**第十二章读书报告**

09118223 吴亦珂

读书进度：第十二章读完一半

1. **问题列表**

（我提出）P3也提到的terminal symbols和non-terminal symbols具体指什么？

讨论结果：non-terminal symbols是指还可以进行进一步展开，进一步进行推导，而terminal symbols是指不能再进行进一步的推导。

（我提出）P6页说的推导感觉没理解，具体是如何由一个句子产生另一个句子的吗？

讨论结果：应该是指按照相应规则从左到右一步一步代入展开。语法树生成是应该是自底向上但规则产生时则是自顶向下。

（别人提出）P8最下面段落的clause的概念具体是指什么？是否有例子？  
讨论结果：应该是指从句，最后几段我的理解是S可以指代一个完整的句子，从某种程度上说，使得S规则和其他规则不同的一个原因是它可以与从句有一定关联，有一定的完整性。所以S就是指大句子中套的小句子。比如I think后面的宾语从句应该可以叫做S。

（别人提出）Treebank都有哪些具体应用？可以用来做什么？  
讨论结果：应该可以用来做句子结构的分析以及语料标注

（别人提出）constituency grammars有什么优势和实际应用？  
讨论结果：能较为简单的推导出一个句子结构。可以用于句子产生以及词性标注。但缺点是推导出的结构可能存在二义性。

1. **读书收获**

**12.1 Constituency**

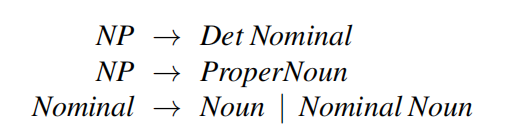
Constituency:基于一组单词可以看成一个整体。

名词词组：可以出现在相似的句法环境之中，例如出现在动词前。、

**12.2 Context-Free Grammars**

Context-free grammer是一系列规则，每一个规则描述了单词是如何聚集的。可以看成产生句子的方法，也可以看成给句子划分结构的方法。

例如：



CFG中的symbol分为两类：terminal symbols和non-terminal symbols。

Parse tree：可以用来表示句子产生的过程。

S：可以被看成start symbol。

动词词组之后跟名词词组的规则：



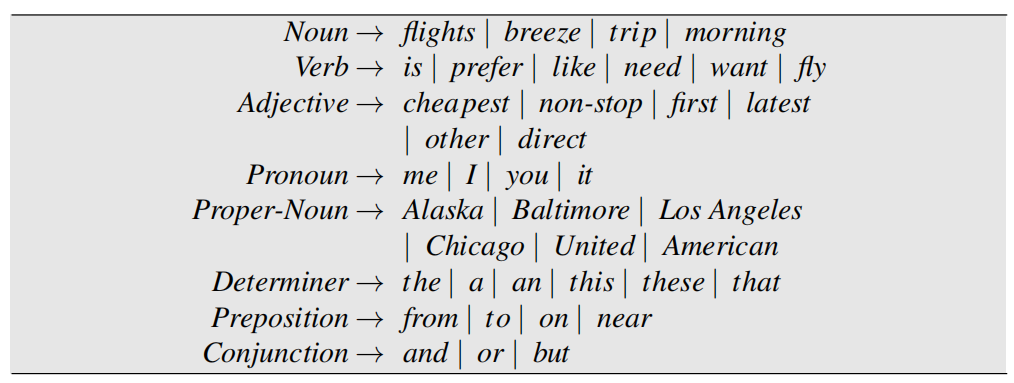


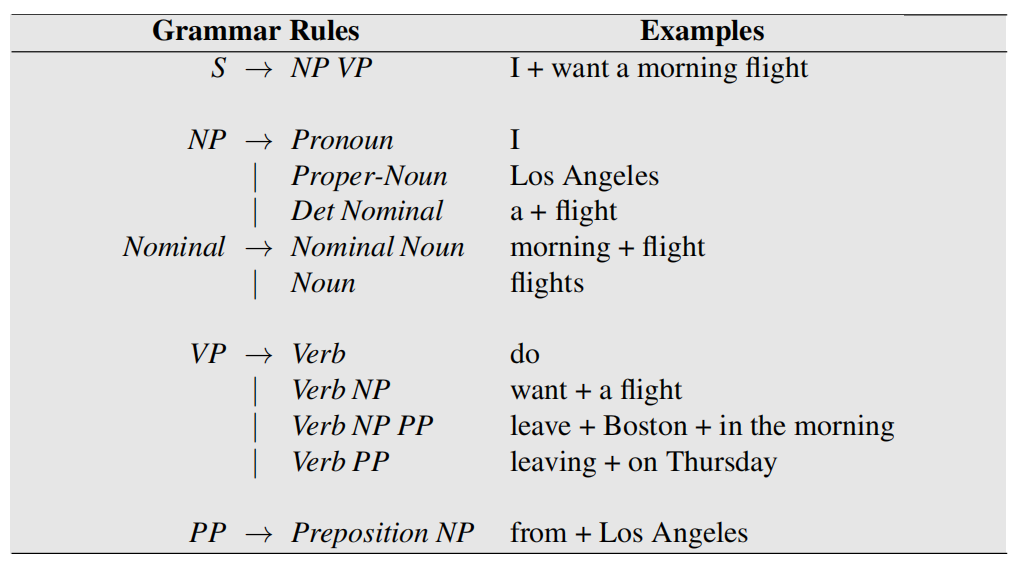




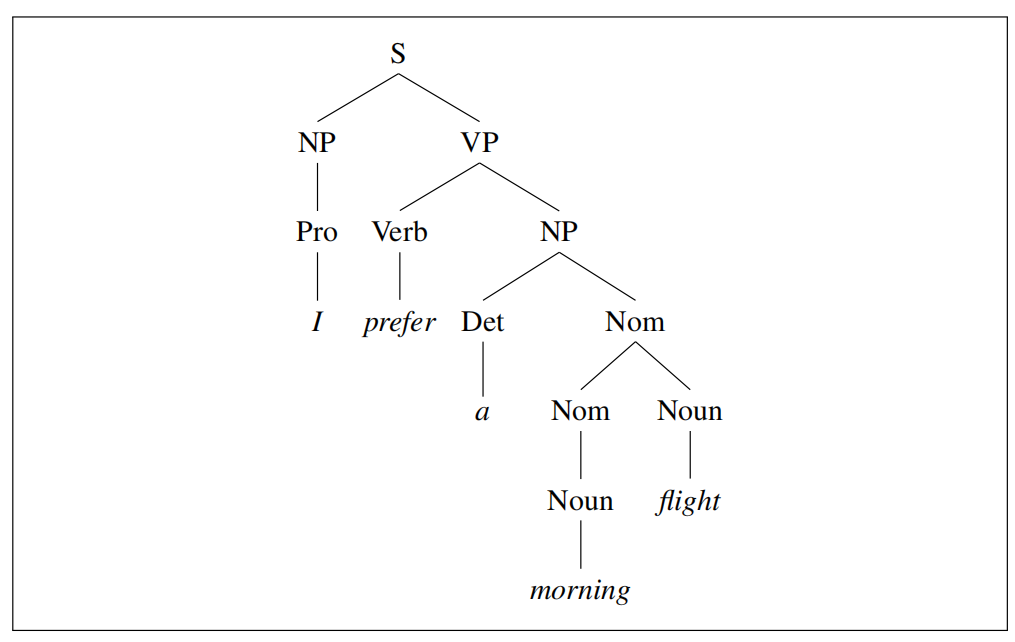


|：表示non-terminal有多个选择。

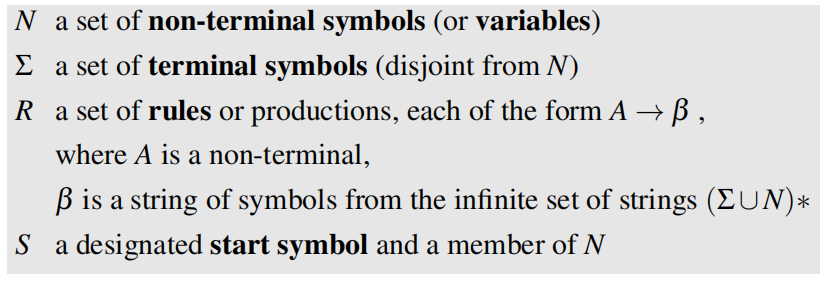


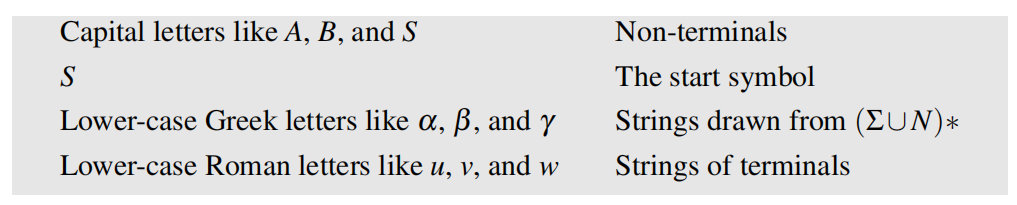


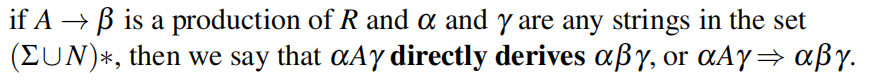
bracketed notation:

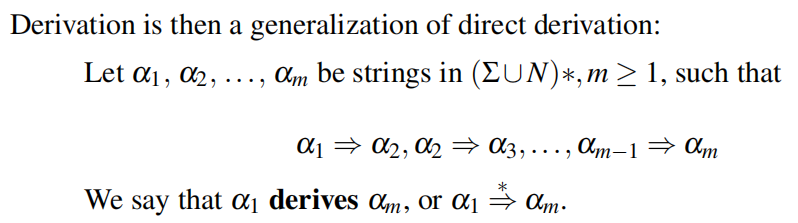


可以通过四个参数定义CFG：











**12.3 Some Grammar Rules for English**

陈述句：主语名词短语后面跟一个动词短语。

祈使句结构：动词短语开始，没有主语。

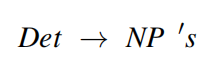
Yes-no question：由助动词开始，之后是宾语名词短语，再跟着一个动词短语。

Wh-phrase：包括wh词，可以分为wh-subject-question：结构和陈述句一致。

Wh-non-subject-question：wh短语不是句子的主语，句子还有一个另外的主语。

Clause：从句，具有一定的完整性，可以用S来表示。

限定词：名词短语可以由简单的限定词开始。也可以是更复杂的形式。



这个规则是可以循环的。

名词性结构：



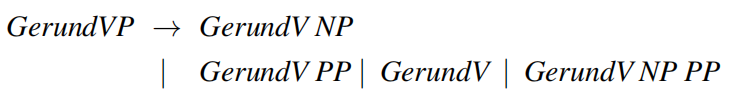
在中心词之前：基数词，序数词，量词，形容词。

在中心词之后：介词短语，非限定从句，定语从句。

三种最常见的non-finite postmodifier：

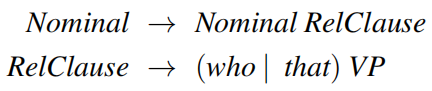
Gerundive postmodifier：由动词的ing形式开始。



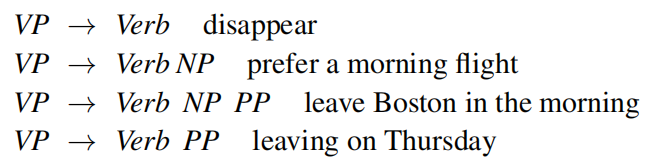


Infinitive clause, -ed form clause

限制性定语从句：以关系代词开始。



动词短语：



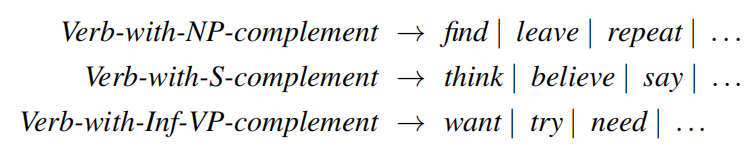
动词短语可以更加复杂：

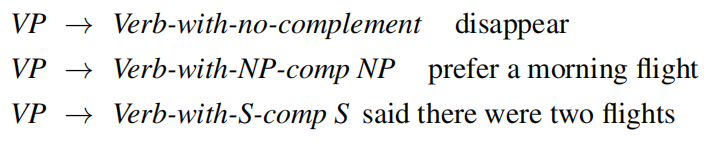


传统的方法将动词分为及物动词或者不及物动词。但现代的方法可以分成100种。

动词和其他成分之间的关系是把动词看作一种逻辑谓词，其他成分看作谓词逻辑参数的成分。

可以将动词和其他成分的关系进行划分：





这种方法的一个问题是增加规则的数目以及降低了泛化性。