

# Общение с компьютером на естественном языке. Моделирование понимания ЕЯ.

Основы теории интеллектуального управления  
Лекция 4.2

# «Понимающий» компьютер должен осуществлять обработку всех уровней языка

- Фонетика (при голосовом общении)
- Морфология
- Синтаксис
- Семантика
- Прагматика
- Дискурс

Решается только одна проблема:  
**НЕОДНОЗНАЧНОСТЬ**

# Два подхода к моделированию понимания смысла ЕЯ

- *Синтаксически-ориентированный подход* основан на детальном синтаксическом разборе предложения. Средствами синтаксического анализа вычленяются связанные понятия, которые объединяются в так называемые атомы смысла (АС). Создание АС идет только на основе данного предложения, определение связанных понятий идет только на основе синтаксических правил.
- *Семантически-ориентированный подход* на основе распознавания семантики. Разбор предложения идет путем вычленения связанных понятий с помощью базы знаний. База знаний хранит АС и определенным образом представленные связи между этими АС. На втором плане стоит синтаксический анализ с помощью которого вычленяются дополнительные АС, те которые не были сгенерированы с помощью БЗ.

# Технологии анализа ЕЯ. Синтаксический анализ.

- **Парсинг** – процесс структурирования линейной репрезентации в соответствии с заданной грамматикой
- **Линейной репрезентацией** предложения естественного языка называется цепочка элементов, где каждый элемент является минимальной синтаксической единицей

# 5 языковых средств синтаксического анализа

- Словоизменительные морфологические средства
- *w1 зависит от w2 по C, если граммема g категории C, характеризующая w1 выбирается в зависимости от слова w2*
- *В русском языке к словоизменительным категориям относятся категории падежа и числа существительного; категория падежа числительного; категории рода, числа, падежа и степени сравнения прилагательного; категории лица, числа, времени, наклонения и рода (в прош. вр. и сослагат. накл.) глагола; категория степени сравнения наречия*
- *пример: **новые** книги*

# 5 языковых средств синтаксического анализа (2)

- Селективные признаки
  - Частеречные признаки (например, *er* в конце слова в англ. языке)
  - Одушевленность
- Служебные слова (*в, к, при, и, а, или, бы, же, уж, в течение; несмотря на то что; пускай, давай*)
- Знаки препинания
- Порядок слов

# Формализмы

- Контекстно-свободные грамматики (грамматика составляющих)
- Head-driven phrase structure grammar (HPSG)
- Грамматика зависимостей
- Link Grammar
- LR-грамматтики

## Грамматический разбор

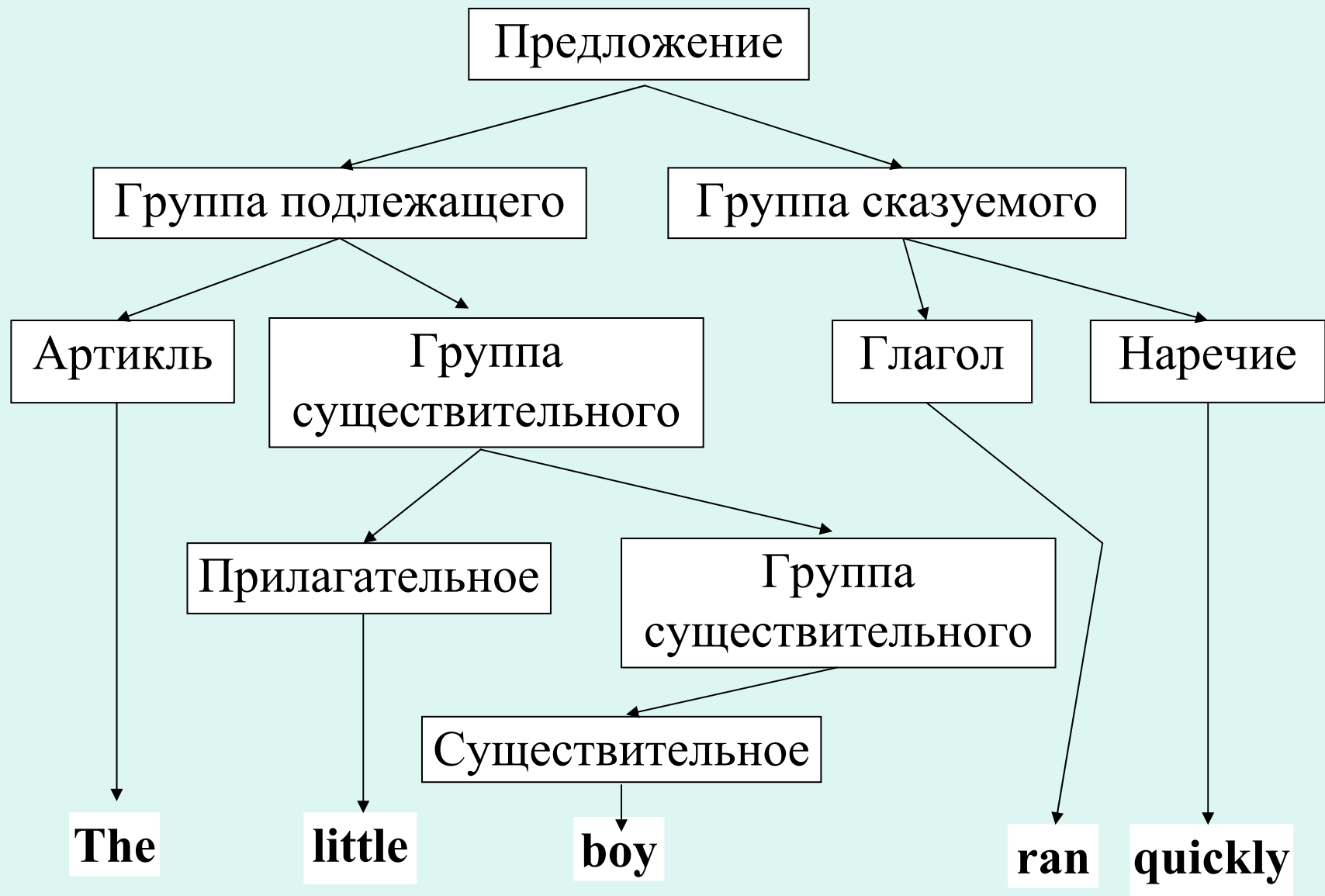
---

При разборе мы имеем дело с *грамматическими категориями*:

‘предложение’, ‘группа существительного’, ‘группа сказуемого’, ‘существительное’, ‘глагол’, ‘наречие’ и т. д. и пользуемся собственно *словами*, составляющими разбираемое предложение.

Например, структуру английского предложения: “The little boy ran quickly” можно изобразить в виде диаграммы.





Синтаксическая структура предложения



# Механизм порождения

Здесь стрелочка  $\rightarrow$  отделяет *левую часть правила* от *правой*, а грамматические термины заключены в *металингвистические скобки*  $<$  и  $>$  для того, чтобы отличать их от слов, составляющих разбираемое предложение.

По этим правилам можно не только *проверять грамматическую правильность предложений*, но также *порождать грамматически правильные предложения*.

## Механизм порождения

### Механизм порождения

Начиная с цепочки, включающей только грамматический термин, являющимся *главным* (< предложение >), каждый грамматический термин, входящий в текущую цепочку, замещается правой частью того правила, которое содержит его в левой части. Когда в результате таких замен в текущей цепочке не останется ни одного термина грамматики, а только слова языка, мы получаем *грамматически правильное предложение* языка.

# Грамматика.

## Язык, порождаемый грамматикой

Ранее речь шла о конкретной грамматике. В ней имеются два словаря:

1) *нетерминалы* — грамматические термины

<предложение>, <группа подлежащего >, ...;

2) *терминалы* — слова, составляющие предложения языка

The, little, boy, ran, quickly;

## Грамматики

3) *правила, левые и правые* части которых состоят из нетерминалов и терминалов;

<предложение> → < группа подлежащего >

< группа сказуемого >

< артикль > → The, ...

4) *начальный нетерминал* — главный грамматический термин; из него выводятся те цепочки терминалов, которые считаются предложениями языка

<предложение>

# Основные виды грамматик

- Контекстно-свободная грамматика – у которой в левой части правил содержится только один нетерминал  
 $A \rightarrow a, b, c.$
- Контекстно-зависимая грамматика – у которой в левой части правил может содержаться помимо нетерминала и терминалы  
 $Ad \rightarrow a, b, c$
- Регулярная грамматика – у которой правая часть каждого правила начинается с терминала  
 $A \rightarrow aB$

# Виды синтаксического анализа (грамматического разбора)

- **Сверху вниз**
  - Программа пытается породить, начиная с главного правила (описывающего структуру предложения) разбираемое предложение (последовательность терминалов)
- **Снизу вверх**
  - Программа пытается на основе текущего слова в предложении (и может быть следующих слов) распознать нетерминалы и в конце концов предложение в целом



# Технологии анализа ЕЯ. Синтаксический анализ. Свободно-контекстные грамматики.

## Недостатки:

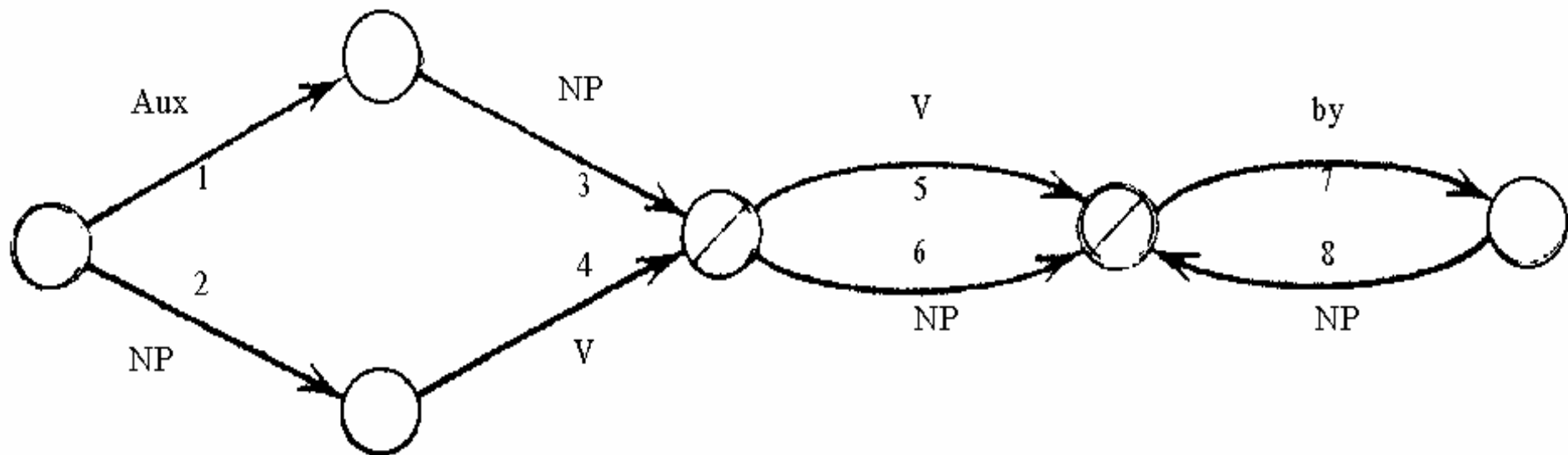
- отсутствие запрета на грамматически неправильные фразы, где, например, подлежащее не согласовано со сказуемым в числе,
- Разрастание грамматики для всех вариантов разбора, в том числе, грамматически неправильных фраз,
- Следствие – непригодны для анализа ЕЯ.

# Технологии анализа ЕЯ. Синтаксический анализ. Трансформационные (генеративные) грамматики.

- Автор – Хомский,
- Для порождения грамматически правильных предложений,
- Центральная идея трансформационной теории состоит в том, что поверхностные формы любого языка - его предложения - являются результатом взаимодействия между несколькими модульными подсистемами
- Трансформационные правила предназначены для описания систематических отношений в предложении, как то:
  - отличия между активным и пассивным предложением
  - глобальные отношения в предложении (например, связь между what и eat в предложении "What will John eat")
  - неоднозначности, причиной которых является одна и та же форма предложения, выведенная из двух различных базовых предложений (например, в предложении "They are flying planes" flying можно рассматривать и как прилагательное и как основной глагол)
- Непригодны для анализа ЕЯ.

# Технологии анализа ЕЯ. Синтаксический анализ. Расширенная сеть переходов.

- Авторы – Бобров, Фрейзер и Вудс,
- Продолжение свободно-контекстных грамматик.
- *The rabbit nibbles the carrot (Кролик грызет морковь).*



# Технологии анализа ЕЯ. Синтаксический анализ. Расширенная сеть переходов.

Недостатки:

- немодульность;
- сложность при модификации, вызывающая непредвиденные побочные эффекты;
- хрупкость (когда единственная неграмматичность в предложении делает невозможным дальнейший правильный анализ);
- неэффективность при переборе с возвратами, т.к. ошибки на промежуточных стадиях анализа не сохраняются;
- неэффективность с точки зрения смысла, когда с помощью полученного синтаксического представления оказывается невозможным создать правильное семантическое представление.

# Технологии анализа ЕЯ. Семантические грамматики.

- Вместо синтаксических категорий в грамматике используются семантические,
- Используются в узких предметных областях,
- Недостатки:
  - для каждой предметной области нужна своя грамматика,
  - Имеют тенденцию быстро увеличиваться в размерах.

S → <present> the <attribute> of <ship>

<present> → what is | [can you] tell me

<ship> → the <shipname> | <classname> class ship

*Can you tell me the class of the Enterprise? (Enterprise*  
- название корабля).

НГТУ, кафедра АППМ  
А.В.Гаврилов

# Технологии анализа ЕЯ. Подбор шаблона.

Диалог программы Eliza, имитирующей психиатра:

**X всегда Y:** Вы можете привести какой-нибудь пример?  
Когда?  
Действительно, всегда?

**X Вы Y меня:** Почему Вы думаете, что я Y Вас?  
Вам нравится думать, что я Y Вас, не правда ли?  
Представьте, что я Y Вас. Что Вы об этом думаете?

В результате *Элиза* способна вести такой диалог:

<b>Пац.:</b> Люди <b>всегда</b> пристально смотрят на меня.	подбор по 1-му шаблону с учетом <i>всегда</i>
<b>Эл.:</b> Вы можете привести какой-нибудь пример?	первый ответ по 1-му шаблону
<b>Пац.:</b> Например, сейчас ты <b>смотришь</b> на меня.	подбор по 2-му шаблону, Y= <i>смотреть</i>
<b>Эл.:</b> Почему Вы думаете, что я <u>смотрю</u> на Вас?	первый ответ по 2-му шаблону
<b>Пац.:</b> Потому что ты <b>ненавидишь</b> меня.	подбор по 2-му шаблону, Y= <i>ненавидеть</i>
<b>Эл.:</b> Вам нравится думать, что я <u>ненавижу</u> Вас, не правда ли?	второй ответ по 2-му шаблону

# Язык шаблонов AIML – язык для Alice-подобных диалоговых систем

- AIML – язык разметки
- Набор категорий <Category>
- Категория состоит из
  - Паттернов <pattern>
  - Шаблонов <template>
- При анализе предложения ЕЯ интерпретатор AIML ищет подходящий паттерн, который сопоставим с предложением, а в ответ выдает один из соответствующих шаблонов
- Имеется возможность запоминать слова из предложения и использовать их в виде переменных в шаблонах и при проверке условий

# Технологии анализа ЕЯ. Падежные фреймы.

- Падежный фрейм состоит из заголовка и набора ролей (падежей), связанных определенным образом с заголовком

[Заголовочный глагол

[падежный фрейм

*агент*: <активный агент, совершающий действие>

*объект*: <объект, над которым совершается действие>

*инструмент*: <инструмент, используемый при совершении действия>

*реципиент*: <получатель действия - часто косвенное дополнение>

*направление*: <цель (обычно физического) действия>

*место*: <место, где совершается действие>

*бенефициант*: <сущность, в интересах которой совершается действие>

*коагент*: <второй агент, помогающий совершать действие>

]]



# Технологии анализа ЕЯ. Падежные фреймы.

Например,  
для фразы *Иван дал мяч Кате* падежный фрейм  
выглядит так:

[Давать

[падежный фрейм

*агент:* Иван

*объект:* мяч

*реципиент:* Катя]

[грам

*время:* прош

*залог:* акт]

]

# Технологии анализа ЕЯ. Падежные фреймы.

Анализ текста с помощью падежных фреймов состоит из следующих шагов:

- Используя существующие фреймы, подобрать подходящий для заголовка. Если такого нет, текст не может быть проанализирован.
- Вернуть в систему подходящий фрейм с соответствующим заголовком-глаголом.
- Попытаться провести анализ по всем обязательным падежам. Если один или более обязательных заполнителей падежей не найдены, вернуть в систему код ошибки. Такой случай может означать наличие эллипсиса, неверный выбор фрейма, неверно введенный текст или недостаток грамматики. Следующие шаги используются уже для анализа и исправления таких ситуаций.
- Провести анализ по всем необязательным падежам.
- Если после этого во введенном тексте остались непроанализированные элементы, выдать сообщение об ошибке, связанной с неправильным вводом, недостаточностью данного анализа или необходимостью провести другой, более гибкий анализ.

# Технологии анализа ЕЯ. Падежные фреймы.

Преимущества использования падежных фреймов таковы:

- совмещение двух стратегий анализа (сверху вниз и снизу вверх);
- комбинирование синтаксиса и семантики;
- легкая встраиваемость в интеллектуальные системы на основе фреймов;
- удобство при использовании модульных программ.