

# Категориальные грамматики

## Введение.

Степан Кузнецов, Мати Пентус, Алексей Сорокин

МГУ им. М. В. Ломоносова, межфакультетский курс,  
весенний семестр 2015–2016 учебного года

## Применение математики в лингвистике

- Строгое описание лингвистических явлений и понятий (фонема – класс эквивалентности по отношению аллофонии).
- Формальное моделирование лингвистических процессов,
- Статистическое моделирование лингвистических явлений.

## Два способа моделирования лингвистических явлений

Два метода применения математики в лингвистике:

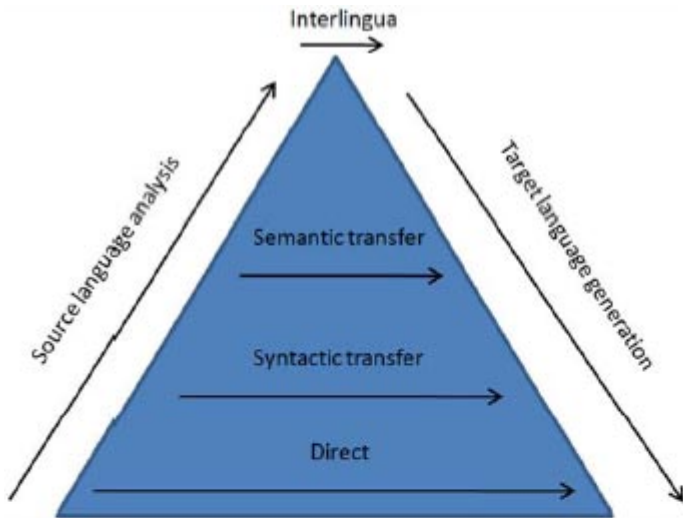
- Строится формальная модель явления, исходя из которой делаются предсказания.
- Само явление не моделируется, моделируется лишь его исход.

Классический пример: правилковый и статистический машинный перевод.

## Общая схема статистического перевода

- $t = \operatorname{argmax}_s P(t|s) = \operatorname{argmax}_s P(s|t)P(t)$ .
- $P(s|t)$  – модель перевода (с целевого языка на исходный), характеризует точность перевода.
- $P(t)$  – языковая модель (для исходного языка), характеризует качество получившегося результата.
- Оценка вероятностей: сбор статистики с больших наборов текстов.

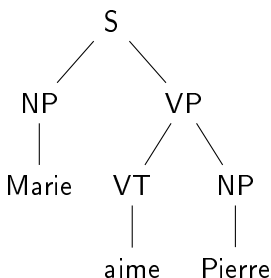
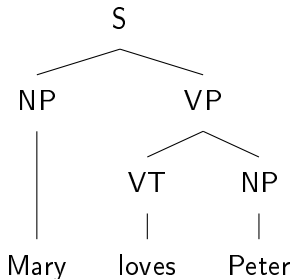
# Общая схема правильного перевода



## Пример перевода

*Mary loves Peter*  $\mapsto$  *Marie aime Pierre*.

Синтаксическое представление:

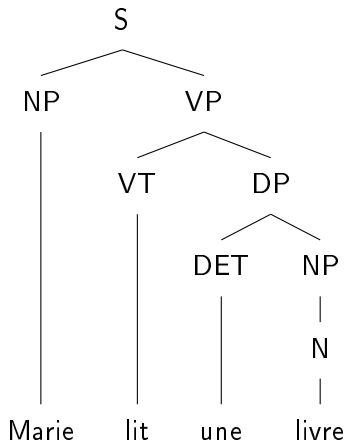
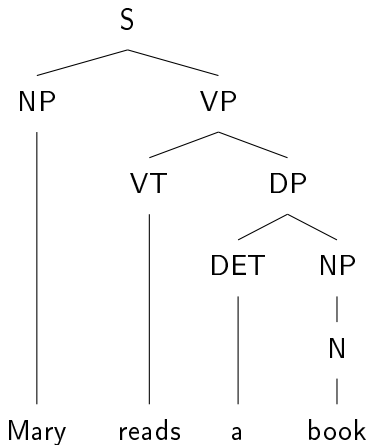


Семантическое представление:  
ЛЮБИТЬ(Мария, Пётр)

## Пример II.

*Mary reads a book*  $\mapsto$  *Marie lit un livre.*

Синтаксическое представление:



## Пример II.

*Mary reads a book*  $\mapsto$  *Marie lit un livre.*

Семантическое представление:

ЧИТАТЬ(Мария, книга) **ПЛОХО!**

Книга — не константа предметного мира.

$\exists x(\text{КНИГА}(x) \wedge \text{ЧИТАТЬ}(\text{Мария}, x))$



## Пример II.

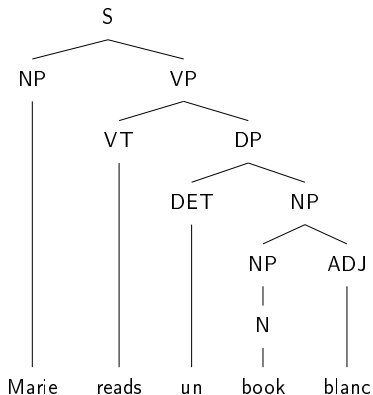
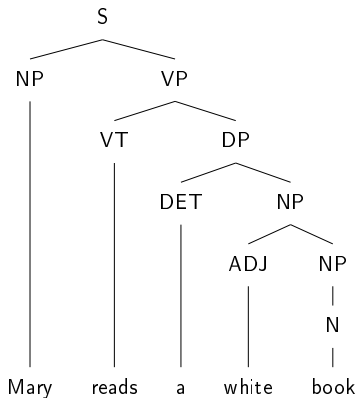
*Mary reads a white book*  $\mapsto$  *Marie lit un livre blanc.*

Семантическое представление:

$\exists x(\text{БЕЛЫЙ}(x) \wedge \text{КНИГА}(x) \wedge \text{ЧИТАТЬ}(\text{Мария}, x))$

## Пример II.

Синтаксическое представление:



## Сложности при правилочном подходе

- При переводе с английского на французский меняется порядок слов в именной группе.
- Способы обработки:
  - Правильная модель: завести правило  $ADJ_e N_e \mapsto N_f ADJ_f$ .
  - Можно не вводить никаких правил, а всё обрабатывать “внутри языка”.
  - Статистическая модель: последовательность *livre blanc* более вероятна, чем *\*blanc livre*.
  - Фразовая статистическая модель: фраза *white book* переводится как *livre blanc*.
  - Можно дополнительно переставить слова на стадии постпроцессинга.

## Семантическое представление: сложности с эллипсисом

*Маша знает, что Петя сдал экзамен.*

$\exists x(\text{ЗНАТЬ}(\text{Маша}, \text{СДАТЬ}(\text{Петя}, x)) \wedge \text{ЭКЗАМЕН}(x))$

*Маша знает, что она сдала экзамен.*

$\exists x(\text{ЗНАТЬ}(\text{Маша}, \text{СДАТЬ}(\text{Маша}, x)) \wedge \text{ЭКЗАМЕН}(x))$

*Маша знает, что сдала экзамен.*

$\exists x(\text{ЗНАТЬ}(\text{Маша}, \text{СДАТЬ}(\text{Маша}, x)) \wedge \text{ЭКЗАМЕН}(x))$