speech and language processing 第十二章报告

09017244 郑健雄

1. 自己提出的问题的理解：
2. P8最下面段落的clause的概念具体是指什么？是否有例子？？

讨论后的理解：clause可以理解为从句，比如说S可以作为语法中的从句对象，例如I think S这种组合，是一种重要的构成方式。

1. Treebank都有哪些具体应用？可以用来做什么？？

讨论后的理解：Treebank作为一种句子结构标注的数据库，其可以用来进行相关的NLP工作以及对于句子结构进行深入的理解，是比较传统的思路。

1. constituency grammars有什么优势和实际应用？

讨论后的理解：该语法的优势在于简单直接，效率很高，可以用于直接分析句子的结构或者模式，并且准确率较好，缺点是其需要人为总结，人为分析，所以很多语言现象可能反而不能理解，甚至产生二义性。其最广泛的应用领域是编译原理中的编译器，因为其语言的简洁性和固定性，其语法不是特别复杂。

1. 别人提出的问题的理解
2. P3也提到的terminal symbols和non-terminal symbols具体指什么？。

自己的理解：这两个概念其实在编译原理中用的比较多，terminal symbols不能再进行推导，比如说英语中某个具体单词，而nonterminal可以进一步展开，也可以理解为不是单词的的那些概括符号都是nonterminal，比如说N代表名词，而其可能会有N->things这样的推理规则。

1. 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？

自己的理解：production的话，其实应该可以理解为具体化，比如一开始的形式是vp，然后vp又由verb+np组成，而verb可以是prefer，np可以是a morning flight，就是展开。实际上一般是给定一个英语句子，倒推其语法树，此处的推导就是把倒推后形成的结果再自顶向下展开，但实际上语法树应该是自底向上构造的。

1. 读书计划

1、本周完成的内容章节：阅读完第12章上一部分。

2、下周计划：阅读完speech and processing第12章部分内容。

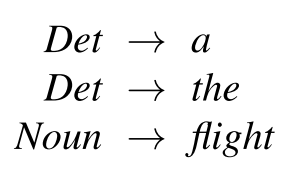
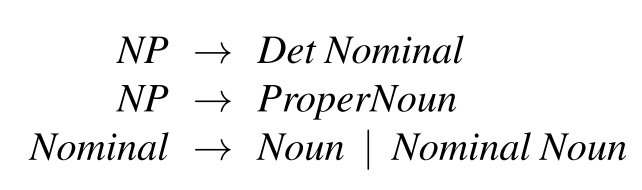
四、读书摘要总结：

1. syntax and Constituency：

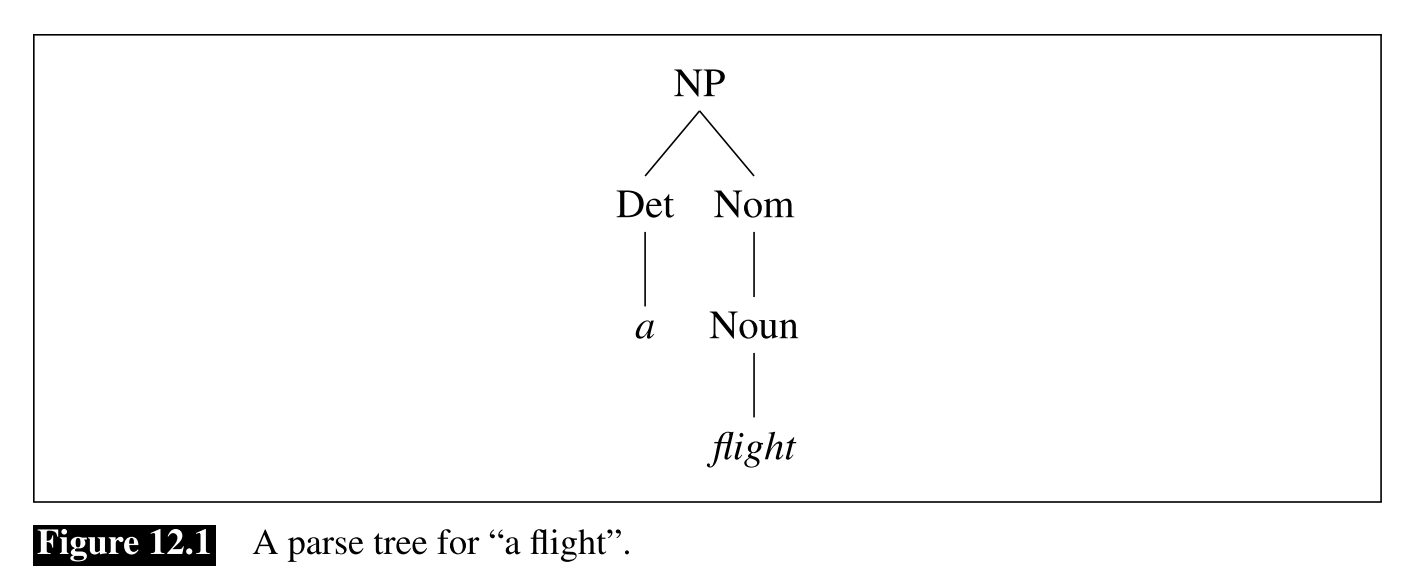
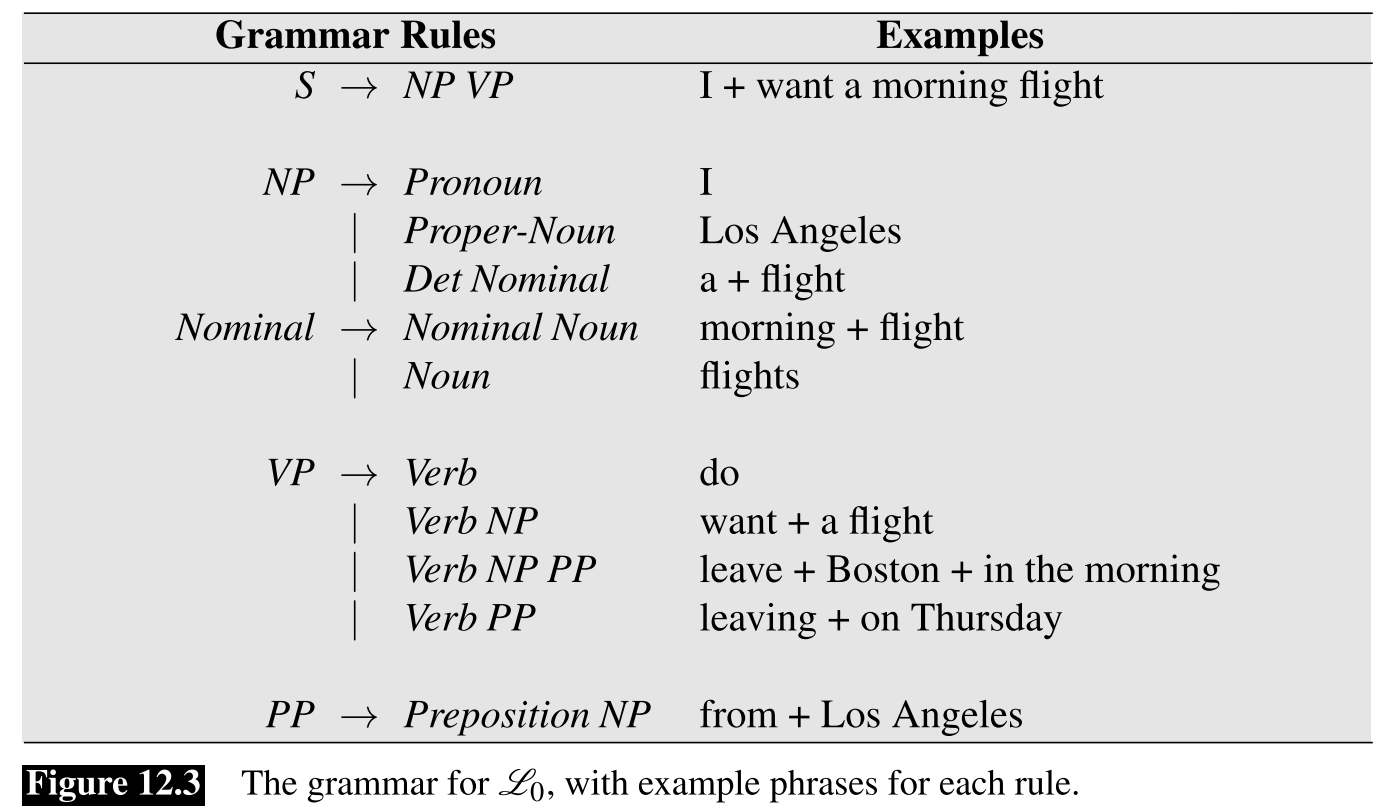
语法是人类对于自然语言构成规律的一种研究，其描述了句子的结构，词性等信息，是一种更加抽象的结构，现在一般有两种语法解析形式，一种是将句子是为多种成分的组合体，也就是constituency grammars，而另外一种则是考虑了句子中词语之间的联系性，是dependency parsing。

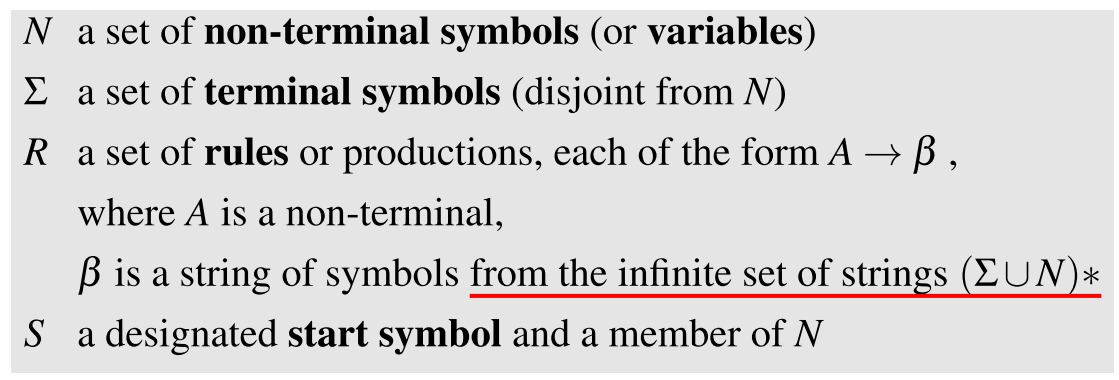
constituents可以理解为独立的可以产生意思的句子单元，比如名词短语或者postposed constructions产生的成分。

2. Context-Free Grammar

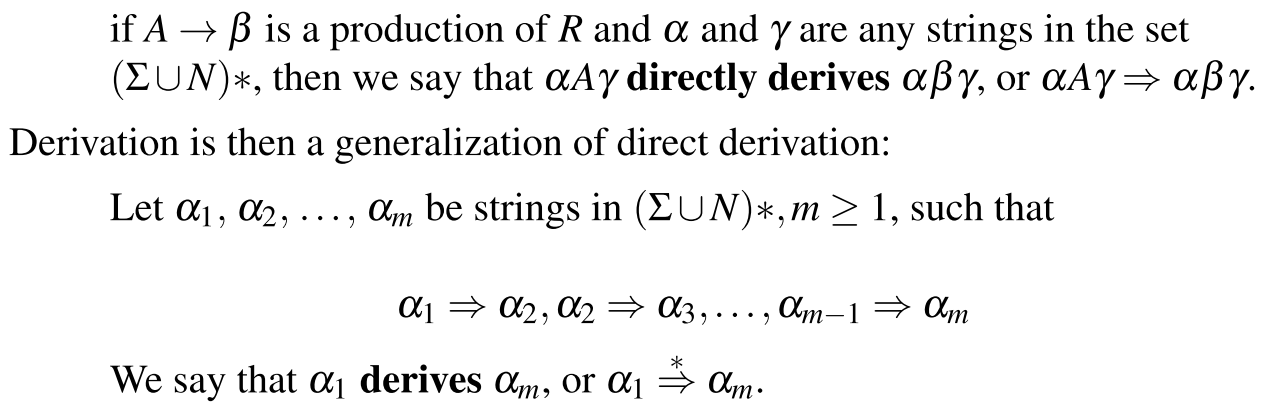
CFG是一种语法结构，其基础是BNF范式，其构成基础是一系列的rules或者说production。例如下面两种： 

语言中的词语被称为terminal symbols，而抽象成分为nonterminal symbols。

CFG可以理解为一个句子的生成器，其产生句子的过程是树形的，称为parse tree，但是也可以根据CFG对特定的句子进行归纳，产生其对应的语法树，所以从这个角度来看的话，分析可以是双向的。 

一般来说，CFG由四个部分构成，

其本质上是状态机。同时其生成规则也比较固定：



从一个句子产生语法树可以被理解为是语法解析。

3. Some Grammar Rules for English

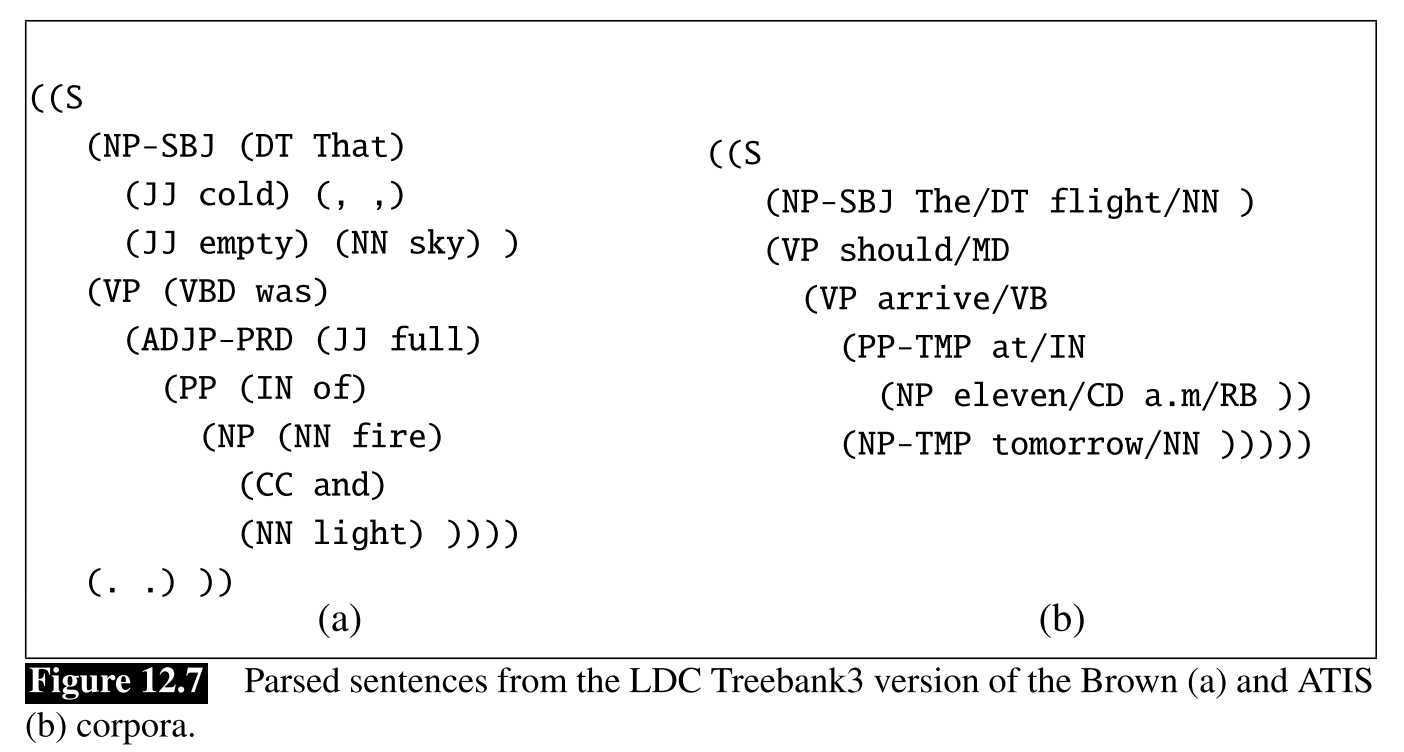
对于英语来说，存在一些常用的语法规则：

对于句子来说，有下面四种规则declaratives, imperatives, yes-no questions, and wh-questions.。其中wh-questions存在longdistance dependencies，因为主语和谓词之间距离很大。

clause可以理解为从句，是语法中比较重要的成分。除此之外，还有noun phrase，verb phrase和连词等成分。

4. Treebanks

其可以理解为分析过语法树结构的语料库，是一种重要的数据资源。比如著名的Penn Treebank数据库，有着丰富的数据。



总结：尽管近期在NLP领域中有着深度学习的系统，其可以在诸多应用中产生重要的作用，但是如果进一步帮助计算机理解语言的结构和意义，进一步的对原理进行研究是十分必要的。比如对于语法的研究，对于语言意义原理的研究，对于其表达意义的单位的研究都是很重要的。其会为未来的研究提供新的思路，以及对深度学习的建模产生影响，帮助人们考虑更多的要素。或者也可以结合深度学习模型辅助研究语法模式和潜在的知识。