Синтаксический и семантический анализ текста и речи.

Выполнила:студентка гр.БИН-41(оз)
Сидикова Д. А.
Проверила:Каратонова С. А.

Синтакси́ческий ана́лиз **в лингвистике и информатике — процесс сопоставления линейной последовательности слов естественного или формального языка с его формальной грамматикой. Результатом обычно является дерево разбора (синтаксическое дерево). Обычно применяется совместно с лексическим анализом.**

Синтаксический анализатор— это программа или часть программы, выполняющая синтаксический анализ. В ходе синтаксического анализа исходный текст преобразуется в структуру данных, обычно — в дерево, которое отражает синтаксическую структуру входной последовательности и хорошо подходит для дальнейшей обработки.

Как правило, результатом синтаксического анализа является синтаксическое строение предложения, представленное либо в виде дерева зависимостей, либо в виде дерева составляющих, либо в виде некоторого сочетания первого и второго способов представления

Структура синтаксического разбора.

Просто о синтаксическом разборе предложения
1)Охарактеризовать предложение по цели высказывания: повествовательное, вопросительное или побудительное.
2)По эмоциональной окраске: восклицательное или невосклицательное.
3) По наличию грамматических основ: простое или сложное.
4)Затем, в зависимости от того, простое предложение или сложное:

Если простое:

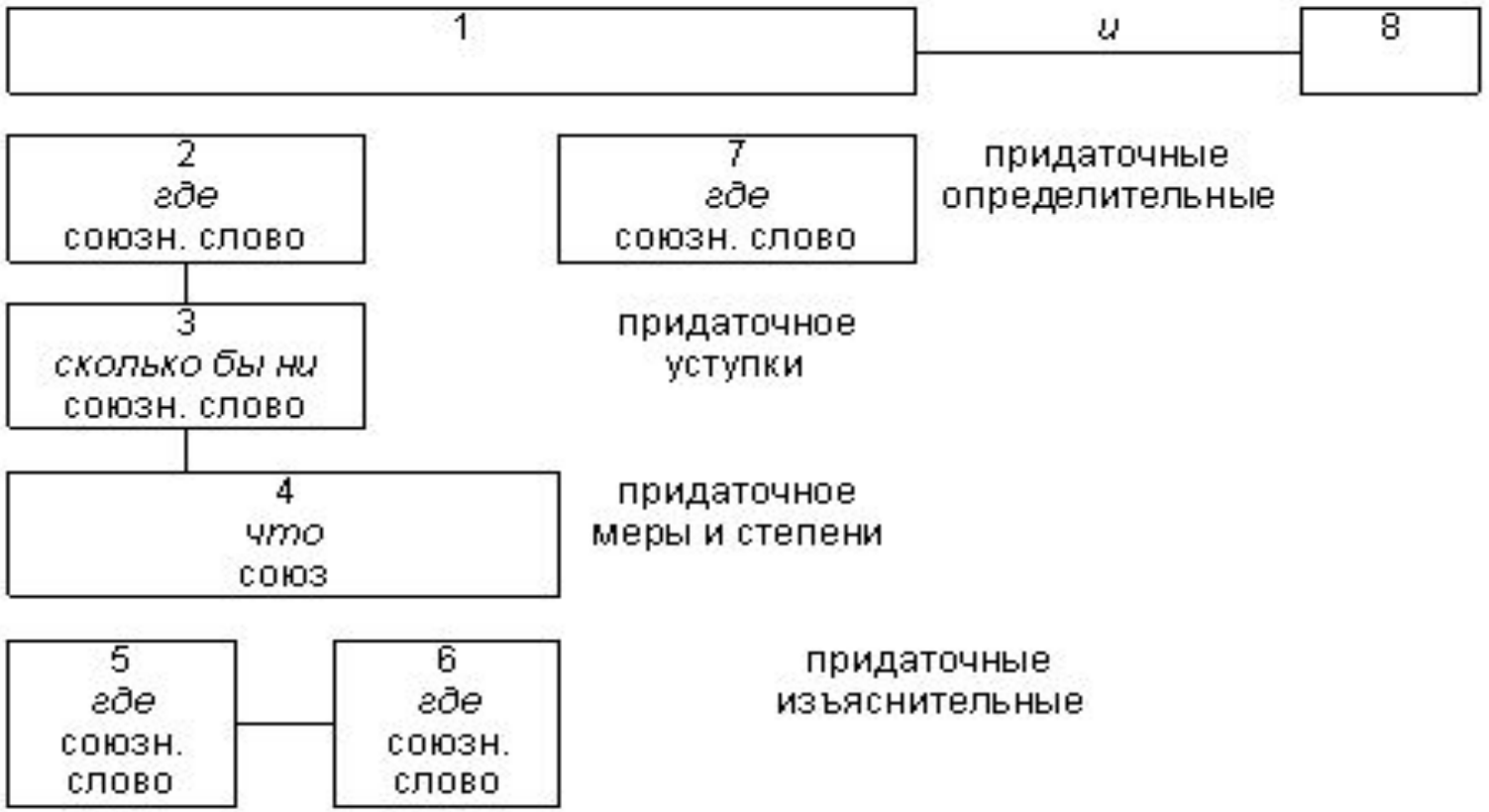
5)Охарактеризовать предложение по наличию главных членов предложения: двусоставное или односоставное, указать, какой главный член предложения, если оно односоставное (подлежащее или сказуемое).
6)Охарактеризовать по наличию второстепенных членов предложения: распространённое или нераспространённое.
7)Указать, осложнено ли чем-либо предложение (однородными членами, обращением, вводными словами) или не осложнено.
8)Подчеркнуть все члены предложения, указать части речи.
9) Составить схему предложения, указав грамматическую основу и осложнение, если оно есть.

Если сложное:

5)Указать, какая связь в предложении: союзная или бессоюзная.
6)Указать, что является средством связи в предложении: интонация, сочинительные союзы или подчинительные союзы.
7)Сделать вывод, какое это предложение: бессоюзное (БСП), сложносочинённое (ССП) сложноподчинённое (СПП).
8)Разобрать каждую часть сложного предложения, как простое, начиная с пункта №5 соседнего столбца.
9)Подчеркнуть все члены предложения, указать части речи.
10)Составить схему предложения, указав грамматическую основу и осложнение, если оно есть.

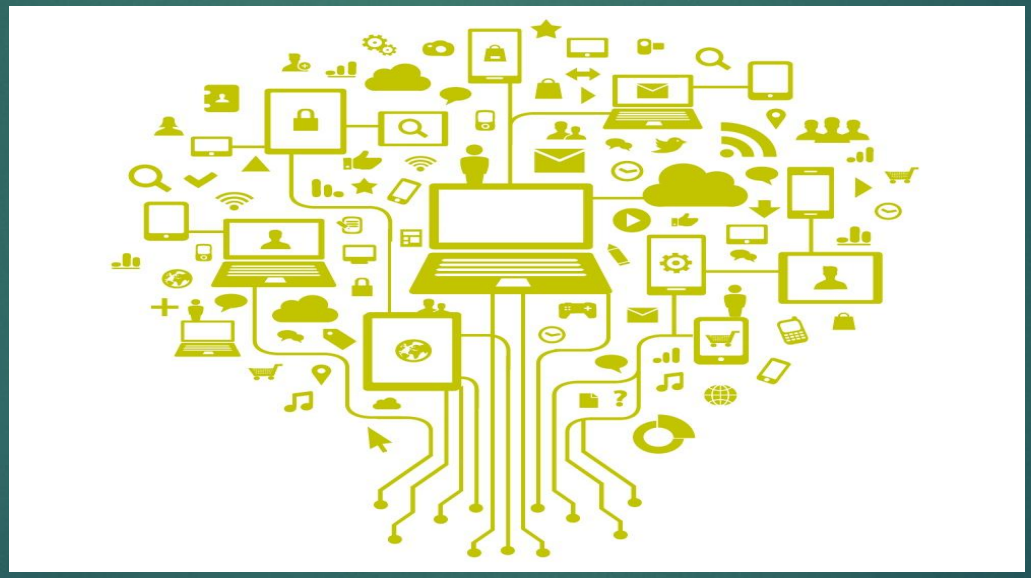
Пример схемы (предложение, после него схема)Источник:

В бывшем кабинете деда,[1]/ где даже в самые жаркие дни была могильная сырость,[2]/ сколько бы ни открывали окна, выходившие прямо в тяжелую, темную хвою, такую пышную и запутанную,[3]/ что невозможно было сказать,[4]/ где кончается одна ель,[5]/ где начинается другая,[6]/ — в этой нежилой комнате,[1]/ где на голом письменном столе стоял бронзовый мальчик со скрипкой,[7]/ был незапертый книжный шкап,[1]/ и в нем тяжелые тома вымершего иллюстрированного журнала[8]/



Семанти́ческий ана́лиз

Слово семантика имеет древнегреческое происхождение, в дословном переводе означает «значительный». Впервые этот термин использовал французский филолог Мишель Бреаль. Под этим понятием принято подразумевать науку, что изучает суть текста, смысл слов и предложений, а также отдельные буквы древних алфавитов. Проще говоря, эта наука пытается понять лингвистический и философский смысл языка, проводя семантический анализ текста.



Особенности семантики и семантического анализа

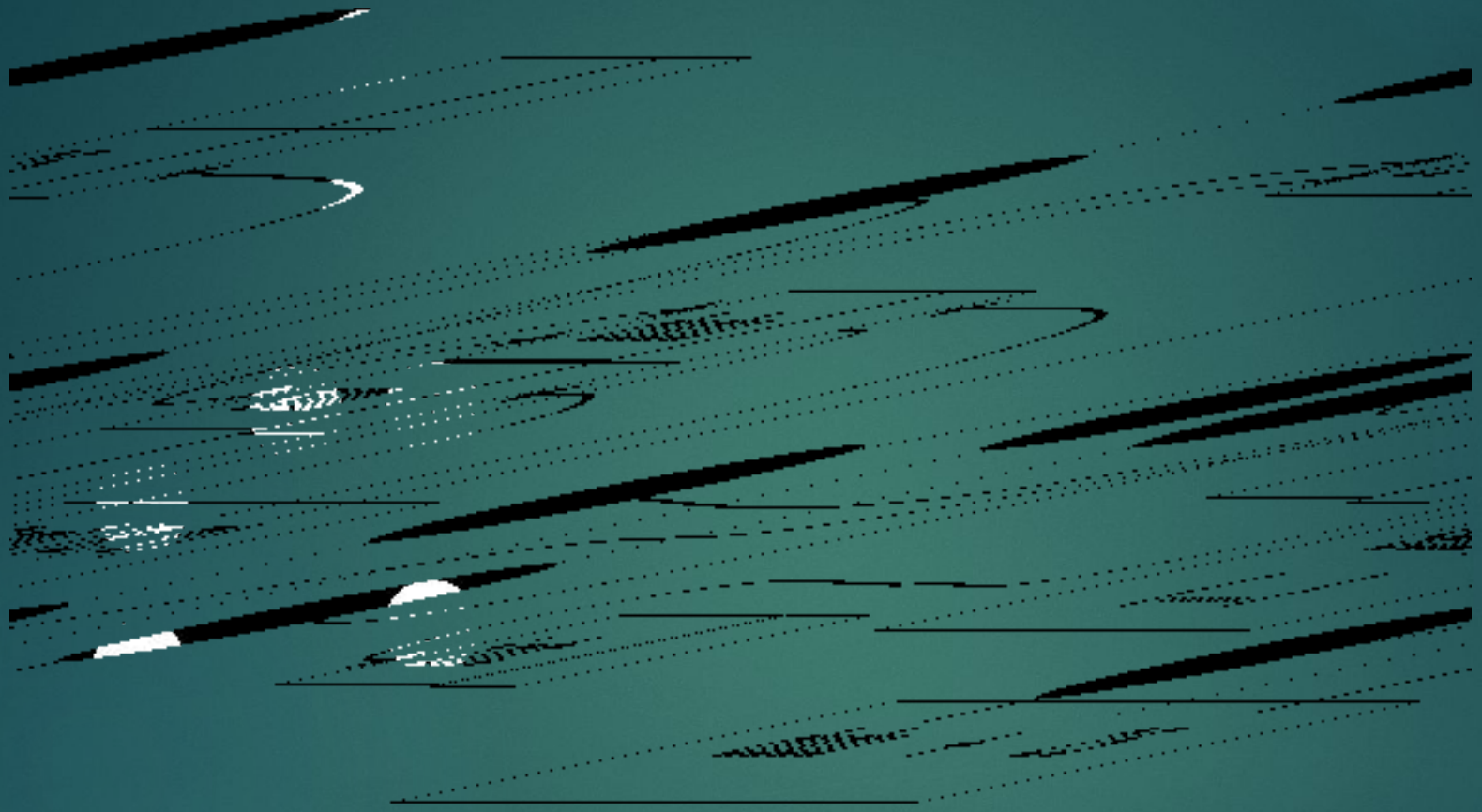
Семантический анализ, существенно отличается от синтаксического. И дело не столько в том, что фаза семантического анализа реализуется не формальными, а содержательными методами (т.е. на данный момент нет универсальных математических моделей и формальных средств описания «смысла» программы).Синтаксический анализ имеют дело со структурными, т.е. внешними, текстовыми конструкциями языка. Семантика же, ориентированная на содержательную интерпретацию, имеет дело с внутренним представлением «смысла» объектов, описанных в программе. Для любого, имеющего опыт практического программирования, ясно, что формальные конструкции языка дают описание свойств и действий над внутренними объектами, с которыми имеет дело программа. Для начала перечислим все, что их касается и лежит на поверхности:

- *большинство объектов являются именованными. Имя объекта позволяет его идентифицировать, существуют различные области действия имен, соглашения об именах, различные умолчания и т.п.. Все это относится к семантике;*

- *виды, сложность и набор характеристик объектов различаются в разных языках программирования и сильно зависят от области приложения языка (в этом смысле семантика языков программирования более разнообразна, нежели синтаксис и лексика). Например, классический Си, ориентированный на максимальное приближение к архитектуре компьютера, работает с такими объектами, как типы данных, переменные, функции. Все они имеют различные свойства и характеристики. Например, переменная характеризуется именем, типом данных, размерностью, областью действия, временем жизни, текущим значением;*

- *объекты связаны между собой (ссылаются друг на друга). В том же Си переменная ссылается на описание того типа данных, к которому она относится, далее производный тип данных ссылается на базовый и т.п.. Можно сказать, что семантика программы во внутреннем представлении выглядит как система взаимосвязанных объектов;*

- *внутреннее представление семантики программы не совсем удачно называется семантическими таблицами. На самом деле структура данных, соответствующая представлению семантики, может быть любой. Термин «таблицы» говорит о том, что имеются множества объектов различных типов, для каждого из которых заведена отдельная таблица, но нельзя забывать, что элементы различных таблиц связаны между собой. Наиболее близкий термин для описания подобной системы – база данных.*



Задача семантического анализа, т.е. «описания смысла» фразы относится скорее к области искусственного интеллекта. Ее неформализуемость означает, что она **не имеет формальных средств описания,** например, языков**.** Следовательно, семантическая модель языка разрабатывается в каждом случае уникально, здесь отсутствует общий подход, а имеет место набор частных решений и рекомендаций. Отсюда и уникальность семантики языка.

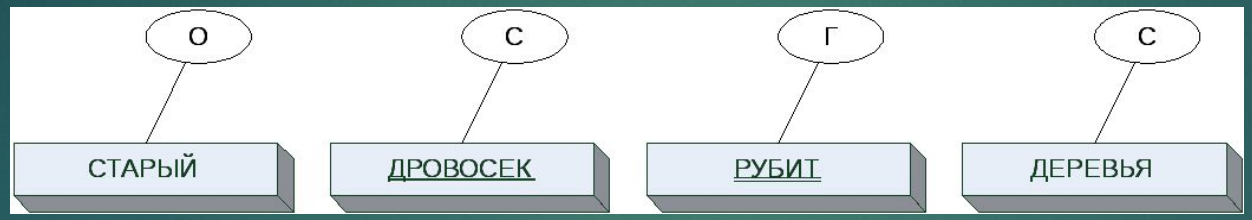
Задачи семантического анализа:

1. Построение семантической интерпретации слов и конструкций;
2. Установление «содержательных» семантических отношений между элементами текста.

Результат семантического анализа — семантическая структура предложения или текста.

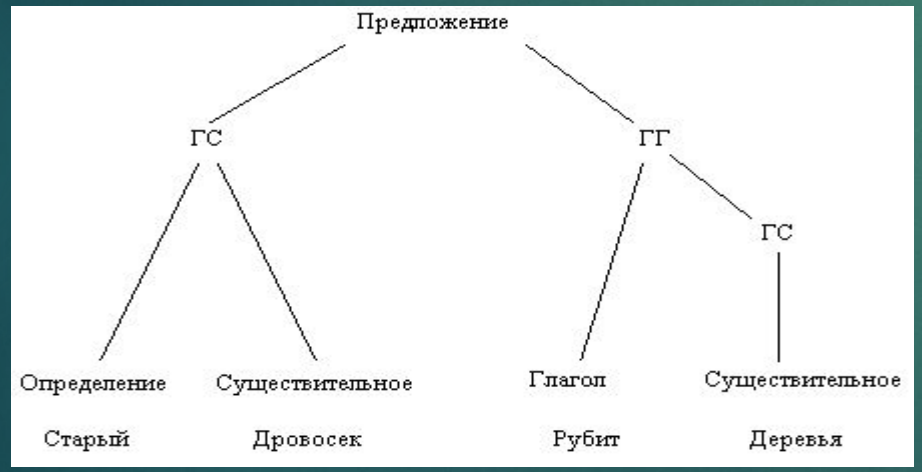
Пример

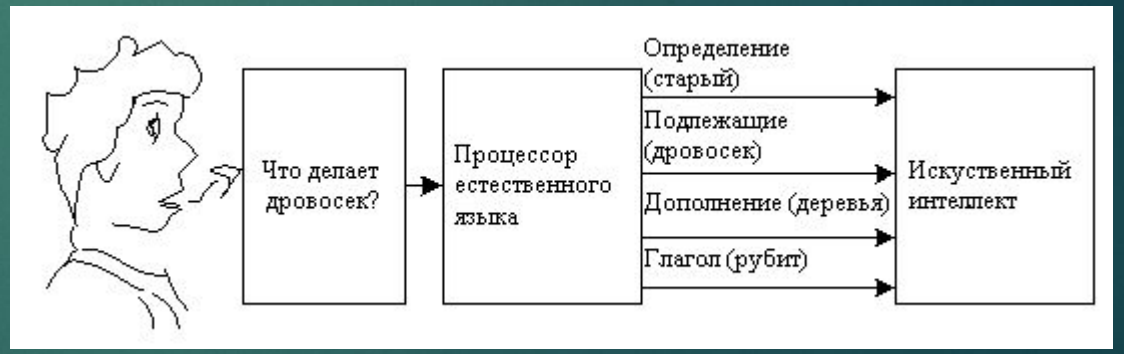
Разбив предложение на составные части, компьютер проводит его семантический анализ, т.е. пытается понять его смысл. В системах искусственного интеллекта применяется некоторая совокупность правил, позволяющая компьютеру понять смысл предложения:



Для интерпретации предложения в базе знаний семантического анализатора должен быть следующий набор правил.

Правило 1: ЕСЛИ определение стоит на первом месте и за ним идет существительное, ТО существительное является подлежащим.
Правило 2: ЕСЛИ за подлежащим идет глагол, ТО этот глагол является сказуемым и поясняет, что делает подлежащее.
Правило 3: ЕСЛИ за подлежащим идет сказуемое, а за ним следует существительное, ТО это существительное является дополнением.
Правило 4: ЕСЛИ предложение имеет следующий порядок слов: подлежащее, глагол, дополнение, ТО вся фраза говорит о том, что подлежащее делает (действие, выраженное сказуемым) по отношению к дополнению.





Поясним сказанное на примере. Предположим, что система искусственного интеллекта должна решить следующую задачу: узнать, что делает дровосек и что является объектом его действия. Семантический анализатор обращается к *правилу 1*, с помощью которого определяет, что слово “дровосек” - это подлежащее. С помощью*правила 2*определяется, что слово “рубит” - это сказуемое. Объект действия, выраженный словом “дерево”, устанавливается с помощью*правил 3 и 4*. Данный пример показывает, как процессор естественного языка обрабатывает или “понимает” предложение, используя лексические, синтаксические и семантические правила своей базы знаний.

Процессор естественного языка может служить промежуточным звеном между пользователем и другой системой искусственного интеллекта, позволяя человеку устно общаться с компьютером (см. рис.Процессор ЕЯ). По существу, обработка естественного языка может освободить пользователя компьютера от необходимости изучать сложные языки программирования. Если удастся создать программы, которые позволят компьютеру и пользователю общаться на естественном языке, то будет сделан крупнейший шаг на пути создания подлинно “интеллектуального” компьютера.

Зачем нужен семантический анализ?

Благодаря ей можно определить особые комбинации слов, что будут формировать основную нить повествования. Умея грамотно и гармонично сочетать слова, можно создать интересную статью, которая наверняка заставит читателя действовать.

К тому же поисковые системы используют основы семантики, чтобы отвечать на запросы пользователей. Благодаря семантическому анализу поисковые роботы могут моментально определить смысл статьи и поставить ее на соответствующую позицию в поисковой выдаче

Автоматическая помощь

Каждый копирайтер может воспользоваться специальными программами, которые проводят структурно-семантический анализ текста совершенно бесплатно. Существуют программы, проверяющие статьи на уникальность, они обладают определенными характеристиками структурно-семантического анализа. Она покажет количество символов, процент воды, количество стоп-слов и ошибок.
Проще говоря, для любого примера семантический анализ текста будет произведен почти в полном объеме, вне зависимости от пожеланий пользователя. Эти программы работают по стандартному алгоритму вычислений. Сегодня семантический анализ нашел себе применение в различных категориях исследований. Он активно используется в информатике, информационных технологиях, развитии техники и других областях, хотя изначально был объектом размышления только в психологии и лингвистике.
Возможно, причиной всему технический прогресс, который развивается так быстро, что появившиеся пробелы знаний приходится закрывать достижениями прошлого. А может, из-за своей простоты - целое состоит из фрагментов, которые нужно исследовать исключительно в рамках этого целого

Программы используемы для анализа текста и речи.

Название	Авторы,организация
word2vec	© Tomas Mikolov, etc., Google, 2013.
Apache OpenNLP	The Apache Software Foundation, Incubator
Russian Morphological Dictionary †	Sergey Sikorsky
Mystem	Илья Сегалович, Виталий Титов компания Яndex
Программные продукты фирмы LingSoft	LingSoft, Финляндия
Система StarLing	С.А.Старостин
Galaktika-ZOOM	корпорация Галактика, Москва

http://rvb.ru/soft/catalogue/c01.html

Спасибо за внимание.