**1. What is NLP?**

NLP = NLU + NLG

NLU: 语⾳音/⽂文本 -> 意思(meaning)

NLG: 意思-> ⽂文本/语⾳音

**2. The Challenge**

Multiple Ways to Express (多种表达⽅方式)

Ambiguity(一词多义)

🡪 how to solve ambiguity:

从单词到语境(context) interest --- a financial interest in IBM

更新认知

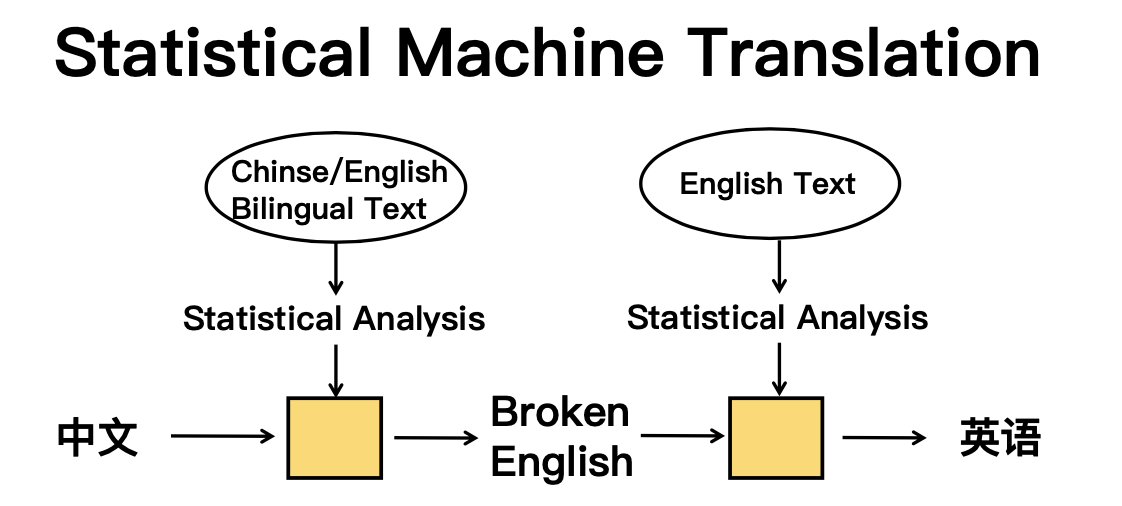
**3. Case Study: Machine Translation**

图片包含 屏幕截图

描述已自动生成

在语料库中进行统计，找单词配对。

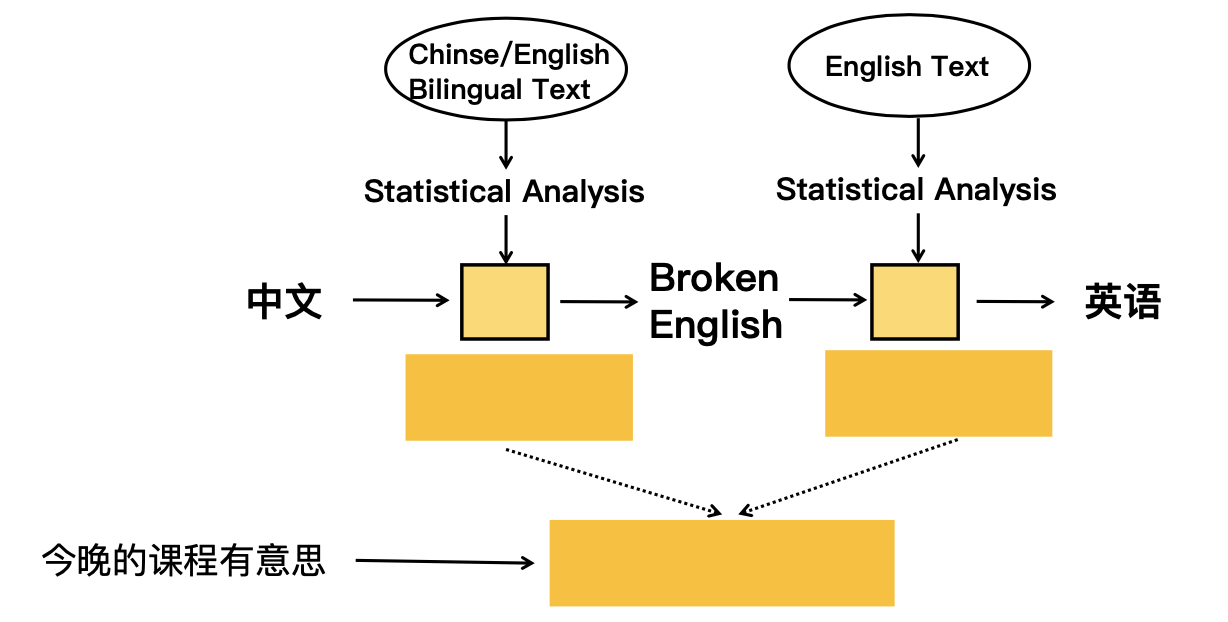
缺点：慢、语义语境、上下文考虑不足、语法错误



今晚的课程有意思---tonight of course interesting ( broken English )

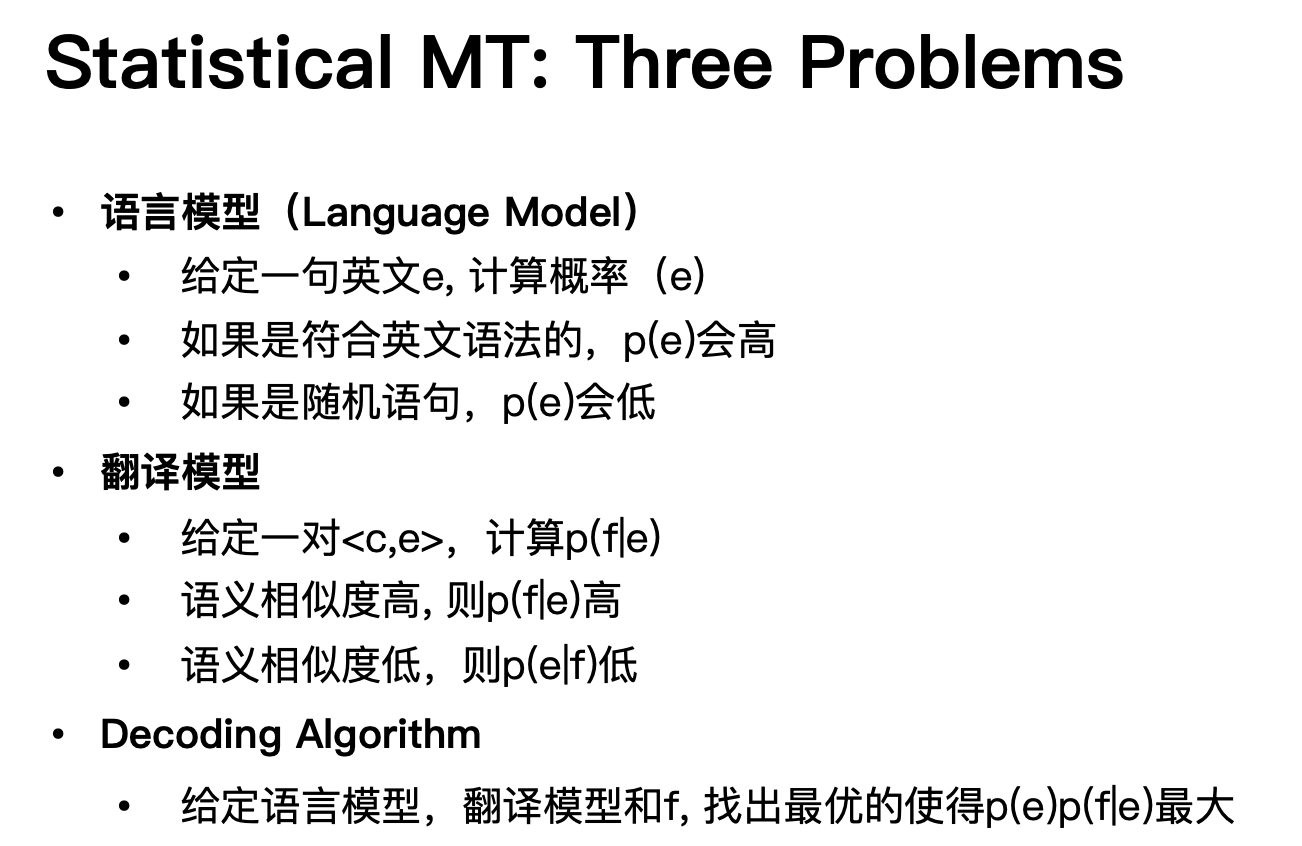
排列组合一共有4！中句子

Language Model: 求解每一种排列组合的概率---最高的来寻找合适的句子



Translation model 🡪 language model🡪decoding algorithm

**4. Statistical MT: Three Problems**

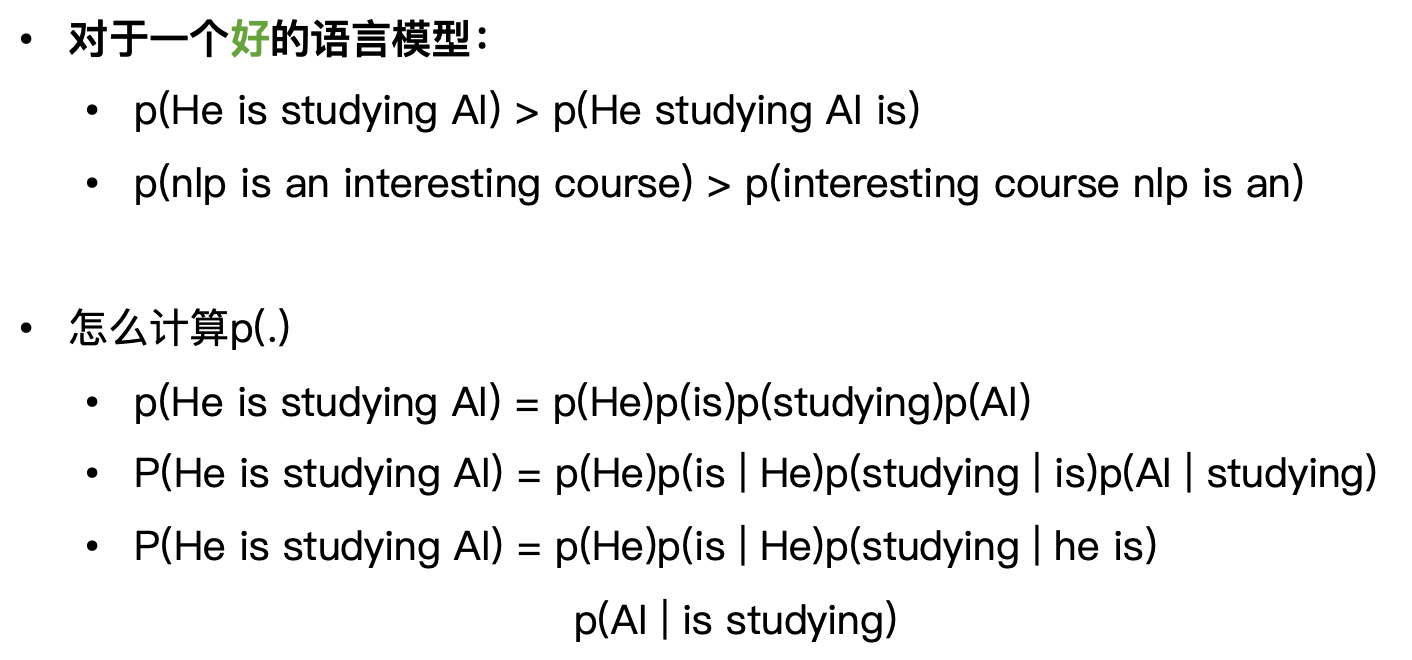


语言模型是需要提前训练出来的。

翻译模型，可以起到一个词典的作用。

Decoding = 语言+翻译

**4.1 Language Model (语言模型)**

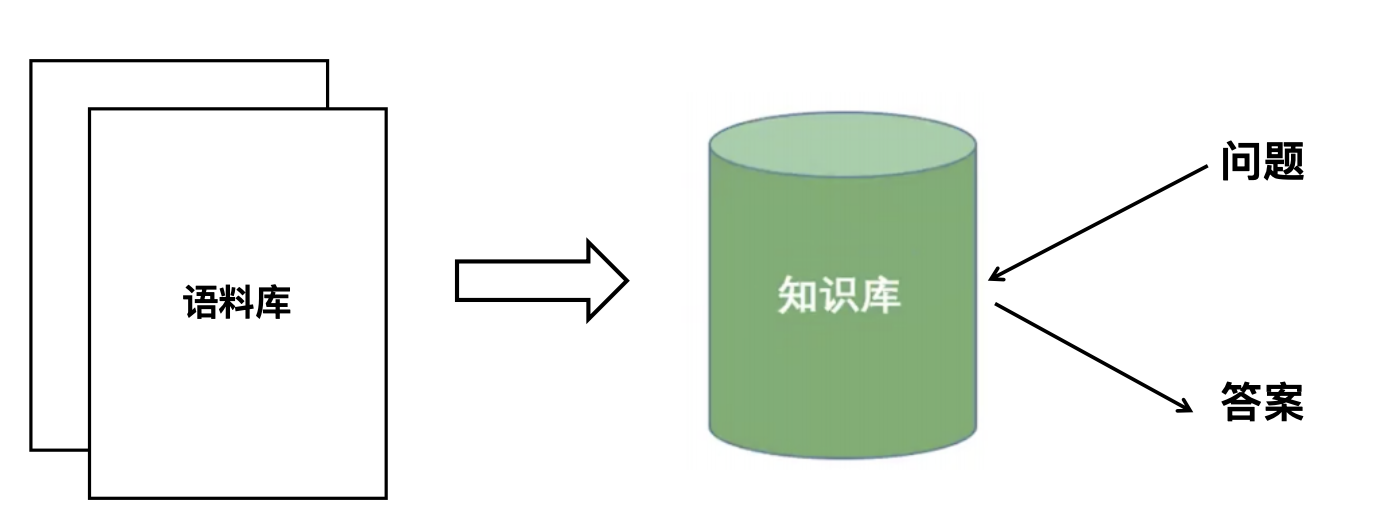


Unigram, bigram, trigram, ---, N-gram 在考虑一个单词的时候，需要参考前面多少个单词

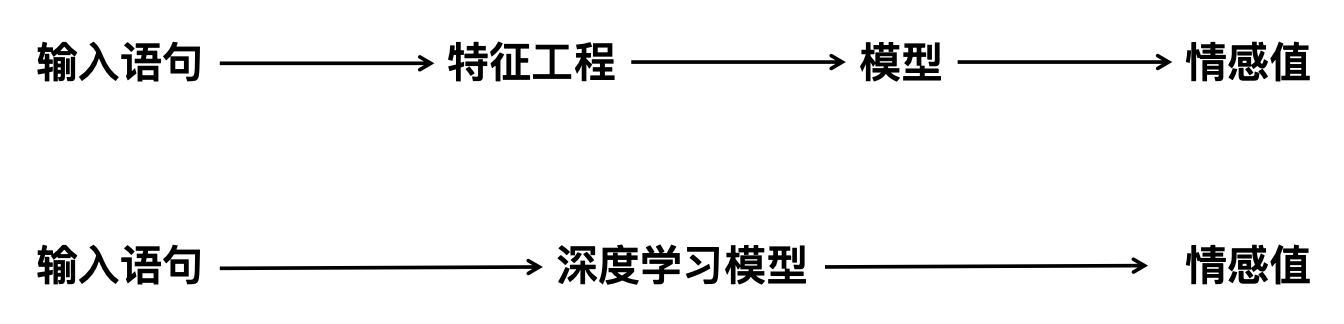
核心是不知道每一个p

**5. NLP的经典应用场景**

Question Answering(问答系统)



Sentiment Analysis(情感分析)



Machine Translation(机器翻译)

Text Summarization(自动摘要)

Chatbot (聊天机器人)

Information Extraction(信息抽取)

**6. NLP关键技术**

四大维度：

Semantic语义（NLU、机器翻译）

syntax句子结构（句法分析、依存分析

morphology单词（分词、pos词性、NER）

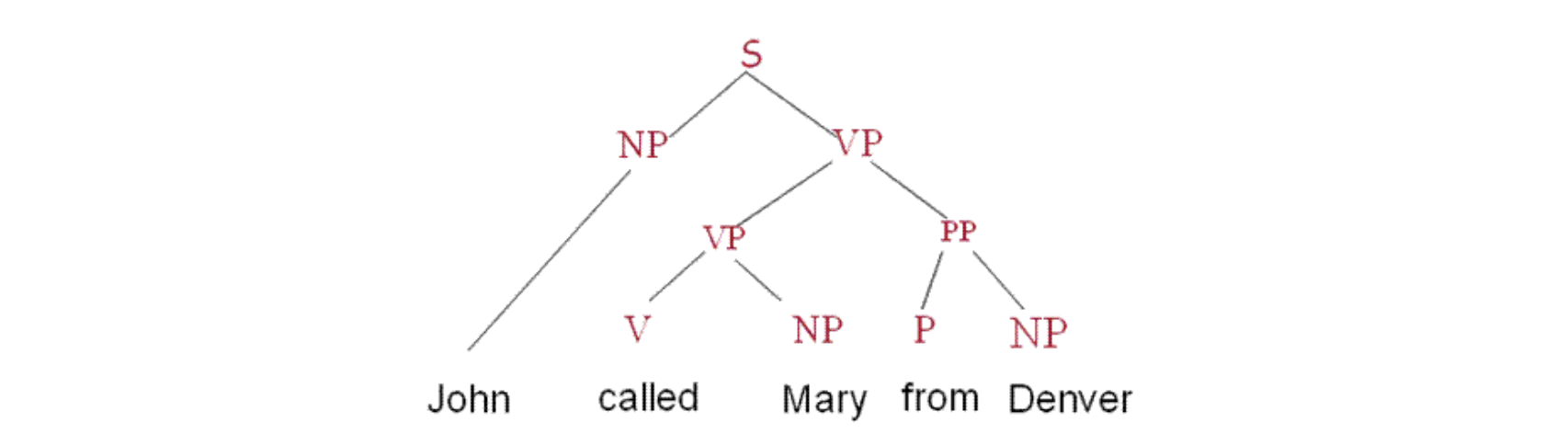
phonetics声音

word segmentation分词

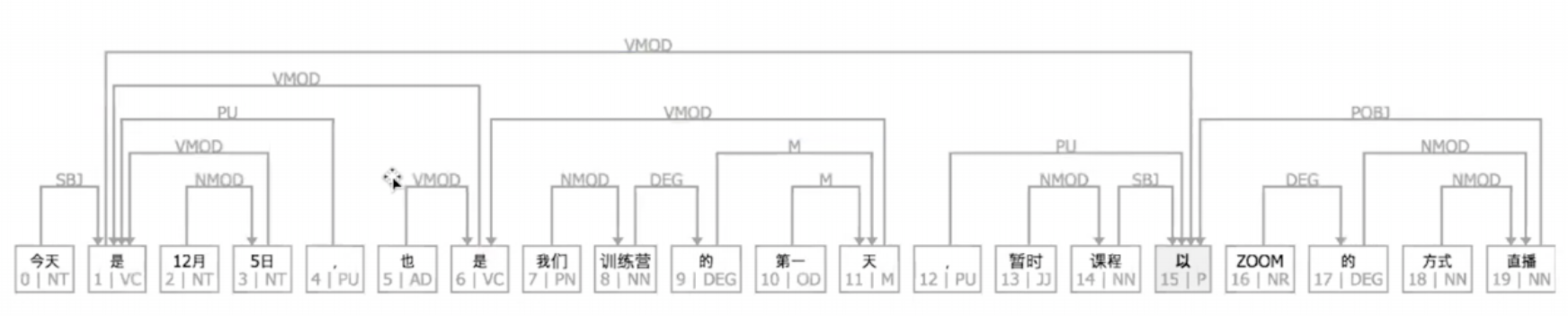
part of speech词性分析，一个单词在不同句子里扮演不同角色

named entity recognition命名实体识别，抽取关键信息---知识图谱 问答系统

parsing句法分析



Dependency parsing依存分析



Relation extraction关系抽取

