

Лингвистические основы информатики

Часть 1

Рубашкин Валерий Шлемович

д. т. н., профессор
кафедры информационных систем
факультета филологии и искусств
СПбГУ

ТЕМАТИЧЕСКИЙ ПЛАН

Введение. Лингвистические модели в современных информационных технологиях.

Раздел 1. Лингвистические модели и методы формализации текста.

1. Язык как система. Уровни описания языка.
2. Синтаксическая структура предложения и способы ее формализации.
3. Автоматизация словарных работ и статистический анализ текста.

Раздел 2. Логико-семантические модели текста

4. Основные понятия и проблемы семантики.
5. Языки представления знаний.
6. Семантическое описание лексики. Семантические словари.

ЛИТЕРАТУРА

(к части 1)

Основная:

1. Лингвистический энциклопедический словарь. – М., 1990; М., 2000
2. Беляева Л. Н. Лингвистические автоматы в современных информационных технологиях. – СПб.:Изд-во РГПУ, 2001.
3. Баранов А. Н. Введение в прикладную лингвистику. – М., 2001
4. Г.Г. Белоногов, Ю.П. Калинин, А.А. Хорошилов. Компьютерная лингвистика и перспективные информационные технологии - М.: Русский мир, 2004.
5. Антопольский А. Б. Лингвистическое обеспечение электронных библиотек". - М., 2003. - 302 с
6. Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы. – М: Издательский центр «Академия», 2006

Дополнительная:

1. Левицкий Ю.А. Основы теории синтаксиса. – М.: URSS. 2005.

Введение. Лингвистические модели в современных информационных технологиях.

- 1. Автоматический перевод – исторически первая "лингвистическая" технология.**
- 2. Лингвистические модели в документальных информационных системах.**
- 3. Орфографические и грамматические корректоры.**
- 4. Лингвистические модели в системах распознавания письменных текстов и устной речи.**
- 5. Системы понимания (смыслового анализа и синтеза) текста.**

1. Автоматический перевод – исторически первая лингвистическая" технология.

1.1. Этапы развития

1.2. Основные проблемы

1.3. Методы и процедуры

1.1. Этапы развития

1) Начало:

- 1946: А. Бут & У. Уивер – формулировка задачи – перевод как дешифровка текста
- 1952: Первая конференция по АП. Возможен ли АП?
- 1954: Начало работы - Джорджтаунский эксперимент.
- 1954: Начало работ в СССР. - ИПМ; О. С. Кулагина

2) Конец 50-х – 60-е гг.: Осознание трудностей и проблем; разочарование и отступление.

3) 70-е – 80-е гг.: «Правильная осада».

4) С начала 90-х: Период коммерческих систем:

ProMT Translation Office - фирма ProMT;

Сократ – фирма «Арсенал»;

карманные переводчики: “Partner” - «Эктако»,

“Rocket_ProMT”; “СократСЕ”; “СловоЕд”

5) Прогноз:

2010-е гг.: литературный перевод деловых текстов.

1.2. Основные проблемы АП

1) Недостаточная полнота и точность описания языка в традиционной лингвистике

Основная причина – традиционная лингвистика ориентирована на человека, владеющего (родным) языком.

«Простой» пример - словоизменение:

Абзац – абзац

Конец - конц

Стрелец - стрельц

Читать - читаю

Спать - сплю

Быть – будь – будящий – буденный

... - сидящий ... - облущенный

Слово – часть речи = ?

А. А. Зализняк. Грамматический словарь русского языка. – М.: 1977 (Около 100 000 слов).

2) Лексическая многозначность

<i>Он открыл дверь своим ключом</i>	(1)	
<i>Дай мне ключ 12 на 18</i>	(2)	
<i>Он сумел подобрать ключ к шифру</i>	(3)	
<i>Ключ данных; ключевое поле</i>	(4)	(= <i>key</i>)
<i>Музыкальный ключ</i>		(= <i>klef</i>)
<i>В овраге бьют ключи</i>		(= <i>source, spring</i>)

предлог - I (отговорка) pretext, pretence; (повод) ground

предлог - II грам, preposition

предложение - I offer, suggestion

предложение - II грам, sentence;

(часть сложного предложения) clause

Омонимия частей речи:

- Light – 1) сущ.: = свет, лампа;*
2) прил.: = светлый, легкий;
3) глаг.: = освещать, зажигать;

3) Невозможность пословного перевода:

He is reading.

One has to wash his face every morning.

4) Неоднозначность значения грамматических форм

Пример – многозначность русского творительного падежа:

Письмо написано ребенком. («агентивность»)

Он ел икру ложкой. («инструментальность»)

Он ел икру дурОгой. («локативность»)

Он ел икру утром. («темпоральность»)

Он ел икру пудами. («квантитативность»)

Различие грамматического строя разных языков

- невозможность однозначного перевода однотипных грамматических конструкций:

Stone bridge □ $[A + N]$

Carriage return □ $[N + N_{Gen}]$

Power struggle □ $[N + P + N_{Acc}]$

Трудности перевода предлогов:

Самолет прибывает в 12:10.

The plain arrives at 12:10

Я не работаю в среду.

I don't work on Wednesday

В июле очень жарко.

It is very hot in July

Вы должны следовать за проводником.

You have to follow your guide / conductor (нет предлога)

NB: в путешествии – *guide*;

в вагоне – *conductor*;

в культуре и науке – *champion*.

Ср. также *под предлогом*:

- *under preposition* (грам.)

- *under / on the pretext of* (используя как повод)

Трудности перевода предлогов:

Самолет прибывает в 12:10.

The plain arrives at 12:10

Я не работаю в среду.

I don't work on Wednesday

В июле очень жарко.

It is very hot in July

Вы должны следовать за проводником.

You have to follow your guide / conductor (нет предлога)

NB: в путешествии – *guide*;

в вагоне – *conductor*;

в культуре и науке – *champion*.

Ср. также *под предлогом*:

- *under preposition* (грам.)

- *under / on the pretext of* (используя как повод)

То же с английского на русский:

The plain flew over the hill. (= над)

The dog jumped over the fence. (= через)

Шутки специалистов:

Дух силен, да плоть слаба.

... *The spirit is strong, but the flesh is weak*
Водка крепкая, а мясо протухло.

1.3. Методы и процедуры АП

1) Основные методы:

- Словарная поддержка; методы быстрого поиска в словаре;
(проблема создания больших словарей – 100 тыс. – 1 млн.)
- Формальные грамматики
(Совокупность точных правил, описывающих способы построения и грамматического разбора предложений и словосочетаний)
- Прецедентно-статистический
(*example-based, case-based*) анализ

TMS – Translation Memory System

2) Процедуры АП:

- 1) От письменного текста к словоформам
- 2) Определение возможных грамматических характеристик словоформы
- 3) Грамматический разбор – определение возможных синтаксических связей. [+ Прецедентный анализ]
- 4) Выбор переводного эквивалента словоформы - с учетом синтаксического контекста
- 5) [Построение смыслового представления; разрешение лексических и синтаксических неоднозначностей]
- 6) Перевод синтаксических связей на выходной язык
- 7) Морфологический синтез
- 8) Линеаризация

2. Документальные ИПС

Лингвистические проблемы:

- 1) Отождествление разных грамматических форм слова: *формат – формата - *формам*
- 2) Поисковая эквивалентность связанных по смыслу слов:
*Word -> текстовый процессор; стрела -> подъемный кран;
жесткий диск -> компьютер*
(словари-тезаурусы)
- 3) Смысловые связи между словами в тексте:
базы данных ? на базе данных моделей...

Расплодившиеся гусеницы уничтожили листву яблонь. Для борьбы с ними были использованы установленные на тракторе распылители пестицидов

Пример - для п.п. 2) и 3):

Деятельность английских разведчиков в немецких колониях во время 1-ой мировой войны.

Деятельность немецких разведчиков в английских колониях во время 1-ой мировой войны.

Британский шпионаж в германских владениях в 1915 -16 г.г.

3. Орфографические и грамматические корректоры

- 1) Орфографический контроль: словарь основ + моделирование словоизменения
- 2) Отсутствие орфографических замечаний не гарантирует от ошибок правописания:

Целью организации является...

*Пуск осуществляется крас**К**ой кнопкой.*

Частый детектив.

*Он не /**может**/ решит**Ь**ся на это.*

NB: Опасность команды «Заменить все»:

*Все купивш**И**Е путе**вк**И...*

*Ка**ж**дый купивш**И**Й путе**вк**У...*

3) Возможности грамматического корректора:

- пока ограничены!

-

- формальные проверки (лексические повторы, скобки...)
- проверка согласования (именные группы, подлежащее - сказуемое)
- проверка управления (ограниченно - предлоги и др.)
- узус

придает большое внимание:

Замечание корректора: *"Нарушение лексической сочетаемости.*

*Смешение двух устойчивых выражений "придавать значение"
"уделять внимание".*

Перспективы...

4. Распознавание письменного текста и устной речи

OCR (Optical Character recognition)

– от графического файла к текстовому.

Коммерческие системы:

- широкие возможности, высокое качество распознавания.

- **FineReader** (фирма **ABBYY**);
- **Cuneiform** (фирма **Cognitive Technologies**)

SR (Speech Recognition), TTS (Text To Speech) - распознавание устной речи)

Применение:

- голосовой ввод текста;
- "контакт -центры";
- протоколирование официальных заседаний;
- медицина; следствие и судопроизводство;
- военное дело и криминалистика;
- Voice Commander

Коммерческие системы:

- **Dragon** («Горыныч» – неудача русификации);
- *ViaVoice* (фирма IBM).

Основные ограничения: чтение отдельными словами; необходимость настройки на конкретного пользователя.

Пример коммерческого применения: "Контакт-центры":

- службы продажи билетов;
- справочные службы;
- службы технической поддержки

Исходный пункт – IVR-технологии (*Interactive Voice Response*)

Следующий шаг – технологии *Open Speech*

Функциональность и характеристики:

Разработчик TTS - *Nuance (ScanSoft)*;

интегратор в технологии контакт-центров "под ключ" – *Avaya*

- точность распознавания речи (заявленная) – 98%;
- способность к самообучению в процессе работы;
- определение начала и конца реплики собеседника;
- умение отличать паузы от завершения фразы;
- адекватная реакция на перебивание собеседником голосового сообщения системы;
- отсеивание посторонних шумов;
- "живой отклик" (*responsiveness*);
- распознавание на нескольких языках.

Лингвистическое содержание технологий распознавания:

1. Этап распознавания отдельных элементов (буквы в **OCR**, звуки или целые слова в **SR**)
2. Лингвистические фильтры:
 - орфографический;
 - грамматический;
 - семантический;
 - **Прецедентно-статистический.**

5. Системы понимания (смыслового анализа и синтеза) текста.

Что значит «понимать текст»?

- уметь отвечать на вопросы по тексту (ср. пример в #2)

В пустую комнату вошла высокая женщина, за руку которой держался маленький мальчик.

Сколько человек находится в этой комнате?

Диалоговые системы, имитирующие понимание:

- **ELIZA** - диалог с потенциальным пациентом от лица психоаналитика [Weizenbaum 1977];
- **A.L.I.C.E.** – см. в Интернет;
- конкурс Лёйбнера - проводится Университетом острова Флиндерс (Южная Австралия):

<http://www.loebner.net/Prizef/loebner-prize.html>

Если эксперт считает, что его собеседник некоторую часть времени «ведет себя как человек», а не как машина, то конкурсная оценка возможностей тестируемой системы возрастает.

Современная инженерная постановка задачи:

Information Extraction

Глобальные задачи:

- Автоматизированная справочная система с речевым интерфейсом («Служба 09»; «врач-консультант»; «юрист-консультант»; «шеф-повар-консультант» и т.п.).
- «Универсум знаний (оракул)» - интерактивная энциклопедия.

Справочная служба - основные процедуры:

- 1) Получить вопрос, убедиться, что он понятен системе (взаимодействие с подсистемами 2, 3 и 4), если необходимо – уточнить вопрос в диалоге с пользователем
- 2) Преобразование речевого вопроса в текстовую строку (*SR-технология*)
- 3) Определить релевантную БД (?? – глубокий смысловой анализ)
- 4) Построить по тексту вопроса SQL-запрос (смысловой анализ – *имеются экспериментальные системы*)
- 5) Выполнить запрос (технология реляционных СУБД)
- 6) Преобразовать ответ СУБД из табличной в текстовую форму (смысловой синтез - *имеются экспериментальные системы*)
- 7) Преобразовать текстовую строку ответа в звуковую речь (звуковой синтез - *технологии приближаются к коммерческому уровню*)

Т е м а 3. Автоматизация словарных работ и статистический анализ текста.

- 1. Типы словарей**
- 2. Частотные словари, их создание и использование.**
 - 2.1. Основные направления использования
 - 2.2. Технология создания частотного словаря средствами стандартных офисных технологий

#1. Типы словарей

1.1. По типу лексических единиц

- Лингвистические словари: объект описания – **слово**
(орфографические, толковые, переводные,...)
- Энциклопедические словари: объект описания – **понятие**
(ср. БСЭ, Политехнический словарь,)

1.2. По социальной функции

- Описательные
(синонимов, толковые, переводные,...)
- Нормативные
(орфографические, «Вместе или отдельно», словари рекомендуемых терминов)

1.3. По адресату

- Человекоориентированные
 - общего назначения;
 - для определенных социальных групп
- Машиноориентированные
 - = словарь, ориентированный на использование программами
 - обычно ***.DLL - библиотека**
 - (**NB:** НЕ то же, что «словарь на машинном носителе»!)

1.4. По широте охвата лексики

- Общеязыковые
- Специализированные
(словари имен, блатной лексики, словарь туриста, словарь Пушкина, словарь глагольного управления,...)

1.5. По числу языков

- Одно-, двух-, многоязыковые

1.6. По способу упорядочения

- Алфавитные (прямой и обратный порядок; упорядочение по концам слова - ср. грамматический словарь Зализняка)
- Частотные
- Идеографические
(Русский семантический словарь, Толковый словарь русских глаголов)

1.7. По типу описываемых языковых явлений

- Орфографические (правописание)
- Фонетические (произношение и ударение)
- Грамматические (словоизменение и грамматические характеристики словоформы)
- Частотные
 - общеязыковые;
 - для определенного стиля речи: научный, публицистический, разговорный, поэтический;
 - для определенной предметной области (научной дисциплины);
 - авторские, персонажей,...
- Исторические и этимологические (происхождение слов и развитие значений слова)
- Смысловые соответствия (межязыковые – переводные словари и внутриязыковые – толковые и энциклопедические словари)

- Словари сочетаемости

- словари глагольного управления,

- словари лексических функций (определяют способы нестандартного выражения заданного смысла для слова X):

- ✓ **Magn(X)** - *‘в большой степени X’*

X – молчание, Magn(X) – гробовое;

X – обыск, Magn(X) – тщательный;

X – дурак, Magn(X) – круглый;

X – негодяй, Magn(X) – отъявленный;

X – брюнетка, Magn(X) – ...?;

- ✓ **Real(X)** - *‘X осуществляется, происходит’*

X – победа, Real(X) – одержать;

X – поражение, Real(X) – потерпеть;

X – трагедия, Real(X) – пережить;

X – потрясение, Real(X) – ...?;

2. Частотные словари, их создание и использование.

2.1. Основные направления использования

- Изучение особенностей разных языковых стилей и языка социальных групп: (разговорный, литературный, профессиональный стили; подростковый жаргон, ...)
- Изучение авторского языка и стиля («язык Достоевского», «язык Солженицына» и т.д.)
- Выявление наиболее употребительной лексики ("русский как иностранный")
- Выявление основного содержания большого текста (технологии типа *Text Mining*)
- Ранжирование документов при поиске в Интернет (!)
- Изучение сочетаемости слов и терминов
- Разделение лексики на общеязыковую и специальную

2.2. Технология создания частотного словаря средствами стандартных офисных технологий

- 1) Преобразование текста в словник
- 2) Нормализация словоформ (метод сверток)
- 3) Импорт словника в таблицу БД *Access*
- 4) Создание частотного словаря (запрос *Access* с группировкой **по словам - сверткам**)
- 5) Оценка точности грамматической нормализации методом сверток
- 6) Анализ влияния грамматической нормализации на частотное ранжирование лексики
- 7) Разделение лексики на предметную и общеязыковую
- 8) Выделение понятий, выраженных двух- и трехсловными словосочетаниями.
- 9) Сравнение «ручной» и «частотной аннотаций». Подготовка отчета.

1) Преобразование текста в словник

□ Удаление ненужных символов -
использование режима *Подстановочные знаки* - поиск и
замена по шаблону (использование регулярных
выражений); *заменить* на пустую строку или пробел):

? – любой одиночный символ (к?t)

* – любое число любых символов (к*t)

[ио] – любой из указанных символов (к[ио]т)

[A-z] – любой символ в указанном диапазоне

^# – любая цифра

...

□ Преобразование в формат «все словоформы в 1 столбец» -
замена пробелов символом «конец абзаца»

Найти и заменить



Найти

Заменить

Перейти

Найти:

Заменить на:

Меньше ▲

Заменить

Заменить все

Найти далее

Отмена

Параметры поиска

Направление:

- Учитывать регистр
- Только слово целиком
- Подстановочные знаки
- Произносится как
- Все словоформы

Заменить

Формат ▼

Специальный ▼

Снять форматирование

2) Нормализация словоформ (метод сверток)

Алгоритм построения свертки:

- в свертку включаются три первые буквы слова;
- в свертку включаются остальные согласные буквы слова;
- при свертывании отбрасываются конечные буквы *в, з, м, х* (если они входят в окончания существительных и прилагательных)

Макрос *ДобавитьСвертки2*

Примеры:

Словоформа	Свертка
<i>анализ</i>	<i>аналз</i>
<i>анализа</i>	<i>аналз</i>
<i>анализе</i>	<i>аналз</i>

Словоформа	Свертка
<i>данная</i>	<i>данн</i>
<i>данной</i>	<i>данн</i>
<i>даннОМ</i>	<i>данн</i>
<i>даннОМУ</i>	<i>данн</i>
<i>даннЫХ</i>	<i>данн</i>

5) Оценка точности грамматической нормализации методом сверток

Ошибки:

- a) *богослОВ, женИХ, запАХ, систЕМУ*
- b) *конЬ – кон; потОМУ – пот; косОМУ – косА - косоЙ;
comОМУ – comЫ*
- c) *нерегулярность словоизменения:*

Словоформа	Свертка
<i>времЕНИ</i>	<i>времН</i>
<i>время</i>	<i>врем</i>
<i>время</i>	<i>врем</i>

Всего просмотрено слов (существительных и прилагательных) – **М**

Из них имеют не совпавшие свертки – **Н**

Точность отождествления = **N/M**

В примере: N=1, M=3; точность = 1/3 (~30%)

6) Анализ влияния грамматической нормализации на частотное ранжирование лексики:



ЧастСловарьНаСвертках : таблица

Словоформа	Свертка	Частота	РангСврт	ИзмРанга
в	в	115	1	0
и	и	109	2	0
семантическая	семнтчск	63	3	9
словаре	словр	42	4	9
язык	язык	40	5	6
как	как	34	6	3
не	не	30	7	3
единиц	единц	28	8	6
с	с	28	8	3
это	это	28	8	6
быть	быть	27	9	3
для	для	26	10	3
на	на	26	10	3
что	что	25	11	4
или	или	23	12	4
логическая	логчск	22	13	6
понятие	понт	21	14	6
к	к	20	15	5
описание	описн	19	16	3
предлог	предлг	19	16	0
представлен	предствлн	19	16	0
так	так	19	16	2
текст	текст	19	16	2

Запись: 24 из 269

ЧастСловарьнаСловофрм ...

Словоформа	Частота	РангСлвф
в	115	1
и	109	2
как	32	3
не	30	4
с	28	5
быть	27	6
что	26	7
для	26	7
на	26	7
словаря	24	7
или	23	8
к	20	10
языка	18	11
семантического	15	12
может	14	13
словарь	14	13
это	13	14
но	13	14
единицы	13	14
текста	13	14
при	12	15
...	12	15
семантическом	7	16

Запись: 11 из 100

7) Разделение лексики на предметную и общеязыковую

Словоформа	Свертка	Частота2
так	так	49
которая	котр	30
результат	резльтт	17
процесс	процсс	11
исследование	исслдвн	7
компонент	компннт	4
фасет	фаст	2
инсоляция	инслц	1

8) Выделение понятий, выраженных двух- и трехсловными словосочетаниями.

Словоформа1	Словоформа2	Словоформа3	Частота
И	Т	П	6
В	логическом	представлении	5
отношений	между	понятиями	4
формализованного	семантического	словаря	3
с	точки	зрения	3
морфологический	и	синтаксический	3
И	правил	вывода	3
В	виду	что	3

Т е м а 1.

Язык как система. Уровни описания языка.

- 1. Функциональное описание языка (функциональные стили)**
- 2. Внутрисистемное описание языка**

1. Функциональное описание языка (функциональные стили)

1.1. Деловая речь (функция деловой коммуникации)

1.2. Поэтическая речь (функция построения
художественного образа)

1.3. Другие функции языка:

- Выражение эмоций (Смысл не важен – важна интонация...)
- Ритуал (напр., церковная служба)

1. Символизация принадлежности к социальной группе

[*Разговорная речь* – комплексный феномен]

1. Функциональное описание языка

Погруженность языка в культуру.

(Почему вдруг появилась масса англицизмов?)

Что есть культура?

П. Сорокин: *Совокупность значений, ценностей и норм, которыми владеют взаимодействующие лица, и совокупность носителей, которые объективируют, социализируют и раскрывают эти компоненты.*

Основные компоненты культуры:

- 1) понятия и знания, упорядочивающие социальный опыт (когнитивная и технологическая подсистема);
- 2) идеалы, ценности (подсистема мотивации);
- 3) нормы и образцы (подсистема регуляции поведения);
- 4) материальная среда (орудия и продукты человеческой деятельности - артефакты, обеспечивающие физическое существование социума);
- 5) эстетическая среда (подсистема гармонизации социальной жизни) .

Проявления культуры в языке:

Откуда взялись выражения:

Много воды утекло.

Ваше время истекло.

Бить склянки.

Клепсидра в суде.

Песочные часы.

Еще примеры:

магазин "*Старик Хоттабыч*";

"Итого!..."

Культура как диалог культурных феноменов

В.Высоцкий:

Наконец-то нам дали приказ наступать,

Отбирать наши пяди и крохи,

Но мы помним, как Солнце отправилось вспять

И едва не зашло на востоке.

«Чужой земли не хотим, но и своей земли ни пяди не отдадим.»

Сталин (?)

В зависимости от того, где и для какой цели используется язык, Меняются лексика, формы языковых выражений и их характеристики.

С этой точки зрения можно различать **функциональные стили речи**

Важность правильной оценки **коммуникативного намерения** говорящего / пишущего:

На школьном уроке.

Учитель: *В каком году родился Пушкин?*

Ученик: *Мне бы Ваши заботы, господин учитель!*

На улице.

- *Не знаете ли Вы, который час?*

- *Знаю.*

1.1. Деловая речь (функция деловой коммуникации)

А.П. Ершов: «Феномен деловой прозы» (1978 г.): “деловая проза – это языковой носитель производственных отношений”.

Основные характеристики.

- 1) Точность и однозначность выражения;
- 2) Стандартность, единообразие способов выражения мысли – каждое понятие должно выражаться всегда одним и тем же способом, употребление синонимов не приветствуется;
- 3) Неметафоричность – использование слов (терминов) только в их прямом и буквальном смысле;
- 4) Смысл слова (термина) должен быть определен и известен до его употребления в тексте – ср. энциклопедические словари;
- 5) Максимальная смысловая полнота; отсутствие намеков и недоговоренностей; апелляция только к профессиональным знаниям читателя, но не к воображению, не к ассоциациям;
- 6) Логическая последовательность, непротиворечивость.

Примеры деловой прозы

УК РФ, ст. 219:

“1. Нарушение правил пожарной безопасности, совершенное лицом, на котором лежала обязанность по их соблюдению, если это повлекло по неосторожности *причинение* тяжкого или *средней тяжести* вреда здоровью человека, -

наказывается штрафом в размере от 100 до 200 МРОТ...

2. То же *деяние*, повлекшее по неосторожности смерть человека или иные тяжкие последствия —

наказывается *ограничением свободы* на срок до 5 лет или *лишением свободы* на срок до 10 лет...”

СНИП «Жилые здания»:

2.9. Жилые комнаты общежитий *следует* проектировать из расчета заселения не более трех человек при площади не менее 6,0 кв. м. на каждого проживающего. Комнаты *должны быть* непроходными, шириной не менее 2.2 м., их *следует* оборудовать встроенными шкафами площадью не менее 0.5 кв. м. на каждого проживающего.

Научно-технический текст:

Увеличение напряжения текучести может быть объяснено за счет *вакансий*, которые имеют заметную подвижность при -72°C .

Образование илистого осадка зависит от размера частиц, полученных механическим способом, количества металла в амальгаме и чистоты растворов.

Деловая проза – это профессиональные подъязыки - со своей лексикой, своими характерными оборотами речи и грамматическими конструкциями.

Например:

- юридический: *причинение ...средней тяжести вреда...*
- военный: *открыть огонь на поражение*
- морской: *на флоте*
- математический: *для...необходимо и достаточно...*
- приказ по вузу: *зачислить на 1-ый курс (а не принять)*

И т. д.

Чем ответственнее сфера деятельности, тем более стандартизован язык.

Чем беднее словарный состав делового текста – тем лучше текст!

Кто не владеет профессиональным языком – тот не профессионал!

1.2. Поэтическая речь (функция построения художественного образа)

Поэтическая речь = художественное слово «в чистом виде»

Основные характеристики:

) Многовариантность выражения одного и того же смысла.

Так думал молодой повеса,

Летя в пыли на почтовых,

Всевышней волею Зевеса

Наследник всех своих родных.

Друзья Людмилы и Руслана!

С героем моего романа

Без предисловий, сей же час

Позвольте познакомить вас:

Онегин, добрый мой приятель,

...

Ребенок был резов, но мил

...

Чтоб не измучилось дитя

...

2) Смысловая неполнота как принцип – говорится всегда больше, чем сказано; обращение не только к культурно-энциклопедическому контексту, но и к эмоциональной сфере;

3) Метафорическое использование слова:

Поэты ходят пятками по лезвию ножа

И режут в кровь свои босые души. [В. Высоцкий]

4) Смысл слова может создаваться или изменяться контекстом его употребления:

Пинь-пинь-пинь – тарарахнул зензивер

[В. Хлебников]

Плывет в тоске необъяснимой

*Среди кирпичного **надсада***

Ночной кораблик негасимый

Из Александровского сада. [И. Бродский]

Ожегов: *надсада* = чрезмерное усилие; ж.р. (+ орфографическ. сл.)

5) Логическая непоследовательность – вплоть до абсурдности
буквального смысла и прямого логического противоречия.

Речка движется и не движется,

•••

Песня слышится и не слышится

Поэзия не ставит целью **информировать**, описывать *положение дел* – даже если кажется, что о чем-то сообщает:

*Ночь. Улица. Фонарь. Аптека.
Бессмысленный и тусклый свет.*

(О чем? О плохом освещении улиц?...)

*Живи еще хоть четверть века –
Все будет так. Исхода нет.*

(От внешнего к внутреннему...)

Еще:

*В тот год осенняя погода
Стояла долго на дворе,
Зимы ждала, ждала природа,
Снег выпал только в январе.*

(NB: Роман в стихах – но отнюдь не метеосводка!)

*О чем твои стихи? Не знаю брат.
Ты их прочти, коли придет охота.
Стихи живые сами говорят,
И не о чем-то говорят, а что-то.*

[С. Маршак]

И еще о том же:

*Быть знаменитым некрасиво,
Не это поднимает ввысь.
Не надо заводить архива,
Над рукописями трястись.*

*Цель творчества – самоотдача,
А не шумиха, не успех.
Позорно, ничего не знача,
Быть притчей на устах у всех.*

(Пока звучит почти как речь на съезде писателей.

Но дальше:)

*Но надо жить без самозванства,
Так жить, чтобы в конце концов
Привлечь к себе любовь пространства,
Услышать будущего зов.*

*И надо оставлять пробелы
В судьбе, а не среди бумаг,
Места и главы жизни целой
Отчеркивая на полях.*

...

*Другие по живому следу
Пройдут твой путь за пядью пядь,
Но пораженья от победы
Ты сам не должен отличать.*

*И должен ни единой долькой
Не отступиться от лица,
Но быть живым, живым и только,
Живым и только до конца. [Б. Пастернак, 1956.]*

В чем назначение поэзии (и искусства вообще)?

А. Блок:

“Цель поэзии – внести гармонию в мир.”

Что может и что не может быть объектом информационных технологий?

2. Внутрисистемное описание языка

2.1. Язык как текст и как порождающая система.

2.2. Уровни описания языка:

- Фонетический
- Морфологический
- Синтаксический
- Семантический

2.1. Язык как текст и как порождающая система

Что такое «язык»?

Два возможных ответа: **Текст VS Словарь**

- 1) **Как непосредственная данность** – вся совокупность текстов, написанных (произнесенных) на данном языке – “корпус текстов”

Задача дешифровки:

от текстов к единицам и правилам построения;

(шумерские глиняные таблички; иероглифические книги майя;

- От последовательности символов к смыслу
- Реконструкция звуковой формы (?)

2) **Язык как система правил**, описывающих способ порождения текста:

- список элементов (слов) и их классификация – с точки зрения сочетаемости в тексте (словарь);
- способы их соединения в тексте (грамматика);
- [способы их соотнесения с действительностью (семантика)]

Два типа отношений в языке –

синтагматические и парадигматические

<i>красный</i>	-	<i>шар</i>
<i>синий</i>	-	<i>куб</i>
<i>медный</i>	-	<i>стержень</i>
<i>мягкий-</i>		<i>карандаш</i>
...		...

2.2. Уровни описания языка

2.2.1. Фонетический уровень

Фонетика – раздел языкознания, изучающий звуковые единицы языка, их сочетаемость в потоке речи и их позиционные изменения.(РГА)

Единицы – фонемы (звуки) / буквы

Классификация (парадигматика):

- согласные
 - звонкие – глухие;
 - твердые – мягкие;
 - ...
- гласные (АеёИОУЫІЭюя)

Два звука считаются разными в данном языке, если замена одного другим может вызвать изменение смысла.

Примеры:

кот – год; пот – бот; (глухой - звонкий)

мэрский – мерзкий; забор – собор - запор (твердый - мягкий)

Кто счастлив, тот смеется,

Кто хочет, тот доБьётся,

Кто ищет, тот всегда найдет.

(Дунаевский, Лебедев-Кумач)

Кто хочет, тот доПьётся...

(Венедикт Ерофеев. Москва – Петушки.)

Отношения (синтагматика) - линейное следование фонем в слове.

Возможность разных звукосочетаний и их характерность в языке:

- 3 гласных (длинноее, радиоузел, змеопасный,...);
- 4 согласных (вздрогнуть, встряхнуть, президентский,...);
- 5 согласных (бодрствовать, контрпример, фландрский,...).

Характерные для разных языков звукосочетания:

Русский: город, корова, пароход, беда, ...

Польский: Ястржембский, Збигнев Цибульский, Щецин,...

Английский: Джонсон, Кеннеди, Доусон, Черчилль...

Маяковский: "Вор нагл драл с лип жасмин"

Китайский: дао, Мао, Цзы, Цин, Мяо...

На русский текст объемом ~ 10 п.л.:

АО, ИУ – по 1 слову (наоборот, социум)

УИ – 2 слова (конструировать, градуировать)

ЯО - 0 слов;

ИЕ - 1600 раз, **РО**- 2244 раза!

NB: Важно для технологий **SR** и **OCR**! (Набор **SMS**-сообщений)

Задачи моделирования:

- 1) Распознавание устной речи и письменного текста (***SR*** и ***OCR***);
- 2) Звуковая дешифровка древних языков

2.2.2. Морфологический уровень

Морфология – раздел языкознания, изучающий словообразование и словоизменение, а также способы выражения грамматических признаков в пределах словоформы.

Словоформа – слово, взятое в определенной морфо-грамматической форме.

Слово – совокупность всех своих словоформ:

дом, домаа, домуу, домом, домое, домаа(столы), домов, домами, домах.

Единицы – морфемы (= морфы) (Бодуэн де Куртенэ, 1881)

«Морфема – минимальная значимая часть словоформы»

Классификация морфем - по месту в слове и функции:

- корень (основное предметное значение слова);
- основные словообразовательные аффиксы
(приставка, суффикс);
- прочие словообразовательные аффиксы -
 - интерфиксы (соединительные гласные *-о-, -е-, -и-*):
пароход, овцебык, дозиметр,...
 - префиксоиды:
анти-, полу-, много-, высоко-, двух-, кило-, милли-, авто-, авиа-, гео-, вибро-, метео-, гидро-, теле-,...
 - суффиксоиды:
-метр, -скоп, -трон, -мобиль, -стат, --фон, -граф, -лиз, -навт, -лог, -логия, -фил,...
- словоизменяющие аффиксы -
 - окончания:
домами, красных, читают, работать,...
 - постфлексии:
давайте, смеется, какой-то,...

Семантика морфем

А) Словоизменительные

Основная смысловыражающая функция – указывают на существование смысловой связи между словами.

Предметная семантика грамматических категорий

- иногда есть, иногда нет.

- число: *столы, книги, войны, ...;*
НО *вилы, брюки, носилки, ...*
- род: *мама, девочка, баран, ...;*
НО - *щука, сом, гусеница, врач, день - ночь,*
стул - табуретка, ...;
побежал – побежала...
- падеж: *послал другу (адресат);*
НО *горные породы разрушаются водой*
vs горные породы разрушаются весной
- время: время действия относительно момента речи,
либо время одного действия относительно другого в тексте:
Так думал молодой повеса, летя в пыли на почтовых, ...

Б) Словообразовательные

Можно выделить типовые значения:

- уменьшительные и увеличительные:

ларчик, хлебец, книжечка, медвежонок, котище,...

- принадлежность лица профессии, месту, виду деятельности и т. п.:

летчик, невец, писатель, европец, картежник,...

- характеристика по форме и материалу:

металлический, стеклянный, конический

НО:

ср. суффиксы **-ов-** и **-н-** :

крокодиловые сапоги...;

научный, яблочный, компьютерный,...;

отделочные материалы.

Отношения: (синтагматика) - линейное следование морфем в слове (сочетаемость морфем).

Для окончаний –

описание сочетаемости с основой

через отнесение к *флексивному классу* (~ 1000).

РГА: «К 1-му склонению относятся существительные муж. р. с нулевой флексией в им. п. ед. ч. и сред. и муж. р. на *-o, -e, -ё*»
стол, дом, договор, сахар, купец, ...

Список возможных окончаний (муж. р., «твердое склонение»):

И.: (=0)

Р.: *-a (-y)*

Д.: *-y*

В.: *как И. или Р.*

Т.: *-ом*

П.: *-e(-y)*

Мн. ч.: *-ы(-а), -ов, -ам, -ами, -ах.*

Однако сочетаемость, например, основы и суффикса нормировать много труднее:

*ротор – ротор**Н**ый,*

*куб – куб**ИЧЕСК**ий*

*контур – контур**Н**ый,*

НО *шар – шар**О**вой;*

НО *зуб - зуб**Н**ой; куб**О**вая краска*

НО *шнур – шнур**О**вой;*

Задачи моделирования:

- 1) Отличить словоформу от случайного набора букв/звуков
(обнаружение ошибок правописания, **OCR**- и **SR**-технологии).
- 2) Отождествление (по смыслу) разных форм
одного и того же слова (**ДИПС**).
- 3) Отождествление основного предметного смысла (корня) в
разных словах: *ротор* – *роторный* – *двухроторный*, ... (**ДИПС**).

НО: *лев*, *левый*, *королевский*. ("Борис, ты не прав, ты лев!")

войска □ *савойская капуста*

4) Определение по основе и окончанию грамматических характеристик словоформы для автоматического грамматического разбора предложения (**АП**, грамматические корректоры и фильтры в системах распознавания речи, анализ смысла текста):

дому – дат. п. ед. ч.; *дома* – род. п. ед. ч., им. и вин. п. мн. ч.

5) Определение смысла слова по смыслу составляющих его морфов: *дом, бездомный, домовладелец, домовой, домосед,...*;
НО: *аванс(-?)цена, полу(-?)чили, шок(-о?)лад, кол(-?)лекция,...*

Т е м а 2.

Синтаксическая структура предложения и способы ее формализации.

- 1. Синтаксис простого предложения**
- 2. Межсегментные связи [Слайд 71](#)**
- 3. Анафорические связи**
- 4. Синтаксическая омонимия**
- 5. Представление синтаксической структуры в
алгоритмах и программах**
- 6. Формальные свойства синтаксической структуры**

1. Синтаксис простого предложения

Синтаксическая связь – смысловая связь между словами,
выраженная грамматическими средствами.

Синтаксис – учение о грамматически допустимых способах
соединения слов и предложений.

Единицы – слова и устойчивые словосочетания (фразеологизмы).

Отношения – синтаксические связи.

NB: не обязательно следование в тексте!

1.1. Виды синтаксической связи.

Синтаксические связи:

о Сочинительные - Соч

о Подчинительные:

- согласование - С

 - полное

 - неполное

- управление - У

 - УП – Управление от Предлога

 - (УС) – Управление с Согласованием

 - (связь сказуемое - подлежащее)

 - сильное – слабое

 - беспредложное – предложное

- примыкание - П

Сочинительная связь –

связь между словами, имеющими одну и ту же грамматическую и/или смысловую функцию в предложении (однородные члены предложения),

либо между грамматически равноценными предложениями.

Ср.: *Передайте конверт тайно и в нужные руки.*

Имеет обязательное грамматическое выражение посредством:

- а) сочинительных союзов;
- б) знаков препинания (, ;)

Ср.: *Красный, синий и зеленый шары.*

Большой желтый медный шар. – нет показателей сочинения!

Подчинительная связь –

связь, при которой одно слово ("зависимое", "слуга") уточняет, конкретизирует смысл, выраженный другим словом ("главное", "хозяин").

Может связывать отдельные слова, либо предложения.

Обнаруживается **носителем языка** (человеком, понимающим текст) путем постановки уточняющего вопроса от главного слова к подчиненному.

Ср.: *Белеет*

что? - парус

какой? - одинокий

в чем? - в тумане

чего? - моря

каком? - голубом.

Компьютерная программа (алгоритм) обнаруживает синтаксические связи, учитывая всю совокупность грамматических показателей, – в частности, учитывая порядок слов, расстояние между словами и их грамматические формы.

Подчинительные связи делятся по степени обязательности на сильные и слабые.

Связь считается **сильной**, если при отсутствии подчиненного слова носитель языка воспринимает словосочетание/предложение как незавершенное: *стать специалистом; обработка информации; можно продать; продать рукопись.*

Связь считается **слабой**, если при отсутствии подчиненного слова носитель языка воспринимает словосочетание/предложение как завершенное: *новый дом; дом отца; дом у дороги; дом построен, дом для престарелых.*

Чаще всего это различие применяется к связям управления.

Градации: от абсолютной обязательности (предлоги, фазовые и модальные глаголы, переходные глаголы) до полной необязательности (прилагательное + существительное, конструкции с предлогом *для* и др.)

Согласование –

подчинительная связь, при которой грамматическая форма зависимого слова **уподобляется** грамматической форме главного слова.

Тест: при изменении формы главного слова меняется и форма подчиненного.

Характерно для флективных языков (ср. русский и английский).

Полное согласование – совпадение всех грамматических признаков, характеризующих одновременно оба слова:

утомленное солнце – род, число, падеж;

белеет парус – число;

белел парус – число, род (*белела лодка*);

Неполное согласование – совпадение лишь части грамматических признаков, характеризующих оба слова:

город Москва; врач Петрова;

синий и красный шары (только число!)

Управление –

подчинительная связь, при которой для выражения определенного смыслового отношения главное слово требует постановки зависимого слова в определенную падежную форму и/или определенного предложного оформления (*ср. русский и англ.*)

Тест: при изменении формы главного слова форма подчиненного не меняется.

Утомленные солнцем VS утомленное солнце

Ср. также:

приближающийся к солнцу; удаляющийся от солнца;

обязанный солнцу; лишенный солнца; помещенный на солнце

– для выражения разных смысловых отношений, как правило, должны использоваться разные падежно-предложные формы.

Список падежных форм, которые могут быть подчинены данному слову, называют его (синтаксической) **моделью управления**.

Исчерпывающее описание модели управления можно дать только для предлогов:

К - дат. (*к вам и квас*)

ДЛЯ - род.

НА - вин., предл. (*на улицу, на улице*)

С - род., вин., твор. (*с крыши, с версту, с другом*)

NB: выбор предлогов и падежей трудная сторона грамматики !

Актуально для глаголов и отглагольных существительных (обычно перечисляются только сильноуправляемые позиции):

СООБЩАТЬ (*кто: [им.]; кому: [дат.]; что: [вин.]/о чем:[предл.];*)

ГРОЗИТЬ (*кто: [им.]; кому: [дат.]; чем: [твор.];*)

Не актуально для существительных с предметным значением
(сильноуправляемые позиции отсутствуют):

КНИГА – (*- жалоб, - для детей, - с иллюстрациями, - в переплете*).

Для знаменательных слов модель управления определяется прежде всего **смыслом** управляющего слова.

Поэтому преобладает точка зрения, что для них модель управления – явление скорее семантического уровня.

Примыкание –

подчинительная связь, при которой зависимое слово обычно не изменяемо и установление связи обусловлено лишь взаимным расположением (расстояние и порядок) и/или смыслом связываемых слов.

Присоединяемые части речи:

- наречия (*говорить медленно*)
- частицы (*Он не спит*)
- неизменяемые глагольные формы - инфинитив, деепричастие: (*любил рисовать; спал стоя*)
- неизменяемые прилагательные, компаративы (*костюм хаки; стал умнее*)

Различают контактное и дистантное (падежное) примыкание.

Контактное выражается расположением слов (расположение значимо):

- (а) *Они **вместе** заявили о невозможности работать.*
*Они заявили о невозможности работать **вместе**.*
- (б) *Он **быстро** научился решать эти задачи.*
*Он научился **быстро** решать эти задачи.*

Дистантное выражается лишь смыслом связываемых слов – может быть реализовано на большом расстоянии и при разном порядке слов:

*Он **приедет** из Москвы с двумя сотрудниками поездом «Русская тройка» **завтра** **вечером**.*

1.2. Табличное представление синтаксических связей

Белеет парус одинокий в тумане моря голубом.

Слово	N	Хоз	Связь
<i>Белеет</i>	1	0	-
<i>парус</i>	2	1	УС
<i>одинокий</i>	3	2	С
<i>в</i>	4	1	У
<i>тумане</i>	5	4	УП
<i>моря</i>	6	5	У
<i>голубом</i>	7	5	С

1.3. Средства выражения синтаксической связи

1. Морфологические показатели (род, число, падеж, лицо)
2. Служебные слова (союзы, предлоги)
3. Знаки препинания (наличие и отсутствие)
4. Расстояние
5. Порядок слов

NB: Разная значимость их в разных языках!

Разные стили мышления!

2. Межсегментные связи

2.1. Виды сегментов и способы установления синтаксических связей между ними

Сегмент – любой компонент предложения, для которого правила пунктуации предусматривают фиксацию их границ при помощи знаков препинания. [Т. Ю. Кобзарева]

2.1.1. Виды сегментов:

- 1) Простые предложения в составе сложных
 - а) связь сочинения;
 - б) связь подчинения;
- 2) Обособленные обороты (неполные предложения)
 - а) причастные;
 - б) деепричастные;
 - с) вводные обороты;
- 3) Обособленные члены предложения

2.1.2. Синтаксические связи между сегментами

1а) Простые предложения в составе сложносочиненного:

- Сочинительной связью соединяются главные слова (сказуемые) простых предложений
- Каждое последующее предложение подчиняется предыдущему

Слово	N	Хоз	Связь
...			
<i>Сказуемое 1-го</i>		0	
...			
&			
...			
<i>Сказуемое 2-го</i>		Сказ-1	МССоч
...			

Пример:

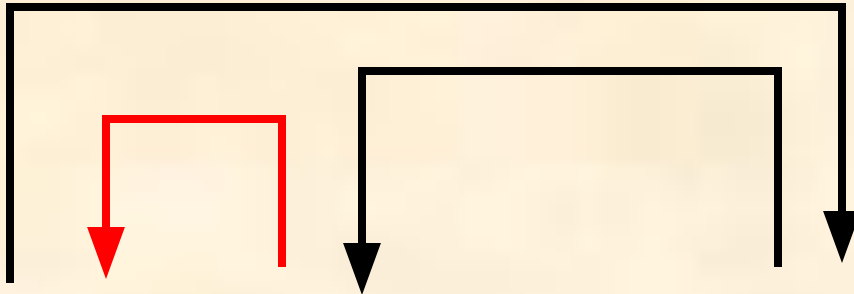
Не продается вдохновенье, но можно рукопись продать.

СЛОВО	N	Х03	СВЯЗЬ
<i>Не</i>	1	2	П
<i>продается</i>	2	0	-
<i>вдохновенье</i>	3	2	УС
<i>но</i>	4	0	-
<i>можно</i>	5	2	МССоч
<i>рукопись</i>	6	7	У
<i>продать</i>	7	5	П

[Назад](#)

1b) Простые предложения в составе сложноподчиненного.

1b-1) Подчинительный союз (союзное слово) является членом придаточного предложения:



А это синица, которая часто ворует пшеницу...

Устанавливаемые синтаксические связи.

- 1) Союзное слово подчиняется одному из слов в придаточном (обычно сказуемому): **ворует** –кто? □ **которая**
- 2) Придаточное предложение (его главное слово) подчиняется определяемому слову главного:
синица –какая? □ **ворует** (которая)
- 3) Если союзное слово является анафором, устанавливается **анафорическая связь** между союзным словом и определяемым главного: **которая** = **синица**

Примеры:

1. *Кирджали представлен был наше, который присудил его быть посажену на кол.*
2. *Кто жил и мыслил, тот не может в душе не презирать людей.*
3. *Легко на сердце от песни веселой, она скучать не дает никогда.*
4. *Пока не требует поэта к священной жертве Аполлон, в заботах суетного света он малодушно погружен.*
5. *Высокой страсти не имея для звуков жизни не щадить, не мог он ямба от хорея, как мы ни бились, отличить.*

СЛОВО	N	Хоз	СВЯЗЬ
<i>Кирджали</i>	1	3	УС
<i>представлен</i>	2	3	С
<i>был</i>	3	0	-
<i><u>наше</u></i>	4	2	У
<i><u>который</u></i>	5	<u>6</u>	УС
<i><u>присудил</u></i>	6	<u>4</u>	МС_П
<i>его</i>	7	6	У
<i>быть</i>	8	6	П
<i>посажену</i>	9	8	У (С)?!
<i>на</i>	10	9	У
<i>кол</i>	11	10	УП
<i>который</i>	5	4	Анаф

1b-2) Подчинительный союз (союзное слово) НЕ является членом придаточного предложения

Все знают, что Волга впадает в Каспийское море

Устанавливаемые синтаксические связи.

1) Придаточное предложение (его главное слово) подчиняется присоединяющему союзу:

что –(_?) □ впадает

2) Подчинительный союз (представляя все придаточное) подчиняется определяемому слову в главном предложении:

знают –что? □ что (Волга впадает...)

Сравнить:

Мы не знали, что он привез. (случай 1b-1)

Мы не знали, что он привез письмо. (случай 1b-2)

Пример: *Я это потому пишу, что сам давно уж не грешу.*
 (потому указательное местоимение, сигнал о продолжении...)

Слово	N	Хоз	Связь
<i>Я</i>	1	4	УС
<i>это</i>	2	4	У
<i><u>потому</u></i>	3	4	П
<i>пишу</i>	4	0	-
<i><u>что</u></i>	5	<u>3</u>	МС_П
<i>сам</i>	6	10	УС
<i>давно</i>	7	10	П
<i>уж</i>	8	10	П
<i>не</i>	9	10	П
<i><u>грешу</u></i>	10	<u>5</u>	Пусто

2а) Причастные обороты (связь типа согласования):

Дверь, ведущая на второй этаж, оказалась заперта.

Устанавливаемые синтаксические связи.

- 1) Все слова причастного оборота прямо или косвенно подчинены причастию: **ведущая** –на что? □ **этаж**
- 2) Причастие подчиняется связью **МС_С** определяемому слову главного: **Дверь** –какая? □ **ведущая**

2б) Деепричастные обороты (связь типа примыкания):

Служив отлично-благородно, долгами жил его отец.

Устанавливаемые синтаксические связи.

- 3) Все слова деепричастного оборота прямо или косвенно подчинены деепричастию.
- 4) Деепричастие подчиняется связью **МС_П** сказуемому главного предложения: **жил** –как? □ **служив**

2с) Вводные обороты и вставные конструкции (связь типа примыкания):

По всем приметам, будет жаркое лето.

Устанавливаемые синтаксические связи.

Вводный оборот подчиняется (формально) сказуемому главного предложения.

3) Обособленные члены предложения:

А он, мятежный, ищет бури...

Устанавливаемые синтаксические связи.

Подчинение не отличается от подчинения не обособленных.

3. Анафорические связи

1) **Анафор** – служебное слово, не имеющее самостоятельного значения и отсылающее к полнозначному слову или выражению. (Значение ‘отсылка к предыдущему имени’ входит в смысл анафорического элемента.)

Анафоры: личные (*он, они, ...*) и указательные (*это, этот, ...*) местоимения; союзные слова (*который, поэтому, так, ...*).

2) **Антецедент** – полнозначное (знаменательное) слово или целое предложение, замещаемые анафором.

Отношение **замещения**: анафор можно заменить антецедентом без изменения смысла текста.

Отношение между языковыми выражениями, состоящее в том, что в смысл одного выражения входит отсылка к другому. (ЛЭС)

А) Вербальная анафора – замещается отдельное слово
(существительное).

синица, которая...

Грамматически – связь типа (неполного) согласования – в роде и числе.

Представление: направленной связью с пометой *Анаф.*

Б) Сентенциальная анафора – замещается целое предложение.

*Волга впадает в Каспийское море. Это известно каждому.
...когда же черт возьмет тебя. Так думал молодой повеса...*

Согласование отсутствует, связь явно не фиксируется.

Катафорическая связь:

Ясно одно: я должен уехать.

4. Синтаксическая омонимия

1. Виды синтаксической омонимии:

- Реальная – формальная
- Локальная - глобальная
- Омонимия адреса - содержания

2. Омонимия разных видов связи:

- Омонимия подчинительных и сочинительных связей
- Омонимия анафорических связей
- Омонимия межсегментных связей

Явление, состоящее в том, что синтаксические связи в предложении могут быть установлены или грамматически описаны несколькими альтернативными способами.

Влечет за собой, как правило, и смысловую неоднозначность.

Л.Н. Иорданская:

"Под синтаксически омонимией (неоднозначностью) фразы понимается возможность приписать фразе более чем одну правильную синтаксическую структуру. Естественно, эта возможность зависит от того, как определена правильная синтаксическая структура и, в частности, от того, как проведена граница между синтаксической и семантической характеристиками фразы."

*Человек стрелял **из ружья** vs Человек стрелял **из окна***

Лесник присматривает за не особенно аккуратными туристами.

- "выделяющая" интерпретация;
- "квалифицирующая" интерпретация.

Реальная – формальная омонимия

Реальная:

Он из Германии туманной привез учености плоды.

Наблюдения над языком маленьких детей.

Формальная: Обнаруживается, если устанавливать синтаксические связи без учета смысловых характеристик слов и / или контекста целого предложения

Лифты для высотных зданий со скоростью 30 м/мин.

парень в голубой рубашке с зелеными глазами

парень в голубой рубашке с зелеными рукавами

Возьмите деревянный брусок с отверстием диаметром 30 мм.

Возьмите деревянный брусок с отверстием весом 300 г.

"Временная" омонимия – возникает и разрешается по ходу анализа:

Мальчишек радостный народ коньками звучно режет лед.

Еще примеры:

Простой солдат вызвал суматоху.

Маркизу нельзя есть руками.

а) нельзя --(кому?)--> маркизу

б) есть --(кого? что?)--> маркизу

Локальная – глобальная омонимия

Локальная: Выбор одной из альтернативных связей для данного слова не влияет на установление связей между другими словами предложения

Глобальная: Выбор одной из альтернативных связей для одного слова влечет изменение связей между другими словами предложения

(1) *Автобус догнал трамвай*

(2) *Он видел их семью своими глазами*

а) Он видел

кого? семью

чьей? их

чем? своими глазами

б) Он видел

кого? их

чем? своими глазами

сколькими? семью

3) *Простой солдат вызвал суматоху.*

а) вызвал

кто? солдат

какой? простой

что? суматоху

б) вызвал

что? простой

кого? солдат

что? суматоху

Омонимия адреса - омонимия формы

Омонимия адреса: Альтернативные связи по разному определяют хозяина для данного слова

Black power struggle

*Fred saw the plane **flying** over Zurich*

*Fred saw the mountains **flying** over Zurich*

*Я **опять** хочу [поехать] в Париж.*

Омонимия формы: Альтернатива состоит в разном определении вида связи для данной пары «слуга – хозяин»

Перевод Тютчева.

Платформа, нагруженная рабочими.

Борец с пережитками.

Выступление адвоката Иванова

адвокат [чей?] – Иванова (управление)

адвокат [имеет фамилию?] – Иванов (согласование)

Благодаря этому инженеру удалось избежать брака.

Казню и милую.

(Примеры Л.Н. Иорданской)

□ Омонимия разных видов связи:

■ Омонимия сочинительных связей:

Вошли два человека в шляпах и пальто.

Вошли два человека в шляпах и мальчик.

■ Омонимия анафорических связей:

*Девочка уронила карандаш на пол и сломала *его*.*

*Взрослые огорчены тем, что дети читают не то, что **им**
хочется.*

...Магомеда Халилова...

*Танки генерала Шаманова расстреляли семерых членов *его*
семьи, а еще соседку и *ее* двоих детей.*

■ **Омонимия межсегментных связей:**

Кто-то застрелил служанку актрисы, которая стояла на балконе.

Необходим контроль за крупными расходами граждан, которые толкают сегодня вверх стоимость жилья.

■ **Более сложный пример (3 варианта сочинения):**

Он постоянно видел отца, красящего забор соседа, старый дом и сарай.

1) отец – сосед – дом – сарай;

2) отец – дом – сарай;

3) забор – дом – сарай.

■ **Омонимия семантической интерпретации синтаксической связи:**

Таблица стандартных размеров:

- 1) 'Таблица имеет (характеристика) стандартный размер'
- 2) 'Таблица содержит сведения о стандартных размерах'

книга сестры:

- 1) 'книга, принадлежит сестре'
- 2) 'книга написана сестрой'

4. Формализация синтаксической структуры предложения

4.1. Синтаксическая структура как математический объект

Математические объекты:

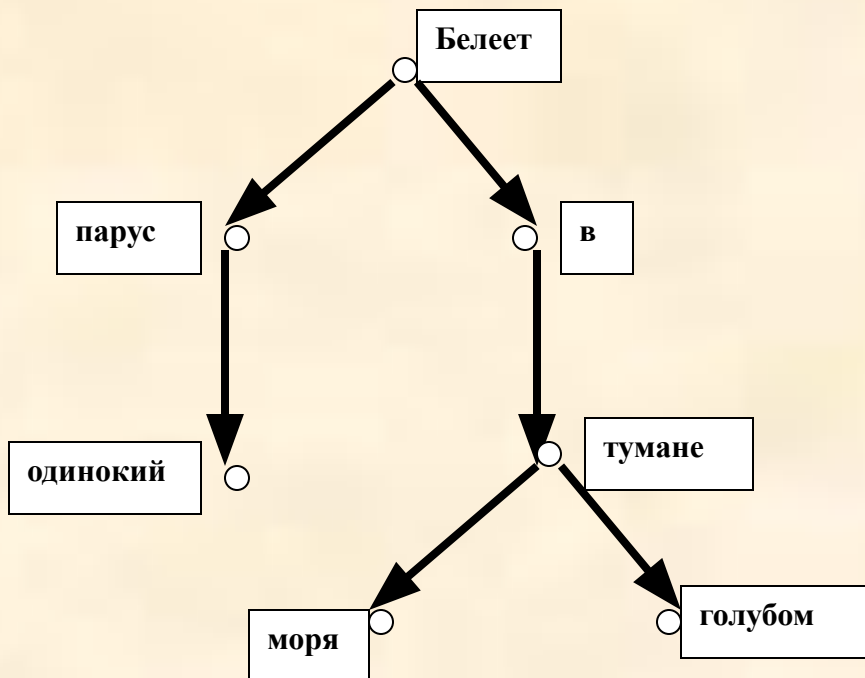
Граф – множество элементов, между некоторыми из которых установлены связи.

Ориентированный граф – если связи имеют направление.

Ориентированное дерево – если связи образуют иерархию подчинения.

Синтаксический граф =
Дерево синтаксических зависимостей.

Дерево синтаксических зависимостей:



Математические понятия:

Множество: $\{ A, B, C, D, E, F, \dots \}$

Упорядоченная пара элементов (стрелка, соединяющая пару элементов): (A, B)

Ориентированный граф – универсальный способ представления:

$\{ (A, B), (C, D), (E, F), \dots \}$

Представление дерева синтаксических зависимостей как множества упорядоченных пар:

$\{ (белеет, парус), (парус, одинокий), (белеет, в), (в, тумане), (тумане, моря), (тумане, голубом) \}$

Другие способы представления дерева синтаксических зависимостей.

а) Сокращенная скобочная запись:

После каждого главного слова перечисляются в скобках все непосредственно ему подчиненные.

Белеет (парус, в)

Белеет (парус (одинокий), в (тумане))

Белеет (парус (одинокий), в (тумане (моря, голубом)))

б) Матрица инцидентности (матрица связей):

		Хозяин						
		1	2	3	4	5	6	7
Слуга	1	0	0	0	0	0	0	0
	2	1	0	0	0	0	0	0
	3	0	1	0	0	0	0	0
	4	1	0	0	0	0	0	0
	5	0	0	0	1	0	0	0
	6	0	0	0	0	1	0	0
	7	0	0	0	0	1	0	0
			0	0	0	0	1	0

4.2. Представление синтаксических связей в виде таблиц. Дополнительные замечания.

1) Сводный список синтаксических помет (для русского языка):

- **С** - согласование
- **У** - управление
- **УС** - управление с согласованием
- **П** - примыкание
- **Соч** - сочинение
- **МССоч** – межсегментное сочинение
- **МС_П** – межсегментное подчинение
- **Анаф** - анафорическая связь

2) Локальная омонимия и анафора.

Слово	N	Хоз	Связь	Содерж
...				
Ленский	Z
...				
<i>Он</i>	1	Z	Анаф	-
Он	1	5	УС	кто?
Из	2	6	У	откуда?
Германии	3	2	У	-из?
<i>туманной</i>	4	3	С	какой?
<i>туманной</i>	4	6	С	какой?
привез	5	0	--	--
учености	6	7	У	чего?
плоды	7	5	У	что?

3) Омонимия содержания и глобальная омонимия.

Слово	N	Хоз	Связь	Содерж	ГлбВар
Очень	1	4	П	как?	1
Дали	2	4	УС	кто?	1
Не	3	4	П	-?	1
любил	4	0	--	--	1
Пикассо	5	4	У	Кого?	1
Очень	1	4	П	Как?	2
Дали	2	4	УС	Кого?	2
не	3	4	П	-?	2
любил	4	0	--	--	2
Пикассо	5	4	У	кто?	2

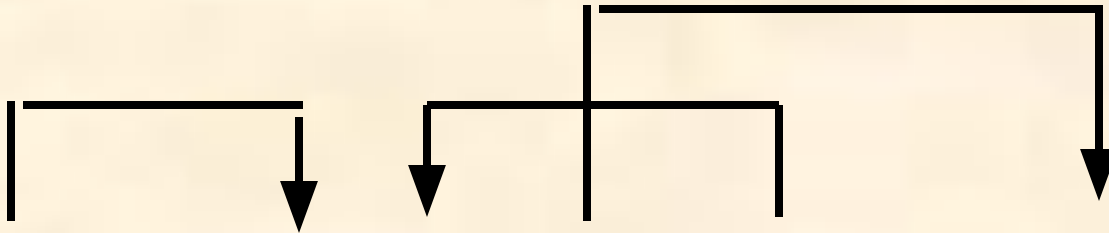
3) Представление сочинительных связей.

Слово	N	Хоз	Связь	Содерж	ГлбВар
В	1	8	П	когда?	1
те	2	3	УС	какие?	1
годы	3	1	П	-в?	1
дальние	4	3	С	какие?	1
глухие	5	4	Соч	--	1
в	6	8	У	в чем?	1
сердцах	7	6	У	-в?	1
царили	8	0	-	--	1
сон	9	8	УС	что?	1
мгла	10	9	Соч	--	1

4.2. Условие проективности для синтаксических структур

- а) Стрелки связей в синтаксическом графе не должны пересекаться.
- б) Дополнительное условие: стрелки не должны накрывать главное слово предложения.

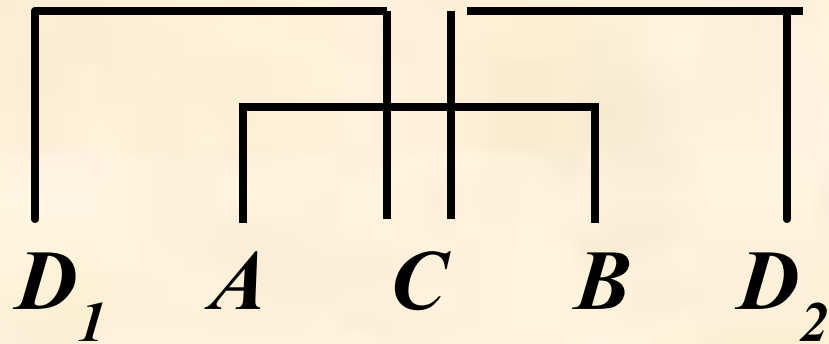
Пример непроективного предложения:



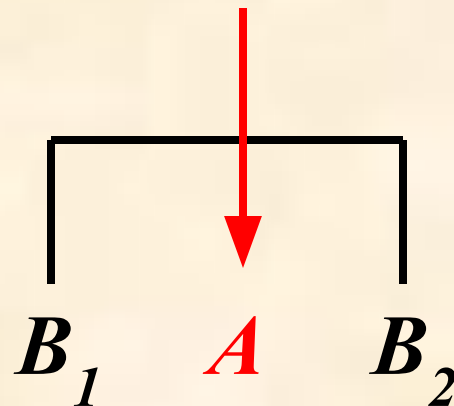
Он из Германии туманной привез учености плоды.

Общая схема:

а) Основное условие:

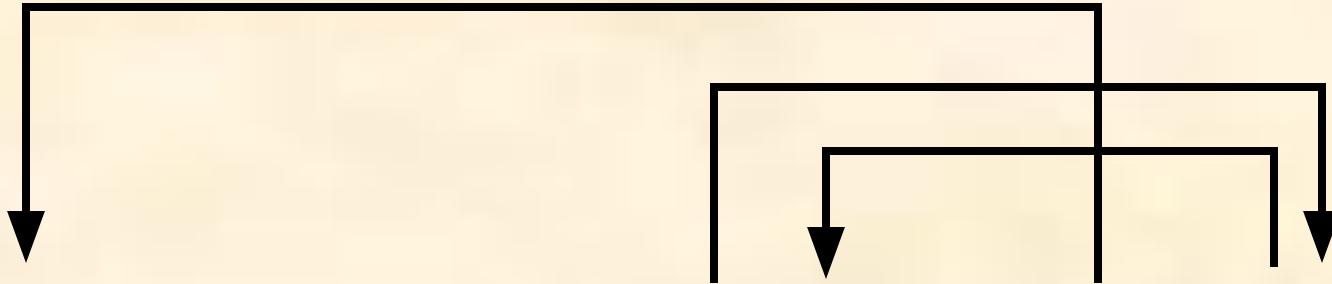


б) Дополнительное условие:



Примеры непроективных конструкций:

(1) Я памятник себе **воздвиг** нерукотворный.

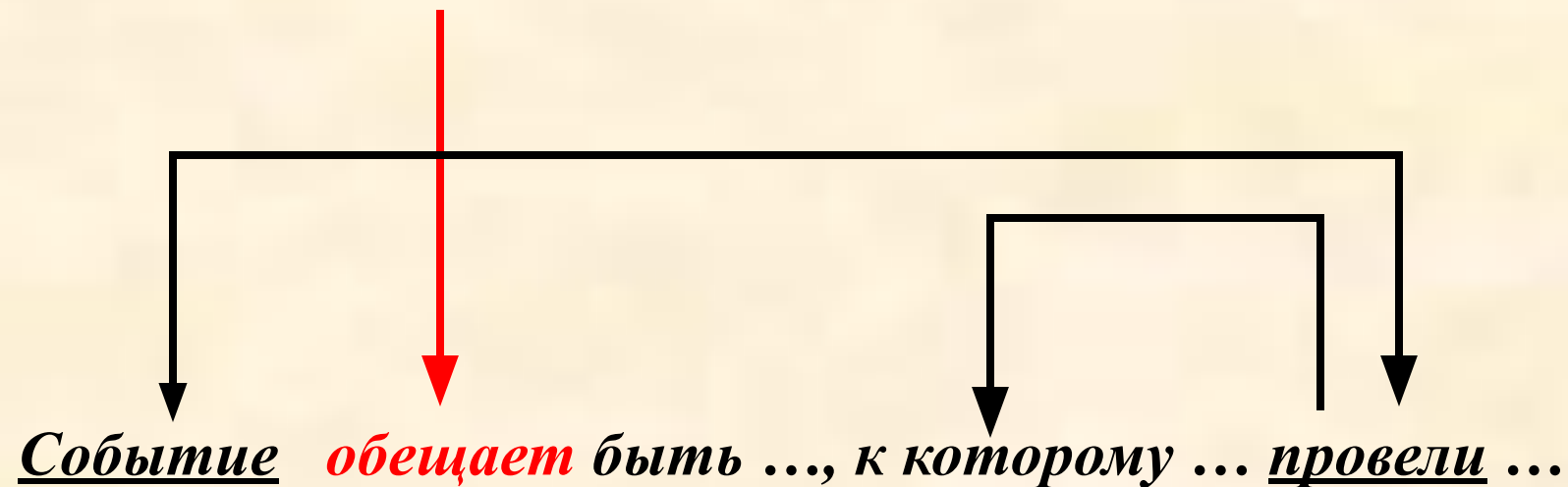


2) И перья страуса склоненные в моем **качаются** мозгу.

3) И каждый вечер в час назначенный
(Иль это только снится мне?)
Девичий стан, шелками схваченный,
В туманном **движется** окне.

4) **Непроективность межсегментных связей:**

(4a) Событие *обещает* быть впечатляющим,
к которому городские власти провели большую предварительную
подготовку.



(4б) *Некоторые из оперетт были специально поставлены в театре в расчете на Зою Виноградову,
в труппе которого она пела с 18 лет.*
[Газета «Телевидение. Радио»]

(4в) *Меня ж никто вредить не может.*

Я злобу твердостью сотру.

Врагов моих **червь** кости **сложет**,

А я ниит – я не умру.

[Державин]

NB: Глобальная омонимия: "червь кости" ~ "гиена подозренья"

[Построить таблицы связей!]