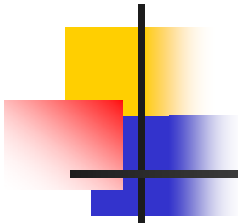


句法结构树的表示

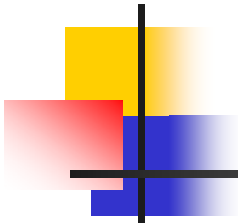
詹卫东

<http://ccl.pku.edu.cn/doubtfire/>



层次、关系、范畴、中心

- 关于一个句子的句法结构树具体该如何画，一种语言中需要定义多少词类，多少短语类，存在许多不同的处理方式。
- 从共性的角度来说，无论句法结构树的具体表现如何不同，句法结构树都需要包含四个要素，才能较为全面地反映一个句子的结构面貌：层次、关系、范畴、中心。



层次、关系、范畴、中心

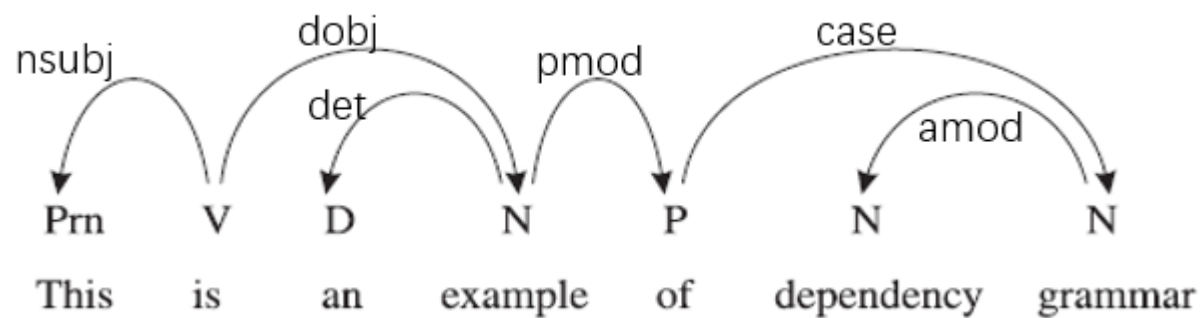
- (1) 句子不是表面的词语线性序列，句子是有层次的，层次是句子的基本结构特征；
- (2) 各个层次上直接成分之间的组合存在不同的关系，句子中哪些词之间没有关系，哪些词之间有关系，具体是什么关系，也是句子的基本结构特征；
- (3) 由于句子的无限性，句子的层次构造需要在对词和短语进行分类（范畴化）的基础上进行描述，才能做到以简驭繁，以有限表达无限；
- (4) 两个直接成分组合时，往往有一个是中心，中心成分的范畴属性更能代表整个结构的范畴性质。



层次、关系、范畴、中心

- (1) 这四个要素中，层次和关系，是最基础的两个要素，独立于语法理论体系，或者说，任何一个语法理论体系，在描述自然语言句子的构造时，必然要考虑这两个要素。
- (2) 范畴和中心，则是为了知识表示的概括性和便利性所做的进一步的理论假设，是补充性的要素。
- (3) 不同的句法理论体系的设计，对于句法结构“范畴”和句法结构“中心”的认识，可能存在比较大的差异，使得不同人画出来的句法结构树看上去不一样。
- (4) 即使在同一语法理论体系内部，不同学者也可能存在不同的认识，比如汉语语法学界对于“这本书的出版”中的“出版”属于动词范畴还是名词范畴，就有争议；再比如“弄-坏”中哪个词是结构中心？“同意去-的”中是“同意去”是结构中心，还是“的”是结构中心？都可能存在不同看法。

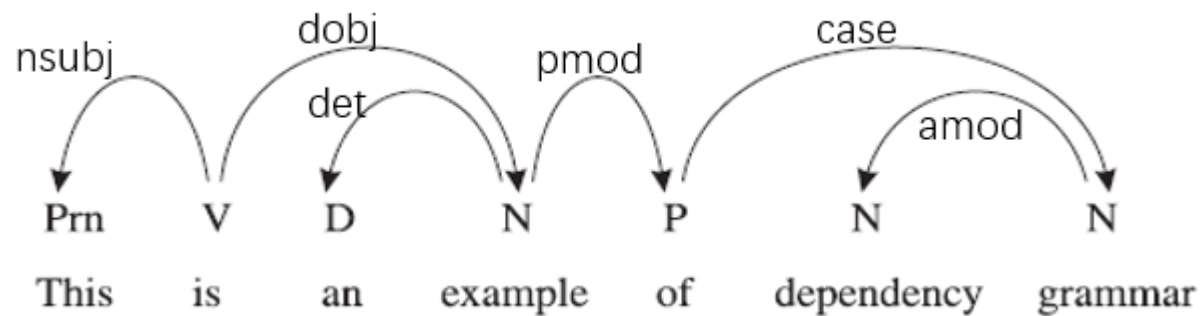
依存语法的表示



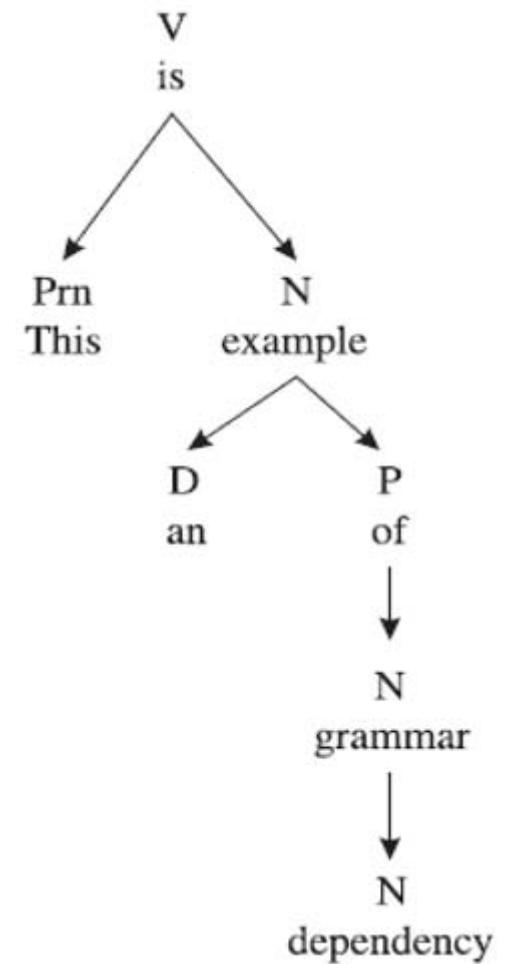
A

Michael A. Covington, 2001, A fundamental algorithm for dependency parsing, In Proceedings of the 39th Annual ACM Southeast Conference.

依存语法的表示



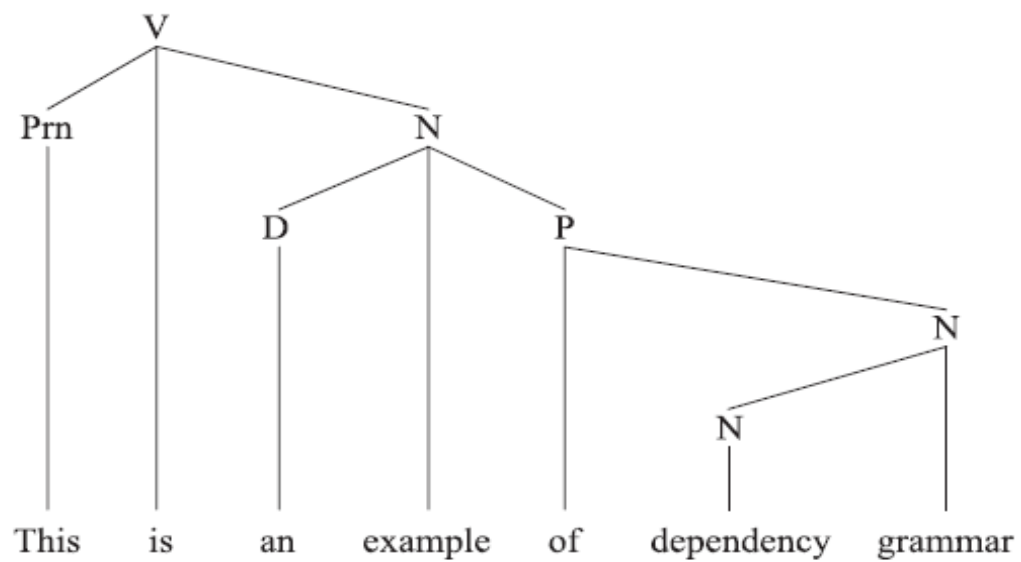
A



B

Michael A. Covington, 2001, A fundamental algorithm for dependency parsing, In Proceedings of the 39th Annual ACM Southeast Conference.

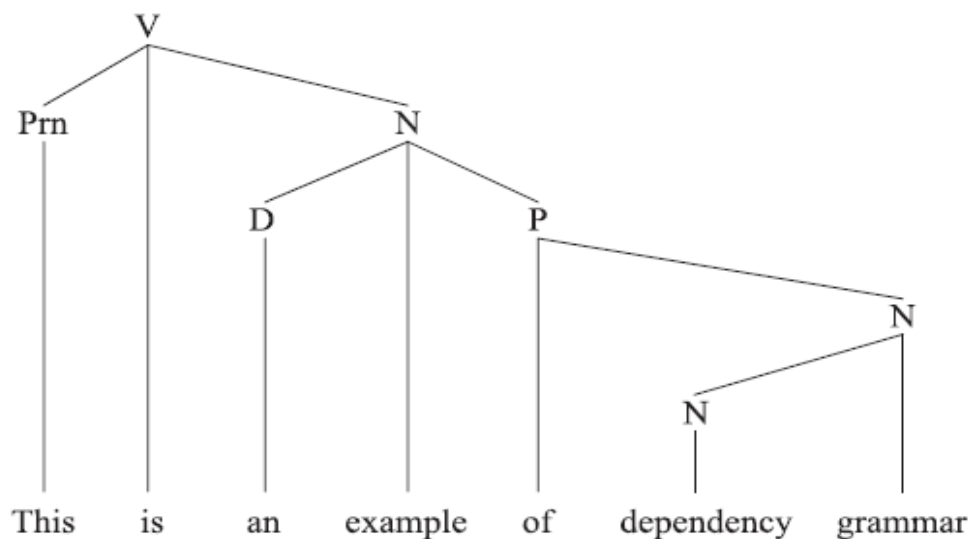
依存语法的表示



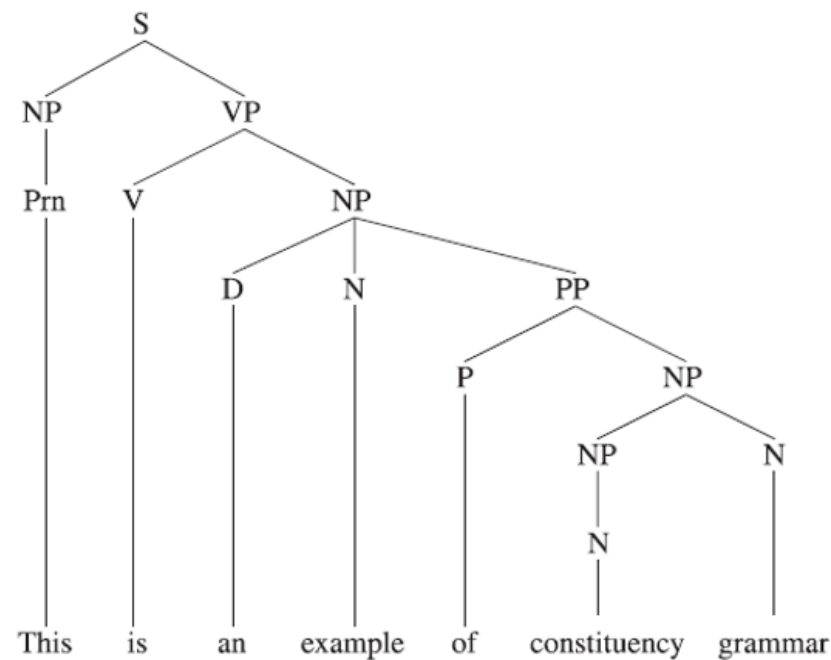
C

Michael A. Covington, 2001, A fundamental algorithm for dependency parsing, In Proceedings of the 39th Annual ACM Southeast Conference.

短语结构语法的表示



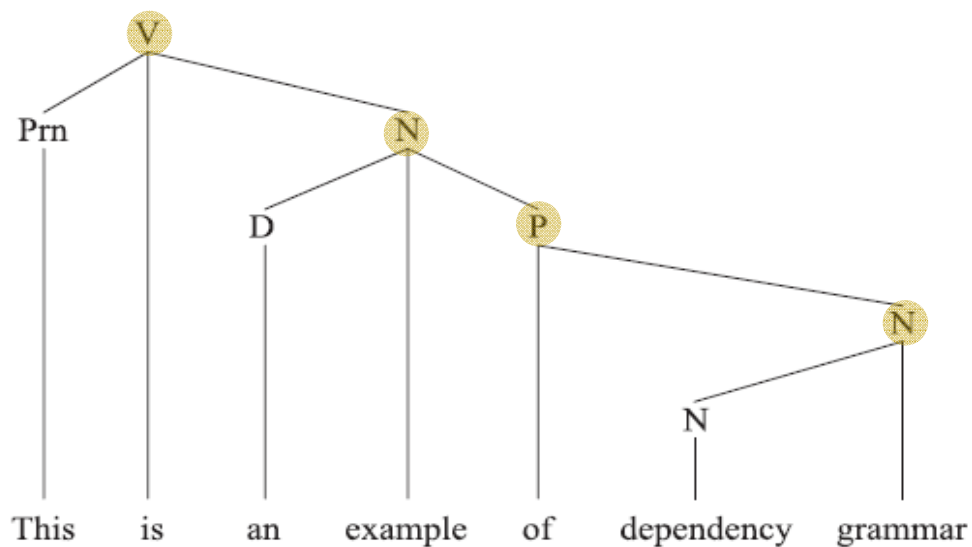
C



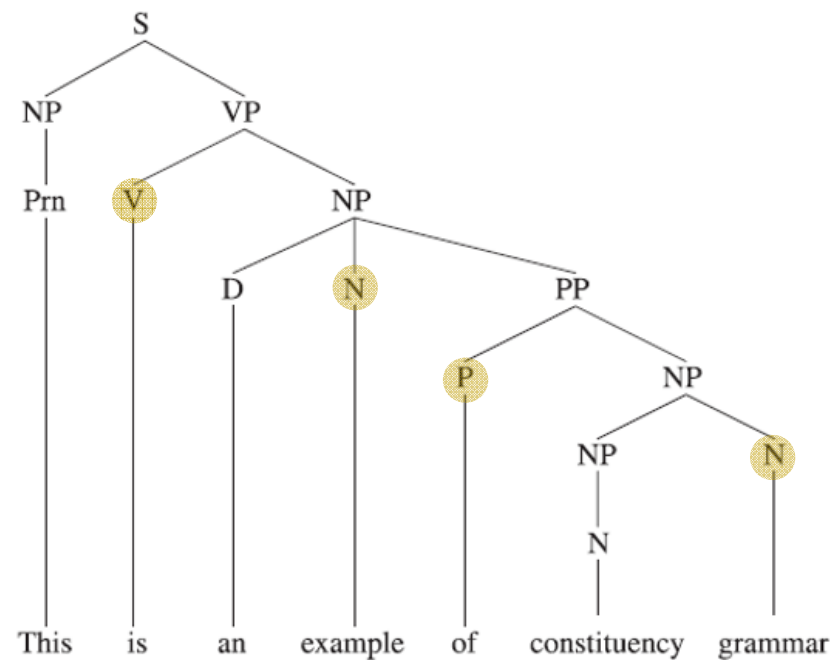
D

Michael A. Covington, 2001, A fundamental algorithm for dependency parsing, In Proceedings of the 39th Annual ACM Southeast Conference.

短语结构语法的表示

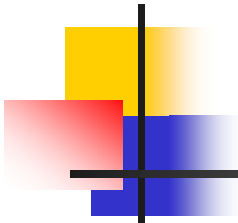


C



D

Michael A. Covington, 2001, A fundamental algorithm for dependency parsing, In Proceedings of the 39th Annual ACM Southeast Conference.



认识四要素的不同视角

- 自顶向下的视角：层次、关系 | 范畴、中心
 - 自然语言句子不是线性结构，是层次结构（树结构）；
 - 树的节点之间有不同关系；
 - 根据关系的不同，节点可以聚类成为不同的范畴；
 - 在关系中，有1个节点为中心成分，其余节点为附属成分。
- 自底向上的视角：关系 → 中心 → 范畴 → 层次
 - 两个成分组合，有不同的关系类型；
 - 在关系中，有1个成分为中心成分，其余成分为附属成分；
 - 一个组合关系中的不同位置上，词语聚合成类，形成范畴；
 - 组合关系可以嵌套，形成层次。