Введение в Теоретическую Лингвистику

Филипп Дудчук, Саша Подобряев

E-mail: seminars@linguistics.msx.ru
http://seminars.narod.ru/spring2005

Лекция 1
18.05.2005

1 *Исходные знания о естественном языке*

Языки

Возможные средства передачи информации: язык дорожных знаков, бенгальский язык, русский язык, азбука Морзе, язык марсиан, язык математической логики, язык пчел, …

Все это языки. Как в общем случае устроен язык?
У всякого языка имеется алфавит, синтаксис и семантика.

- **Алфавит** — список (= множество) символов, из которых составляются слова, словосочетания, предложения. Пока для простоты будем называть слова, словосочетания и предложения одним словом — **формула**.

- **Синтаксис** — список (= множество) таких правил, по которым можно составить любую правильную формулу языка и нельзя составить никакую неправильную формулу. Нестрого говоря, синтаксис — сведения о том, как построить всякую правильную формулу на данном языке.

Для того, чтобы понять, что такое семантика, сперва отбросим часть формул, которые были построены при помощи синтаксических правил, и будем рассматривать только такие формулы, про которые можно сказать, истинны они или ложны. Такие формулы будем называть **высказываниями**.

(1) Сегодня идет дождь.
(2) Если Петя любит Машу, то Маша любит Петю.
(3) Если Петя дружит с Машей, то Маша дружит с Петей.

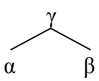
(4) Пусть пойдет дождь!
(5) Петя и Маша пойдут в кино?
(6) Если Петю выгнать из школы.

(1–3) являются правильными формулами и являются высказываниями. (4–5) тоже являются правильными формулами, но не являются высказываниями. (6) не является правильной формулой и, следовательно, не является высказыванием. Теперь можно дать определение семантике.

- **Семантика** — список (= множество) условий для всякого высказывания, при которых это высказывание истинно. Нестрого говоря, семантика — сведения о том, что значат высказывания на данном языке.

2 *Синтаксис естественного языка*

2.1 Грамматика непосредственных составляющих



"α — сестра β"
"β — сестра α"

"γ непосредственно доминирует над α и β", или "γ мать α и β"

VP	глагольная группа:	*любит свою младшую сестру*
NP	именная группа:	*свою младшую сестру*, *пятилетний Петя*
PP	предложная группа:	*с другом*
AP	группа прилагательного:	*гордый за своих детей*
AdvP	группа наречия:	*очень быстро*

Синтаксические критерии выделения составляющих (Constituency Tests)

→ топикализация (вынесение отрезка предложения в его начало):

(7) Так быстро он не справится.
(8) А для Коли почему не взяли?

→ фрагментирование (употребление отрезка предложения в качестве отдельного высказывания):

(9) — Куда он отправился? — В горы.

→ способность замещаться проформами (местоимениями):

(10) — Вы [$_{PP}$ бывали в Петербурге]? — Да, я там бывал;
(11) Разумеется [$_{IP}$ я не Солженицын]. Разве это лишает меня права на существование?
(12) — Что вы скажете об [$_{NP}$ этом таинственном молодом человеке]? — Скажу, что он жулик.

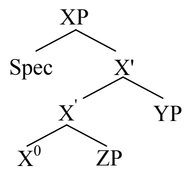
2.2 Порождающая грамматика Хомского

2.2.1 X'-теория

Правила фразовой структуры (Phrase Structure Rules):

A. XP → SpecXP X'
B. X' → X' YP
C. X' → X^0 ZP

(13)



2.2.1 Аргументы и адъюнкты

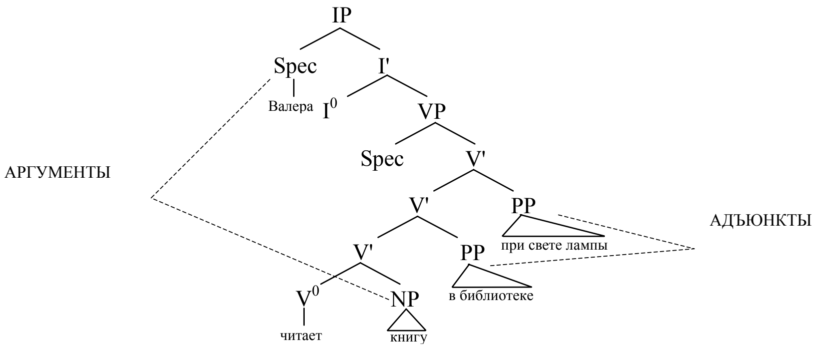
Аргумент — обязательный член (обязательная составляющая) предложения, при его отсутствии возникает «неполнота». При конкретном глаголе число аргументов ограничено.

Аргументы или являются спецификаторами VP, или являются сестрами V^0.

Адъюнкт — необязательный член (необязательная составляющая) предложения, при его отсутствии неполноты не возникает. Адъюнктов в принципе может быть сколько угодно:

(14) Видел своего друга в Александровском саду осенью при большом скоплении народа.

(15)



Валера читает книгу в библиотеке при свете лампы.

3 *Семантика высказывания*

- Как уже указывалось, семантика естественного языка — множество множеств условий истинности для каждого высказывания естественного языка.

- В прошлом семестре мы построили формализм для записи предикаций (= предикатно-аргументных структур) — той части смысла предложений, который указывает на то, о каких предметах мира (индивидах) идет речь, и о том отношении, в котором они находятся по отношению друг к другу (о предикатах). Расширим понятие предиката. Теперь будем говорить так:
  - запись P(x), где P является одноместным предикатом — это указание на то, что x принадлежит множеству P индивидов, которых объединяет некоторое свойство. Например, предикация запись **спать** (*Вася*) означает, что индивид (будем также говорить — **индивидная константа**) *Вася* принадлежит к множеству всех спящих индивидов.
  - запись K(x, y), где K является двухместным предикатом — указание на то, что упорядоченная пара индивидов <x, y> принадлежит к множеству пар индивидов K, для каждой из которых верно, что ее члены находятся в отношении K. Например, запись **дружить** (Маша, Петя) означает, что упорядоченная пара индивидов <Маша, Петя> принадлежит к множеству всех пар, члены которых находятся в отношении дружбы.
  - запись $\Re(x_1, \ldots, x_n)$, где $\Re$ является n-местным предикатом — указание на то, что кортеж $<x_1, \ldots, x_n>$ принадлежит к множеству кортежей индивидов длиной n, для каждого из которого верно, что его члены находятся в отношении $\Re$.

3.1 Формальный язык и истинность высказываний

Будем заниматься переводом предложений естественного языка на специальный формальный язык. Назовем этот язык **языком исчисления предикатов первого порядка**. Условимся: если высказывание ϕ на нашем языке истинно, будем писать (16); если ложно, то (17).

(16) $\llbracket\phi\rrbracket = 1$
(17) $\llbracket\phi\rrbracket = 0$

Временно условимся, что $\llbracket\ \rrbracket$ — функция, область определения которой — все высказывания языка, а область значений — множество, которое состоит из двух элементов: {1,0}. Как уже стало понятно, функция возвращает 1 для истинных высказываний и 0 — для ложных. Позже мы увидим, что области определения и значений другие.

Язык, которым мы будем пользоваться, как и любой язык, имеет алфавит, синтаксис и семантику.

3.2 Алфавит языка

(18) $U = \langle =, \neg, \leftrightarrow, \rightarrow, \wedge, \vee, (\,,\,), \forall, \exists \rangle$

Кроме того, нам понадобится множество M — множество имен всех мыслимых предметов, или индивидов; это множество будем называть множеством **индивидных констант**. Понадобится также множество переменных, область значений каждой из которых — множество M. Эти переменные будем называть **индивидными переменными**. Введем также множество предикатов, или **предикатных констант**. — это упорядоченные двойки, тройки и т. д. Элементов из множества M. Содержательно это — все мыслимые предикаты естественного языка.

- Рассмотрим, как «употребляются» в языке первого порядка символы его алфавита.

(19) = **равенство**
0 = 0, 1 = 1;
если a = b, где a и b — индивидные константы, то a и b — имена одного и того же индивида;
если $x = y$, где x и y — индивидные переменные, то какое бы значение не приняла переменная x, это же значение принимает переменная y и какое бы значение ни приняла переменная y, это же значение принимает x.

(20) $\neg$ **отрицание**
если $\llbracket\phi\rrbracket = 1$, то $\llbracket\neg\phi\rrbracket = 0$
если $\llbracket\phi\rrbracket = 0$, то $\llbracket\neg\phi\rrbracket = 1$
Эту информацию можно записать в виде таблицы. Будем называть такие таблицы **таблицами истинности**.

$\llbracket\phi\rrbracket$	$\llbracket\neg\phi\rrbracket$
1	0
0	1

(21) $\leftrightarrow$ **эквивалентность**
Две формулы эквивалентны, когда они при одинаковых условиях истинны и при одинаковых условиях ложны — см. текст раздела 3 в Курсе лекций.

$\llbracket\phi\rrbracket$	$\llbracket\psi\rrbracket$	$\llbracket\phi \leftrightarrow \psi\rrbracket$
1	1	1
1	0	0
0	1	0
0	0	1

(22) $\rightarrow$ **импликация**
Импликация переводится на русский язык предложениями вида «если …, то…»; импликаци ложна, титтк посылка импликации (**антецедент**) является истинной, а следствие импликации (**консеквент**) является ложным; в остальных случаях импликация истинна.

$\llbracket\phi\rrbracket$	$\llbracket\psi\rrbracket$	$\llbracket\phi \rightarrow \psi\rrbracket$
1	1	1
1	0	0
0	1	1
0	0	1

(23) $\wedge$ **конъюнкция**
Конъюнкций двух формул переводится на русский язык союзом «и». Конъюнкция истинна ровно в одном случае: когда истинны оба ее члена.

$\llbracket\phi\rrbracket$	$\llbracket\psi\rrbracket$	$\llbracket\phi \wedge \psi\rrbracket$
1	1	1
1	0	0
0	1	0
0	0	0

(24) $\vee$ **дизъюнкция**
Дюзъюнкция двух формул переводится на русский язык союзом «или». Для итинности дизъюнкции достаточно истинности одного из ее членов.

$[\![\phi]\!]$	$[\![\psi]\!]$	$[\![\phi \vee \psi]\!]$
1	1	1
1	0	1
0	1	1
0	0	0

На этом временно остановимся. Смысл оставшихся двух символов алфавита поясним чуть позже.

Будем использовать тот способ записи предикации (предикатно-аргументной структуры), который обсуждался на позапрошлой лекции.

3.3 Конкретные примеры

Продемонстрируем некоторое применение описанного языка.

(25) Маша не пришла. $\leftrightarrow$ Неверно то, что Маша пришла.
(25') ⟦Маша не пришла⟧ = ¬ ⟦Маша пришла⟧ = ¬**прийти** (Маша)

(26) Коля всегда плачет тогда, когда мама уходит.
(26') ⟦Коля всегда плачет тогда, когда мама уходит⟧ = **плакать** (Коля)$\leftrightarrow$ **уходить** (мама)

(27) Коля плачет, когда мама уходит.
(27') ⟦Коля плачет, когда мама уходит⟧ = **уходить** (мама) $\rightarrow$ **плакать** (Коля)

(28) Уехали то ли Коля, то ли Маша, то ли оба сразу.
(28') ⟦Уехали то ли Коля, то ли Маша, то ли оба сразу⟧ = **уехать** (Коля) $\vee$ **уехать** (Маша)

(29) Маша и Петя пришли. $\leftrightarrow$ Маша пришла и Петя пришел.
(29') ⟦Маша и Петя пришли⟧ = **прийти** (Маша) $\wedge$ **прийти** (Петя)

То, что мы записывали в примерах со штрихом при номере, будем называть **интерпретацией**.
Пока мы не ввели смысла последних двух символов, на этом языке нельзя записать, например, (15) и (16).

(30) На семинар пришли все слушатели.
(31) На свете бывают розовые слоны.

Смысл двух оставшихся без пояснений символов алфавита, каждый из которых называется **квантором**, приводится в (32) и (33).

(32) квантор всеобщности $\forall$
если формула ϕ содержит индивидную переменную x (для ясности будем писать $\phi(x)$), то
$[\![\forall x\phi(x)]\!] = 1$ титтк для любого значения x $[\![\phi(x)]\!] = 1$
(33) квантор существования $\exists$
если формула ϕ содержит индивидную переменную x (для ясности будем писать $\phi(x)$), то
$[\![\exists x\phi(x)]\!] = 1$ титтк найдется хотя бы одно значение переменной x, что $[\![\phi(x)]\!] = 1$

В (32) и (33) x называется **переменной, связанной квантором**. Формула $\phi(x)$ называется **сферой действия** квантора. Все несвязанные переменные будем называть **свободными**. Главное отличие свободных переменных от связанных в том, что вместо первых можно подставлять индивидные константы, а вместо последних — нельзя. Запишем теперь на нашем языке смысл предложений (30) и (31).

(30') ∀x [**слушатель_семинара** (x) → **прийти_на_семинар** (x)]
(31') ∃x [**слон** (x) ∧ **розовый** (x)]

Обратим внимание на некоторые тонкости обращения с кванторами. В интерпретации (30') требуется именно импликация, а не конъюнкция, как в (31'). Если в (30') поставить конъюнкцию, получится смысл совсем другого предложения, нежели смысл (30). Ср. (34') и (34).

(34') ∀x [**слушатель_семинара** (x) ∧ **прийти_на_семинар** (x)]
(34) Все предметы вокруг нас — это слушатели семинара, которые пришли на семинар; других предметов вокруг нас нет.

Аналогично, если подставить в (31') вместо конъюнкции импликацию получится смысл другого высказывания.

(35') ∃x [**слон** (x) → **розовый** (x)]
(35) Среди слонов найдется хотя бы один розовый.

Для иллюстрации запишем интерпретации других предложений с кванторами.

(36) ⟦Все слоны — розовые⟧ = ∀x [**слон** (x) → **розовый** (x)]
(37) ⟦Всё смертно⟧ = ∀x [**смертный** (x)]
(38) ⟦Не все люди, живущие в России, — русские⟧ = ∃x [**человек** (x) ∧ **жить_в_России** (x) ∧ ¬**русский** (x)]
(39) ⟦Все люди, живущие в России, не русские⟧ = ∀x [**человек** (x) ∧ **жить_в_России** (x) → ¬**русский** (x)]
(40) ⟦Не все люди, живущие в России, — русские, и не все люди, не живущие в России, русские⟧ =
= ∃x [**человек** (x) ∧ **жить_в_России** (x) ∧ ¬**русский** (x)] ∧ ∃x [**человек** (x) ∧ **русский** (x) ¬ **жить_в_России** (x)]
(41) ⟦В России никто не живет⟧ = ∀x [**живое_существо** (x) → ¬**жить_в_России** (x)]

Укажем, что следующие формулы — (42–45) — являются тождественно истинными. Эти формулы называются **законами отрицания квантора**.

(42) $\neg\forall x\ \phi(x) \leftrightarrow \exists x\neg\phi(x)$
(43) $\neg\exists x\ \phi(x) \leftrightarrow \forall x\neg\phi(x)$
(44) $\forall x\ \phi(x) \leftrightarrow \neg\exists x\neg\phi(x)$
(45) $\exists x\ \phi(x) \leftrightarrow \neg\forall x\neg\phi(x)$

Укажем также еще две тождественно истинные формулы; эти формулы называются **законами Де Моргана**.

(46) $\neg(\phi \wedge \psi) \leftrightarrow (\neg\phi \vee \neg\psi)$
(47) $\neg(\phi \vee \psi) \leftrightarrow (\neg\phi \wedge \neg\psi)$

Убедимся в истинности, например, формулы (43).

(48) ⟦Никто не смеялся и не улыбался⟧ = ¬∃x [**смеяться** (x) ∨ **улыбаться** (x)] =
= ∀x ¬[**смеяться** (x) ∨ **улыбаться** (x)]

❶ Почему в (48) записана дизъюнкция? Ведь в предложении русского языка союз *и*!
① Конъюнкция [**смеяться** (x) ∧ **улыбаться** (x)] истинна только тогда, когда истинны все ее члены. Но смеяться и улыбаться одновременно невозможно. Тогда такая конъюнкция не будет истинна никогда — какой квантор перед ней не поставь. Но тогда отрицание этой формулы с любым квантором перед ней будет, наоборот, тождественно истинным высказыванием. Тогда предложение русского языка (48) — тождественно истинное высказывание. Но мы знаем, что ситуация «Никто не смеялся и не улыбался» далеко не всегда имеет место. Мы приходим к противоречию.

По закону Де Моргана (47), интерпретацию (48) можно переписать так:

(49) ⟦Никто не смеялся и не улыбался⟧ = ∀x [¬**смеяться** (x) ∧ ¬**улыбаться** (x)]

В этой формуле как раз и появляется конъюнкция, которая отражает смысл русского союза *и* в (48).
В истинности остальных законов отрицания кванторов убедитесь самостоятельно.