нашем исследовании, непосредственно связанном с англоязычными текстами, без перевода не обойтись. Более того, перевод обязательно должен быть точным и высококачественным, ведь статьи на английском языке с параллельным русским переводом помещаются в Приложение к магистерской диссертации. Это выглядит примерно так, как представлено ниже (Рис. 5):

|  |  |
| --- | --- |
| **Текст на английском языке** | **Параллельный русский перевод** |
| **No wonder Doig has exerted such an influence on recent painting. He has been able to paint ordinary, nearly kitsch subjects (boats, water, weather, people copied from photos) in disarmingly alluring ways that somehow, almost by the by, tend to accede to modernist pictorial criteria of flatness and materiality and avoidance of the anecdotal. These criteria remain in force despite the waning of modernism as an ideology. At the same time, he conveys a contemporary rather than a historicist sensibility, and a yearning for the unfamiliar rather than the known. Doig's work of the last few years, however, registers a shift – the import of which remains hard to define since it may still be in progress – away from the quietly delirious, mutedly overelaborated manner of painting that served him so well throughout the '90s. One might say he's beginning to evade his own influence. It's hard not to see the change as connected with his move back to Trinidad in 2002.** | Не удивительно, что Дойг оказал такое сильное влияние на современную живопись. Он смог нарисовать красками простые, почти китчевые предметы (лодки, воду, погоду, людей, скопированных с фотографий) в такой обезоруживающе-притягательной манере, которая как-то почти между прочим тяготеет к тому, чтобы соответствовать модернистским критериям живописи – плоскостность и материальность, и нелюбовь к сюжетности. Эти принципы остаются в силе, несмотря на спад модернизма как идеологии. В то же время он передает современную, а не основанную на историзме восприимчивость и выражает стремление к незнакомому, а не к известному. Однако, творчество Дойга в последние годы отмечается сдвигом (значение которого трудно определить, поскольку он может все еще длиться) – уходом от тихо-сумасшедшей, приглушенно перегруженной деталями (слишком сложной) манеры рисования, которая так хорошо служила ему в 90-е годы. Можно было бы сказать, что он начинает избегать своего собственного влияния. Трудно не связывать эту перемену с его возвращением в Тринидад в 2002-ом году. |

Рисунок 5 – Пример самостоятельного перевода для магистерской работы

Данный перевод выполнен автором самостоятельно. А теперь посмотрим, какие варианты перевода этого текста предоставляют нам автоматические переводчики.

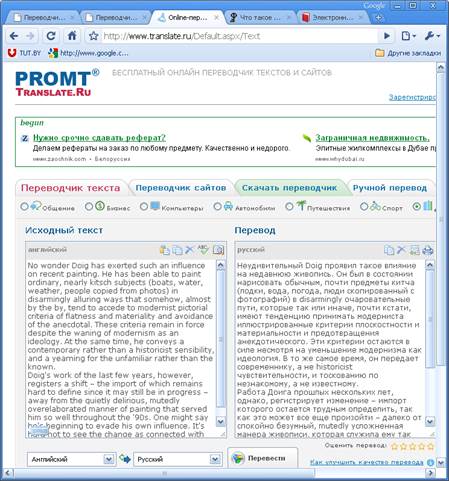


Рисунок 6 – Перевод Translate.Ru

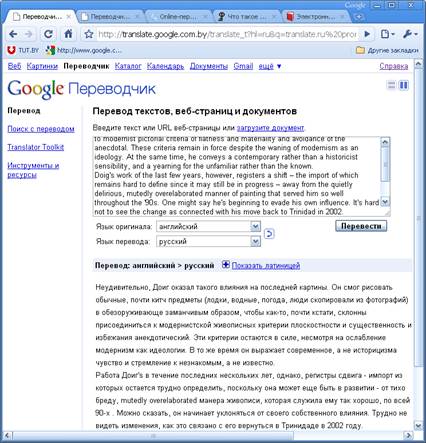


Рисунок 7 – Перевод Google Переводчика

Итак, трудно не заметить, что предложенные варианты перевода очень "шероховатые", во многих местах стилистически и грамматически неточные, уже не говоря о том, что некоторые слова и словосочетания и вовсе остались без перевода ("mutedly overelaborated", "historicist"). Такой перевод несомненно требует доработки, и немалой. На наш взгляд, рассмотренные программы-переводчики скорее будут полезны тому, кто почти не владеет иностранным языком, но хочет понять основную идею текста, ведь они действительно передают общий смысл при переводе. Профессиональному же лингвисту, который хорошо знает язык, общий смысл понятен и так. Ему нужен точный перевод с учетом всех тонкостей и оттенков значений. В это плане на помощь приходит один из самых известных электронных словарей **ABBYY** **Lingvo** (Рисунок 7) с большим количеством встроенных словарей различной тематики, толкованиями, примерами, устойчивыми выражениями и т.д.



Рисунок 8 – ABBYY Lingvo 12

Итак, все описанные выше лингвистические программы действительно могут быть полезны при исследовании англоязычных и русскоязычных искусствоведческих текстов. Однако не стоит всецело полагаться на данные, полученные с помощью таких программ. Лучше еще раз проверить и проанализировать их самостоятельно.

**ЗАКЛЮЧЕНИЕ**

Данная курсовая работа представляет собой самостоятельное исследование возможностей и эффективности применения ИТ в лингвистике в целом и при изучении англоязычных и русскоязычных искусствоведческих текстов в частности.

Все эти направления, находящиеся на пересечении сфер лингвистики и ИТ, реализованы на практике в виде конкретных лингвистических программ и ресурсов сети Интернет, которые можно разделить на 9 групп, рассмотренных выше. Каждая из этих групп, как мы убедились, представлена, в свою очередь, большим разнообразием программ (ресурсов), многие из которых мы кратко охарактеризовали, назвав их основные преимущества. Большинство программ действительно эффективны, а иногда и вовсе необходимы при проведении лингвистических исследований.

Мы также на практике ознакомились с программами TextAnalyst 2.0, Худломер, Translate.Ru и Google Переводчик и оценили их возможности применительно к исследованию русскоязычных и англоязычных искусствоведческих текстов. Выяснилось, что TextAnalyst 2.0 помогает анализировать содержание текста, автоматически формирует семантическую сеть с гиперссылками, создает "смысловой портрет" текста в терминах основных понятий и их смысловых связей, осуществляет реферирование текста. Худломер автоматически определяет стиль текста, однако не всегда делает это правильно. Translate.Ru и Google Переводчик переводят англоязычные искусствоведческие тексты недостаточно точно и корректно – полученный перевод подлежит обязательному редактированию.

ИТ в наше время развиваются очень быстрыми темпами. В перспективе, возможно, появятся технологии, которые будут настолько точно обрабатывать языковой материал, что лингвисту не придется сомневаться в полученных результатах.

**Список литературы**

1. Бесплатный онлайн переводчик текстов и сайтов PROMT Translate.Ru [Электронный ресурс]. – 2003-2009. Режим доступа: http://www.translate.ru/. – Дата доступа: 15.12.2010.

2. Вигурский К.В., Пильщиков И.А. Филология и современные информационные технологии (К постановке проблемы) // Научно-технический центр "Информрегистр" [Электронный ресурс]. – 2007-2009. – Режим доступа: http://feb.inforeg.ru/vigursky-03.html. – Дата доступа: 02.12.2010.

3. Всеволодова А.В. Компьютерная обработка лингвистических данных: учеб. пособие / А.В. Всеволодова. – 2-е изд., испр. – М.: Наука: Флинта, 2007. – 96 с.

4. Зубов А.В. Информационные технологии в лингвистике: Учеб. пособие для студ. лингв. фак-тов высш. учеб. заведений / А.В. Зубов, И.И. Зубова. – М.: Академия, 2004. – 208 с.

5. Зубов А.В., Зубова И.И. Основы искусственного интеллекта для лингвистов: Учеб. пособие. – М.: Университетская книга; Логос, 2007. – 320 с.

6. Информационные системы и применение компьютерной техники в профессиональной деятельности. Основные понятия // Электронный учебник по предмету ИТ [Электронный ресурс]. – 2009. – Режим доступа: http://it-uchebnik.ru/it/glava1/1.html. – Дата доступа: 01.12.2010.

7. Компьютерная лингвистика // Онлайн Энциклопедия "Кругосвет" [Электронный ресурс]. – 2001-2009. Режим доступа: http://www.krugosvet.ru/enc/gumanitarnye\_nauki/lingvistika/KOMPYUTERNAYA\_LINGVISTIKA.html. – Дата доступа: 04.12.2010.

8. Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы: учеб пособие для студ. лингв. фак. вузов / Н.Н. Леонтьева. – М.: Академия, 2006. – 304 с.

9. Логичев С.В. Каталог лингвистических программ и ресурсов в Сети / С.В. Логичев // Русская виртуальная библиотека [Электронный ресурс]. – 1999-2009. Режим доступа: http://www.rvb.ru/soft/catalogue/index.html. – Дата доступа: 25.11.2010

10. Мечковская Н.Б. История языка и история коммуникации: от клинописи до Интернета: курс лекций по общему языкознанию / Н.Б. Мечковская. – М.: Флинта: Наука, 2009. – 584 с.

11. Персональная система автоматического анализа текстов TextAnalyst 2.0 // Microsystems, Ltd [Электронный ресурс]. – 2009. Режим доступа: http://www.analyst.ru/index.php?lang=eng&dir=content/downloads/. – Дата доступа: 05.12.2010.

12. Потапова Р.К. Новые информационные технологии и лингвистика: Учебное пособие. – Изд. 4-е, стереотипное. – М.: КомКнига, 2005. – 368 с.

13. Худломер (автоматический определитель стиля текста) // Тенета (Конкурс русской сетевой литературы) [Электронный ресурс]. – 1996-2003. Режим доступа: http://www.teneta.ru/hudlomer/. – Дата доступа: 10.12.2010.

14. Google Переводчик // Google [Электронный ресурс]. – 2010. – Режим доступа: http://translate.google.com.by/?hl=ru&tab=wT#. – Дата доступа: 04.12.2010.