МИНОБРНАУКИ РОССИИ

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

“ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ”

Факультет *романо-германской филологии*

Кафедра *французской филологии*

*Теория и методика преподавания иностранных языков и культур*

*Курсовая работа.*

*Автоматический синтез и анализ текста.*

Студентка\_\_\_\_\_\_\_\_\_\_\_ Мальцева А.С.

Руководитель\_\_\_\_\_\_\_\_ Донина О.В.

Воронеж 2018

**Оглавление.**

**ВВЕДЕНИЕ……………………………………………………………………**

1. **Автоматический синтез.** 
   1. Семантический синтез.
   2. Синтаксический синтез.
   3. Лексико-морфологический синтез.
2. **Анализ текста.**
   1. Компьютерный анализ.
   2. Цели и задачи.
   3. План.
   4. Виды и приёмы анализа текста.
   5. Методика работы
3. **Заключение.**
4. **Список использованной литературы.**

Автоматический синтез - операция, в которой по заданной грамматической и семантической информации строится содержащий эту информацию текст на естественном языке; операция выполняется по некоторому алгоритму в соответствии с заранее разработанным описанием данного языка.

Автоматический синтез — необходимый этап в разных видах автоматической обработки текстов, в частности при машинном переводе. Автоматический синтез следует отличать от автоматического порождения текстов, при котором строятся произвольные правильные тексты безотносительно к какому бы то ни было предварительному смысловому заданию.

**Автоматический синтез** включает в себя 3 основных этапа. На первом этапе разрабатывается математическая модель (ММ) проектируемой (синтезируемой) системы, представленная, например, системой дифференциальных уравнений. ММ отображает связи между показателями состояния, например быстродействием и надёжностью системы, её точностью, качеством выпускаемой продукции и т. п., и параметрами управления, с помощью которых осуществляется воздействие на проектируемую систему, в результате чего меняются её показатели состояния. Совместно с заданными ограничениями на показатели состояния и параметры управления ММ определяет область допустимых вариантов проектирования. На втором этапе на основе ММ составляется целевая функция (ЦФ), или критерий оптимальности, — математически сформулированная цель С. а. В ЦФ могут быть одновременно отражены несколько показателей состояния с добавлением экономических показателей (например, прибыли, себестоимости и т. п.). В ЦФ также входят параметры управления, набор которых определяет различные варианты проектируемых систем. В некоторых вариантах ЦФ достигает экстремума. Этот экстремум определяет оптимальную область, из которой выбирается один из вариантов для реализации. Второй этап часто объединяется с первым. На третьем этапе разрабатываются способы технической реализации выбранного варианта системы, полученного в виде общих математических зависимостей (рекомендаций), т. е. составляются функциональные, структурные, динамические, электрические, сборочные и т. п. схемы.

**Автоматический синтез** делится на 3 типа:

1. **Семантический синтез**
2. **Синтаксический синтез**
3. **Лексико-морфологический синтез**
4. **Семантический синтез.**

Семантический синтез - переход от смысловой записи фразы к её синтаксической структуре.

Семантический синтез - процесс построения синтаксических деревьев из последовательности фактов. Алгоритм семантического синтеза незамысловат: берём факт, генерируем выходной узел. Из атрибутов факта создаём или атрибуты узла, или дочерние узлы. Продолжаем ту же процедуру для дочерних узлов. Когда дерево сформировано, переходим к следующему факту. И т.д. Атрибут отрицания, например, при преобразовании превращается в самостоятельный узел - частицу "не". А атрибут "время" превращается в одноимённую грамматическую категорию.

1. **Синтаксический синтез.**

Синтаксический синтез - это линеаризация синтаксического уровня языка. На входе имеем последовательность предложений, на выходе каждое предложение заменяется последовательностью атрибутированных лемм. Всё предельно просто. Каждому виду связи между двумя словами поставлен в соответствие порядок их следования в результирующем тексте. Преобразователь, двигаясь от корня к листьям, удаляет из дерева каждую найденную связь, создавая, тем самым, два новых дерева, и размещает их в заданном порядке. Далее преобразование рекурсивно применяется к полученным деревьям до тех пор, пока все они не будут состоять из одного узла. Полученная последовательность и будет результатом преобразования.

1. **Лексико-морфологический синтез**.

Лексико-морфологический синтез обычно проводится на лексико-морфологическом множестве, смоделированном на этапе лексико-морфологического анализа, и осуществляется с помощью поисковых методов лексико-морфологического синтеза, к которым относятся: метод независимой оценки подсистем, метод древовидного конструирования, метод лабиринтного конструирования, метод совершенствования прототипа.

1. **Анализ текста**

Анализ текста — процесс получения высококачественной информации из текста на естественном языке. Как правило, для этого применяется статистическое обучение на основе шаблонов: входной текст разделяется с помощью шаблонов, затем производится обработка полученных данных. Также используется и в литературе.

Анализ - это метод исследования какого-либо предмета или действия, путем выделения отдельных частей, чтобы лучше понять смысл. Анализировать можно все, что угодно: стихотворение, текст, поступок, сказанные слова и так далее. Главное, соблюдать некоторые правила. Что же касается анализа текста по школьным предметам (литературе или русскому языку), то данное занятие помогает не просто читать книги, а читать осмысленно. Так, чтобы после прочтения можно было легко пересказать произведение и уловить мысль автора. Конечно на первых этапах ученик будет задавать себе вопрос, как делать анализ текста. Но в последствие ему станет легче понимать произведения, когда они станут более сложными. Данный метод работы также помогает настраиваться на творческую работу и раскрывает личностное восприятие.

**Компьютерный анализ** текста на естественном языке активно развивается в последние годы многими коллективами. Доступные сегодня вычислительные мощности позволяют применять для обработки больших массивов документов широкий класс математических методов, способствующих эффективному решению задач поиска, классификации, кластерного анализа, выявления скрытых закономерностей в данных и др.

К сожалению, внедрение математических методов в обработку текста происходит в то время, когда собственно лингвистическая составляющая алгоритмов представлена явно недостаточно, и это не позволяет достичь высокого качества работы прикладных систем. Устойчивый уклон в область статистических методов анализа привел к тому, что компьютерная лингвистика оказалась невостребованной. В самом деле, во всех известных русскоязычных системах подобного класса из лингвистического обеспечения используется лишь морфологический словарь, позволяющий отождествлять различные словоформы, тогда как алгоритмы синтаксического анализа реализованы исключительно в автоматических переводчиках и вызывают множество нареканий в связи с невысокой точностью.

**Цель**, которую поставила перед собой автор данной работы - это выяснить, что такое компьютерный анализ текста.

При этом необходимо решить следующие **задачи**:

- ознакомиться с понятием анализ текста;

- рассмотреть, что понимается под компьютерным анализом текста;

- ознакомиться с историей развития компьютерного анализа текста;

- выявить проблемы компьютерного анализа текста;

- привести некоторые программы, используемые при компьютерном анализе текста.

Анализ (греч. "analysis" - разложение) - метод научного исследования (познания) явлений и процессов, в основе которого лежит изучение составных частей, элементов изучаемой системы.

Аналитические методы широко распространены в социуме, поэтому термин "Анализ" часто воспринимается как синоним исследования вообще и особенно при решении познавательных задач. Анализа является составной частью любого научного исследования, образуя, как правило, его первую стадию, когда исследователь выявляет в описании изучаемого объекта его строение, состав, свойства, признаки и т.п. Он используется как метод получения новых результатов в процессе мыслительной деятельности человека.

Мыслительный анализ совершается с помощью понятий и суждений, выражаемых в естественных или искусственных языках. Такой анализ ориентирован на выявление структуры целого, предполагая фиксацию его частей и установление отношений между ними.

В современном обществе важным средством оформления, фиксации, сохранения, передачи информации и обмена ею являются документы.

Анализ документов - это метод сбора первичных данных, при котором документы используются в качестве главного источника информации; это также совокупность методических приёмов и процедур, применяемых для извлечения информации из документальных источников при изучении процессов и явлений в целях решения определённых задач.

Анализ текста - процесс получения высококачественной информации из текста на естественном языке. Как правило, для этого применяется статистическое обучение на основе шаблонов: входной текст разделяется с помощью шаблонов, затем производится обработка полученных данных.

Анализ текста - это процесс получения на естественном языке высококачественной информации из этого текста. Такой анализ осуществляет практически каждый человек не задумываясь, что он делает именно это. Например, любой читатель книги анализирует содержание, читая её по частям. В общем случае всем, особенно трудоспособному населению, постоянно приходится работать с текстовыми и иными видами документов, явно или неявно анализируя их содержание и другие компоненты, например, качество изготовление, форму, размер и т.д. Таким образом, важной задачей практически любых индивидов является нахождение (получение) нужных им документов, а также анализ их содержания на предмет подготовки различных документов (рефератов, аннотаций, справок, отчётов, учебных работ, производственных заданий и др.). При этом нередко возникают ситуации, когда в течение незначительного периода времени необходимо подготовить некоторый (как правило, аналитический) материал. Анализ текстов на естественном языке (ЕЯ) был актуальным практически с момента их появления. При таком анализе необходимо определить правила, с помощью которых, по мнению специалистов, "формальная система (набор структурных элементов текста) преобразуется в систему содержательную (осмысленное сообщение". Анализ текста использовался и продолжает использоваться для классификации текстов, содержащихся в них слов и словосочетаний (например, для определения частоты встречаемости тех или иных терминов в определённых текстах), аннотирования и реферирования текстов, проведения семантически ориентированного поиска текстов по заданным концептам, определения авторского права претендента на соответствующий текст и др.

Считается, что в реальной жизни правила чтения текстов не формализуются. Человек постигает их годами, активно работая в определённой сфере деятельности и предметной области. Специалисты отмечают, что разные контексты порождают различные правила их прочтения, которые со временем меняются, поэтому трудно научить таким правилам компьютер, а значит автоматически учитывать содержательные аспекты обрабатываемого текста.

**План:**

1. **Роль и место анализа текста.**
2. **Виды и приемы анализа текста.**
   1. **Роль и место анализа текста.**

В начальной школе восприятие сформировано у учащихся на разных уровнях, поэтому мы должны включать в урок различные виды анализа текста, для более полноценного восприятия произведения.

Анализ текста идет на этапе повторное чтение и углубленный анализ текста. Цели:

- способствовать полноценному восприятию художественного произведения детьми;

- учить понимать авторскую позицию;

- учить понимать мотивы поступков и характеры героев;

- учить распознавать художественные средства выразительности;

- учить определять жанровые особенности;

- способствовать средствами литературы развитию личности ребенка, его нравственных идеалов;

- развивать воображение детей, навыки работы с текстом.

2. **Виды и приемы анализа текста.**

Методисты выделяют 4 вида анализа литературного произведения:

**1 вид: анализ развития действия** – в его основе лежит работа над сюжетом и его элементами, т.е частями и главами. Предполагает работу над сюжетом литературного произведения и его элементами (эпизод, глава) при этом задача учителя – найти вместе с детьми черты целостности в каждой части произведения и органичную связь части с целым.

Плодотворен бывает анализ, который начинается с ключевой (кульминационной) сцены, где затронуты главные вопросы, волнующие писателя. Большое значение для понимания содержания текста имеют также начало (экспозиция и завязка) и конец (развязка), тк они наиболее полно выражают тему и идею художественного произведения. Именно над этими эпизодами необходимо проводить специальную работу в рамках целостного анализа развития действия.

Обратим внимание ребят на композиционные особенности отрывка из рассказа К. Г. Паустовского «Заячьи лапы»: он имеет рамочную композицию («рассказ в рассказе»).

План текста может выглядеть следующим образом:

1) Ночь на Урженском озере.

2) Неудачная охота.

3) Лесной пожар.

4) Надежда на спасение.

5) Спасены!

6) Чувство вины.

Вопросы: Что, где, когда произошло? Почему? Как поступил при этом герой? Как это его характеризует? А что произошло потом? Можно ли было избежать этого? Почему все-таки так получилось: кто виноват в этом?

**2 вид: проблемный анализ** – во главу анализа поставлен проблемный вопрос в ходе поиска ответа на него могут возникнуть разные точки зрения, которые подтверждаются вчитыванием в текст. Чтобы превратить проблемный вопрос в проблемную ситуацию, надо обострить противоречия, сопоставить разные варианты ответов. Постановка вопросов проблемного характера целесообразна при чтении тех произведений, где есть ситуации, предполагающие различное понимание героев, их действий, поступков, этических проблем, затронутых писателем.

*Черты проблемных вопросов:*

1) наличие противоречия и возможности разных ответов;

2) интерес детей к поставленной теме;

3) возможность сопоставления разных ответов на вопрос.

Пример: Д.Мамин-Сибиряк «Дурной товарищ»:

Правильно ли сделали мальчики, что пошли на лед?

Как произошло, что ребята очутились на льду?

Как Коля относился к отцу?

Какой смысл в названии «дурной товарищ» вкладывали тетушка и отец Коли?

Как вы думаете, Коля поступил бы так еще раз?

Почему Коля боялся отца поле случившегося?

Как вы относитесь к поступку Коли?

**3 вид: анализ художественных образов** – в центре анализа образы героев или пейзаж Анализ художественных образов.

*Основные принципы.*

1. при анализе доводим до детей, что действующее лицо – это представитель определенной общественной группы людей, эпохи, и одновременно это живой конкретный человек.
2. в характере героя определяем ведущие черты.
3. выясняем авторское отношение героев (вычитывая в тексте)
4. способствует сопереживанию детей, их личному отношению к герою.

*Последовательность работы:*

1. эмоциональное восприятие героев:

- после первоначального чтения, что можете сказать о герое? Понравился или нет? Чем?

2. анализ образов идет в ходе повторного чтения:

1) чтение эпизода или слов, говорящих что-либо о герое – беседа: как характеризует это героя, его слова, поступки. Подбор слов называющих черту героя – вывод обобщение, т.е рассказ о герое.

2) учитель или ученики называют качество героя – подтверждение текста и вывод – составление рассказа о герое.

*План рассказа о герое.*

1. кто он такой? (когда, где он жил, живет, его возраст, пол)
2. внешность героя.
3. какие поступки он совершает и как его это характеризует
4. особенности речи героя.
5. авторское отношение к герою.
6. мое отношение.

*Приёмы работы над характеристикой героя:*

1. учитель называет качество героя, дети подтверждают текстом.
2. дети самостоятельно называют свойство характера и подтверждают текстом.
3. сравниваются герои одного произведения или близких по тематике.
4. выделение авторских слов косвенно характеризующих героя.
5. придумывание ремарок автора, передающих душевное состояние героя.
6. лингвистический эксперимент: исключение из текста слов, содержащих авторскую оценку.
7. вывод (главная идея – смысл произведения)

в процессе разбора учащиеся должны понять и характеристику образа (героя, пейзажа), и смысл этого образа, т.е нагрузку, которую он несет в общей структуре произведения.

Вопросы для анализа образа героя, на примере произведения Г.Х.Андерсена «Гадкий утенок»:

Ваше отношение к главному герою? Понравился он вам или нет?

Как он выглядит? Как потом преобразился? Кто он? Где родился?

Что с ним происходит? Отношение окружающих к главному герою.

Как вы думаете, что он чувствовал, когда его обижали?

Какие качества характера помогли пережить ему (гл.герою) все эти обиды и унижения?

Как изменился его внешний облик? Что с ним произошло? А изменился ли его характер? Какое отношение автора к главному герою?

Характеристика образа героя, беседа по тексту, выборочное чтение:

Как появился на свет гадкий утенок? Почему он гадкий? Как отнеслись к утенку жители птичьего двора? Какое отношение братьев, сестер и матери? Как сам утенок относился к себе? На какой поступок решился герой? Смирился ли утенок со спокойной жизнью старушки? Легко ли приходилось бедному утенку? Какие чувства испытывает утенок при виде прекрасных птиц? Глее он очутился весной? Каким был для него сал? Как воспринимает окружающее? А настоящие утята смогли бы увидеть такую красоту? Что выбрал утенок при встрече с прекрасными птицами? Что меняется? Как сам утенок чувствует себя, относится к восторженным взглядам? Какие поступки совершал герой?

Изменилась его внешность, изменилось ли его сердце, внутреннее отношение, состояние? Правильно ли то, что люди воспринимают друг друга по внешности?

Сейчас мы составим рассказ о гадком утенке:

В чаще лопуха утка высиживала яйца, все вылупились, но одно большое вылупилось позже, оттуда появился большой и гадкий утенок.

Мама-утка говорила: Ужасный урод! И совсем не похож на других детей!

Когда они пришли на птичий двор, все только и говорили о том, какой он гадкий, большой и несуразный , все пытались его клюнуть, дразнили, смеялись над ним. Мама-утка защищала его: Он некрасив – это правда – но у него доброе сердце…. Я думаю он вырастет сильным и пробьет себе дорогу в жизнь. Его братья и сестры клевали, щипали его.

Он самокритичен, тк всегда думал, что от него все убегают и птицы, и охотничья собака, именно потому, что он такой гадкий. «Кажется, пронесло, - подумал утенок и перевел дух. – Видно, я такой гадкий, что даже собаке противно съесть меня!»

Когда его обижали, он терпел все эти обиды, боялся, мечтал о самом хорошем и красивом, о свободе. Благородство, гордость, дружелюбность, доброжелательность, добродушие, мужество помогли ему пережить все эти невзгоды.

Наступила весна, на озере он увидел прекрасных лебедей. Красоте их он не завидовал. И вдруг в чистой зеркальной воде, он увидел собственное отражение. Он был уже не гадким темно-серым утенком, а красивым белым лебедем. Теперь утенок был даже рад, что перенес столько горя и бед. Он много вытерпел и поэтому мог лучше оценить свое счастье. А больше лебеди плавали вокруг и гладили его клювами. Но характер его не изменился, он остался таким же скромным, добрым. Изменилось лишь его внутренне состояние: радостное, ликующее, восхищающееся.

Автор вначале его жалел, тк он был беспомощен: Бедный утенок устал, и ему было очень грустно. Бедняжка! В конце автор радовался за утенка.

Мне герой нравится, здорово то, что он мужественно и гордо перенес столько горя и бед, и в награду ему такая красота, счастье, о котором он и не мечтал, когда был еще гадким серым утенком.

**4 вид: стилистический (языковой) анализ** – это анализ изобразительных средств языка используемых автором в данном произведении.

*Цель анализа:* помочь детям уяснить мысли и чувства автора, выраженные прежде всего в образных словах, развитие воображения, выразительное чтение.

*Методика работы:*

1. выделение слова или образного выражения.
2. определение мысли и чувства, заключенного в них (почему поэт так называет…., какую картину вы представляете при этом? Какое чувство испытывает автор? Что с чем сравнивает автор? Почему?)
3. приемы стилистического анализа:
   1. сравнение поэзии и прозы, посвященных одной теме
   2. сопоставление разных произведений одного автора. **Цель:** выделение характерных черт творчества.
   3. Сравнение произведений разных авторов на одну тему. **Цель:** определение своеобразия авторской манеры (3-4кл)
   4. Замена авторского слова синонимом. **Цель:** привлечь внимание детей к авторскому слову.

Например, спрашивая детей о том, какую картину зимы рисует поэт – грустную или веселую, нравится ли ему зима, какие слова убеждают нас в этом, мы стараемся показать роль образного слова в связи с художественным образом – картиной природы.

Итак, работа над словом должна быть направлена на то, чтобы помогать детям уяснять образный смысл произведения, мысли и чувства писателя, которые воплощаются в подборе лексики, в ритме фразы, в каждой художественной детали. Поэтому для языкового анализа выбираются слова и выражения, которые помогают лучше понять образный смысл произведения (рисуют картины природы, выявляют авторские чувства) и в то же время наиболее выразительны и точны. После выделения слова или выражения осознается их роль в тексте, те определяется, какие чувства (мысли) заключены в них.

Часто младшие школьники не понимают переносных значений слов. Так, например, на вопрос: Почему лед сравнивается со стеклом? (Бианки, Синичкин календарь) дети отвечают: потому что он белый. – а стекло белое? – да. вот почему нужна специальная работа над метафорами, сравнениями.

Пример языкового анализа: Федор Тютчев «Весенние воды»

Какую весну изобразил автор? (раннюю) Какое настроение выразил поэт?

Какие слова вызывают и у нас радостное настроение? Какой была природа до появления ручьев? (сонной) Как ручьи разбудили природу? (журчанием) Какие звуки повторяются? (взрывной б) В какой речи используются старые слова? ( в поэтической, придают торжественность) Есть ли здесь старые слова? (брег, гласят) Как называют себя весенние воды? ( мы молодой весны гонцы) И почему? (метафора) Как называют скрытое сравнение? А какой еще прием использует автор? (олицетворение) Почему поэт называет майские дни хороводом? (весна пришла – это праздник) Найдите слова, описывающие хоровод? (румяный, светлый) Почему автор использует эпитет румяный? (от Солнца, его нет, но через эпитет мы увидели) Какое чувство надо передать при чтении этого стихотворения?

**Заключение.**

Главный прием всех видов анализа это беседа над текстом.

Какой выбрать анализ зависит от характера (жанра) произведении, от возможностей учащихся, но все же более распространенным и целесообразным можно считать анализ художественных образов.

1. **Список использованной литературы:**
   1. <https://www.litprichal.ru/slovari/enc_sovet/avtomatic..>
   2. <https://studfiles.net/preview/3192703/>
   3. <https://dic.academic.ru/dic.nsf/bse/132560/Синтез>
   4. <https://studfiles.net/preview/3067182/page:14/>
   5. <https://studfiles.net/preview/2975182/>