

虚词知识表示的形式化框架刍议

詹卫东

北京大学中文系

北京大学汉语语言研究中心

2004.11.7

提 纲

- 一 描述虚词知识的作用是什么？
- 二 需要描述哪些虚词知识？
- 三 如何描述虚词知识？
- 四 虚词知识表示的困难

一 描述虚词知识的作用是什么？

不但我没见过，他也没见过



我不但没见过，他也没见过



他们是不是在开会呢？



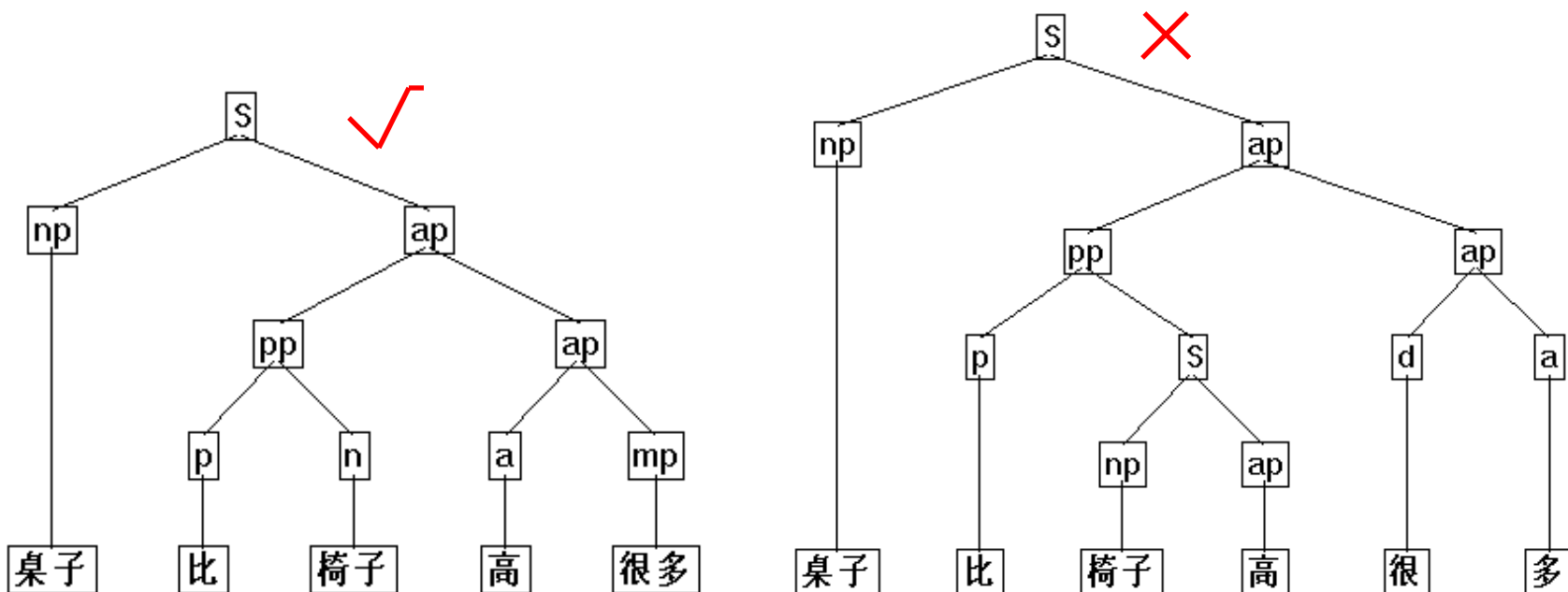
他们是不是在开会吗？



① 判断句子的合语法性

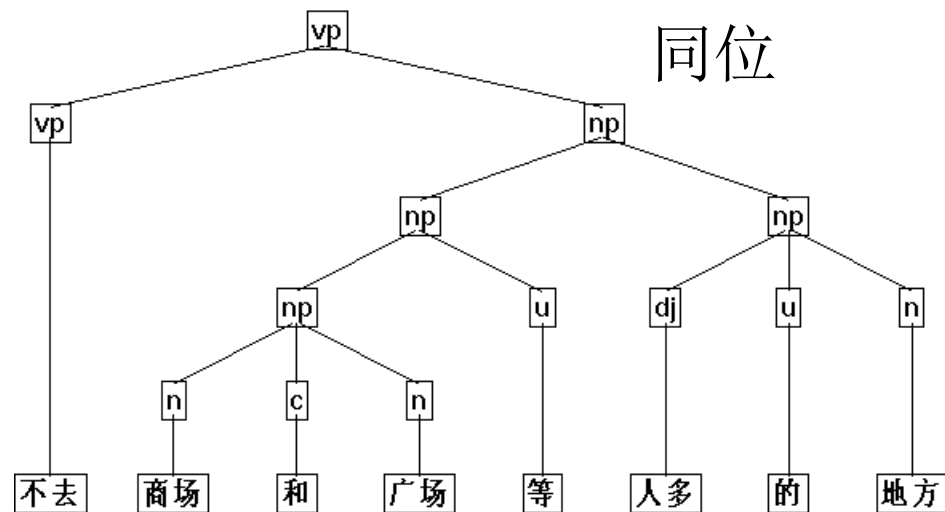
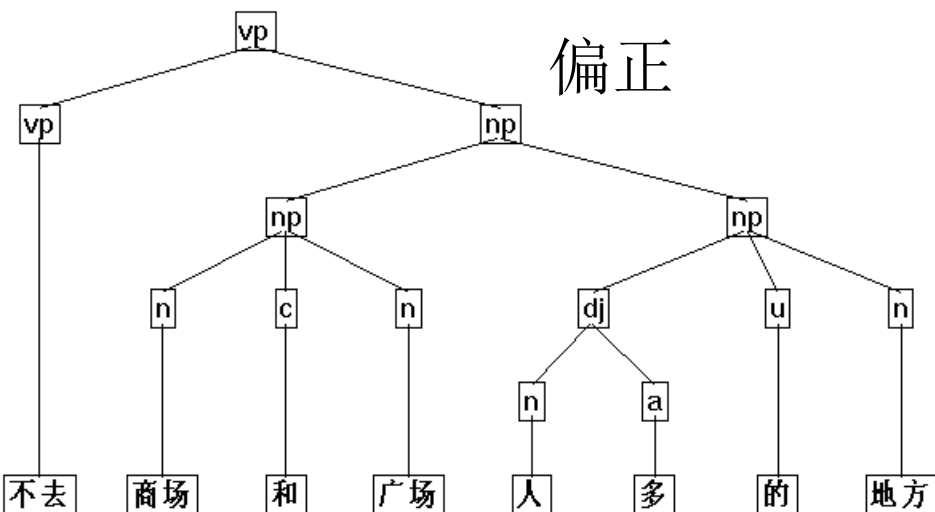
描述虚词知识的作用

- 桌子比椅子高很多



② 帮助分析层次结构

描述虚词知识的作用



③ 帮助分析句法结构关系

描述虚词知识的作用

张三差点儿**没**摔下来 == 张三差点儿**摔**下来
↓ ↓
张三**没**摔下来

张三差点儿**没**考上大学 ≠ 张三差点儿**考**上大学
↓ ↓
张三**考**上了大学 张三**没有**考上大学

④ 句子意思深层理解的需要

二 需要描述哪些虚词知识？

- 虚词的句法结构位置、隐现规律
 - 多义虚词的义项分析及判定条件
 - 对同现实词成分的句法、语义约束
 - 虚词对句子语义深层理解的制约
- } 组合约束

↓
寻找同现成分之间的约束条件，比如：

“了”要附在什么样的成分上？

“X+了”有什么句法或语义性质？

.....

虚词的句法结构位置、隐现规律

- 他连小学都没念过 —— 连小学他都没念过
- 他向窗外看了看 —— * 向窗外他看了看

- * 你这样做是合情合理 —— 你这样做是合情合理的
- 你这样做是偏听偏信 —— * 你这样做是偏听偏信的

一个综合的例子

累计

- A 为了数据正确，我们还（*再）要做一次实验。

重复

- B 为了数据正确，我们要再（*还）做一次实验。

追加

- C 我还（再）要两个包子

	追加	重复	累计
还	+	-	+
再	+	+	-

	语义	句法
要 ₁	将要	后跟vp
要 ₂	索要	后跟np

A 可能要做一个新的试验

倾向

B 已经做过一个相同的试验₉

一个综合的例子

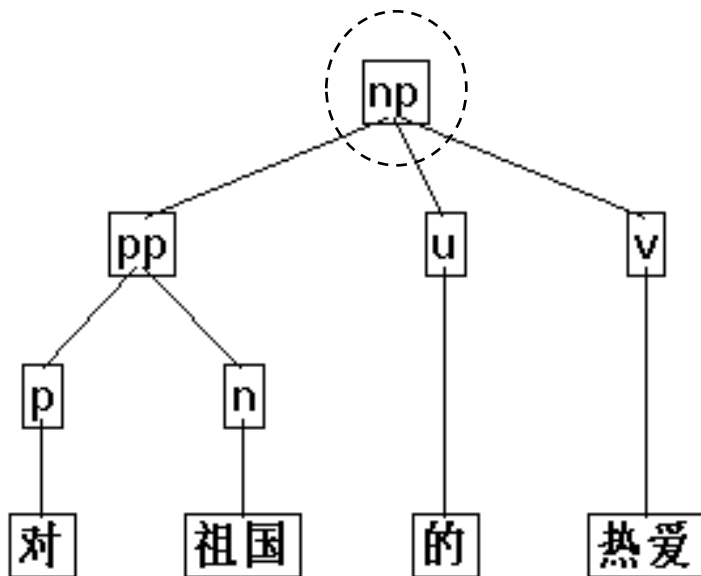
- 你再（*还）坐一会儿吧，外面还（*再）下着雨呢？

还：（不间断）持续 + 进行体 - 时量/动量宾语

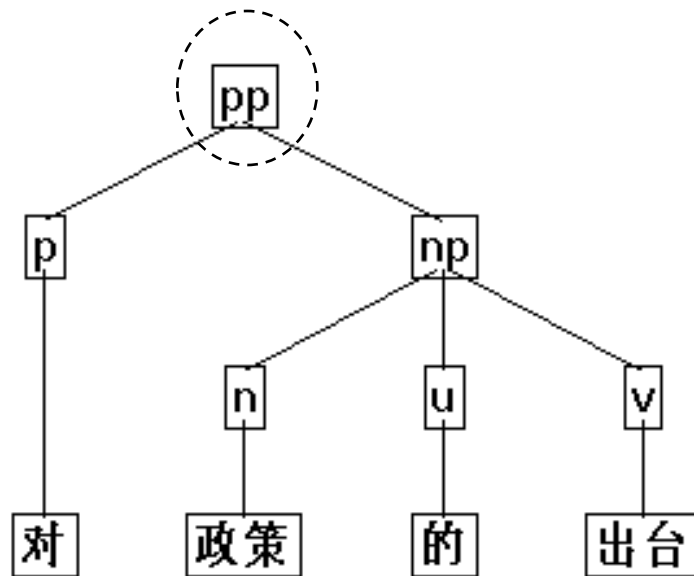
再：（间断）重复 - 进行体 + 时量/动量宾语

可以不依赖虚词知识的例子

对祖国的热爱



对政策的出台



=/=

他对校长的意见很大

他对校长的批评很尖锐

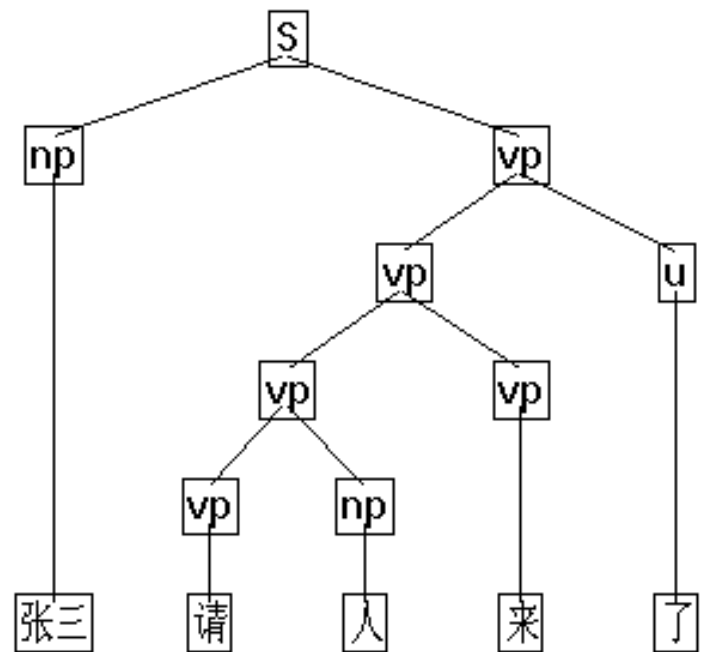
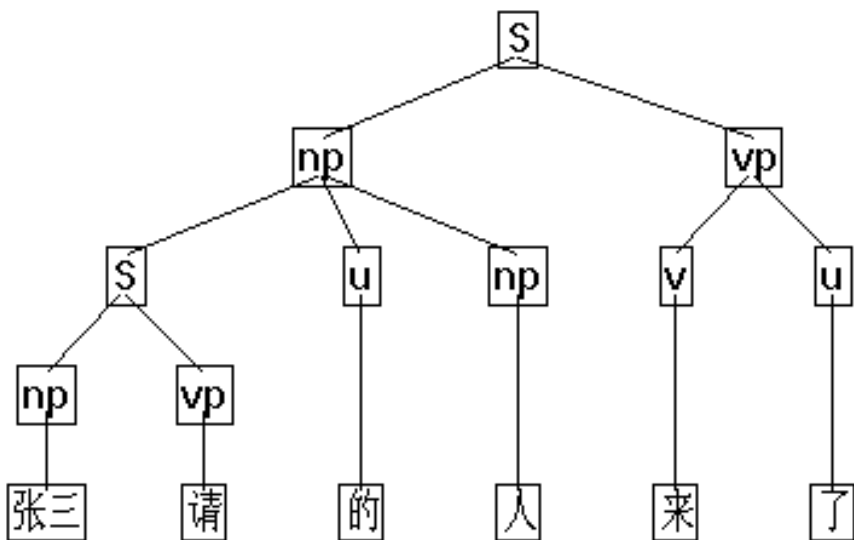
他对校长的意见持否定态度

他对校长的批评充耳不闻

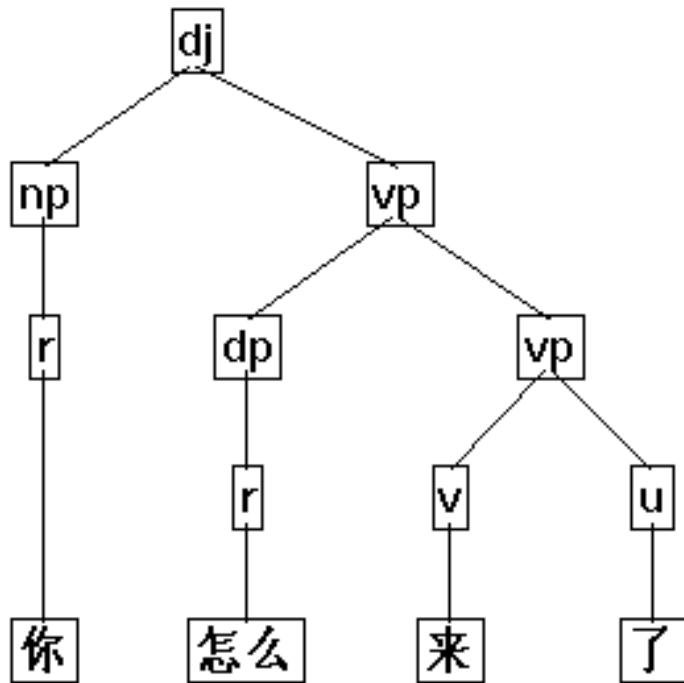
可以不依赖虚词知识的例子

- 张三请的人来了

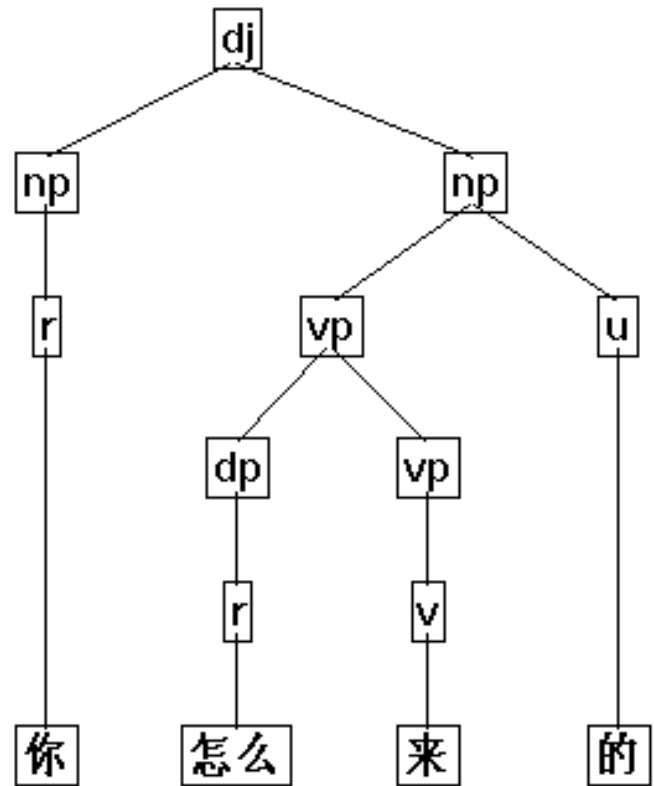
张三请人来了



怎么： why or how?

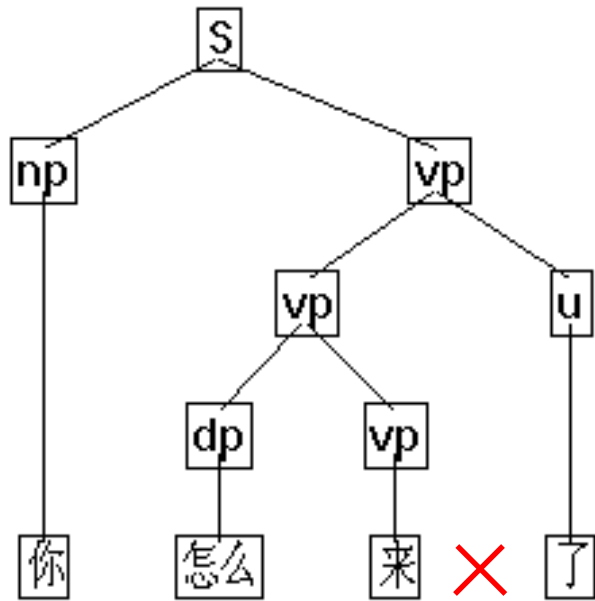


why

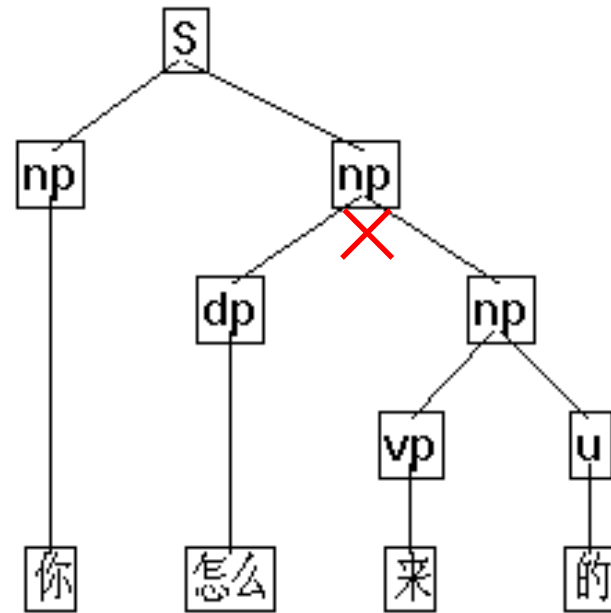


how

怎么： why or how?



?



np ~~np~~ dp np

三 如何描述虚词知识？

- 特征结构
- 带合一约束的CFG规则
- 树结构转换

特征结构

连词的语法属性描述 [特征：值] 形式

词语	前中后	连体	连谓	连句	连主谓
不但	前		谓	句	可
和	中	体	谓		
可	后		谓	句	
与其	前		谓	句	可
...					

《现代汉语语法信息词典》

带合一约束的CFG规则

ap-> pp<比> ap :: IF %ap.可比性 = NO FALSE

ap-> d<很> a :: \$.可比性 = NO

ap-> d<不> a :: \$.可比性 = NO

张三很胖 -- * 张三比李四很胖

张三不胖 -- * 张三比李四不胖

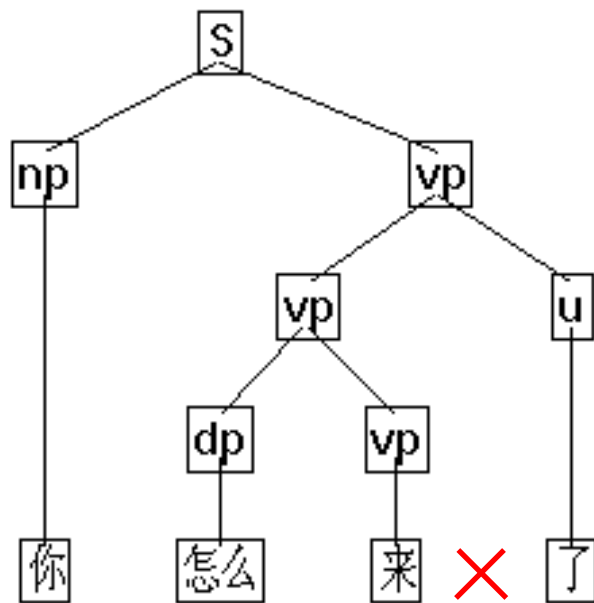
带合一约束的CFG规则（续）

```
np -> np np  :: IF #GetLeafNode(%np,-1).原形=等, %np.语义类=%np.语义类 THEN $.内部结构=同位 ENDIF,  
    IF #GetLeafNode(%np,-1).原形=等等, %%np.定语.cpcat=mp  
FALSE
```

虚词包括副词、介词、连词、助词、语气词等五个词类。

* 虚词包括副词、介词、连词、助词、语气词等等五个词类。

怎么： why or how?

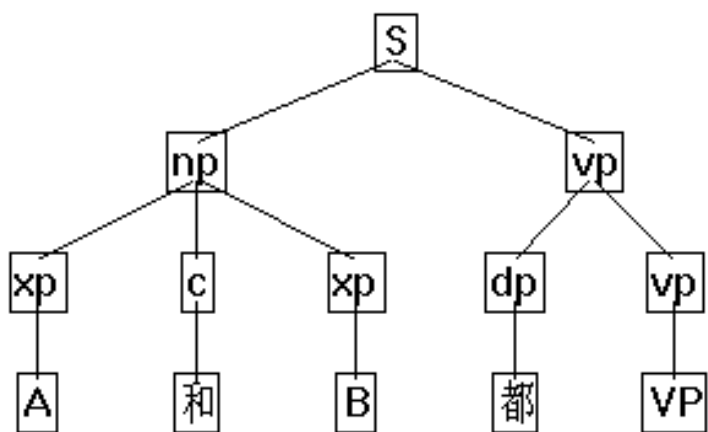


vp -> vp u<了> :: IF %vp.Tense=将来 FALSE

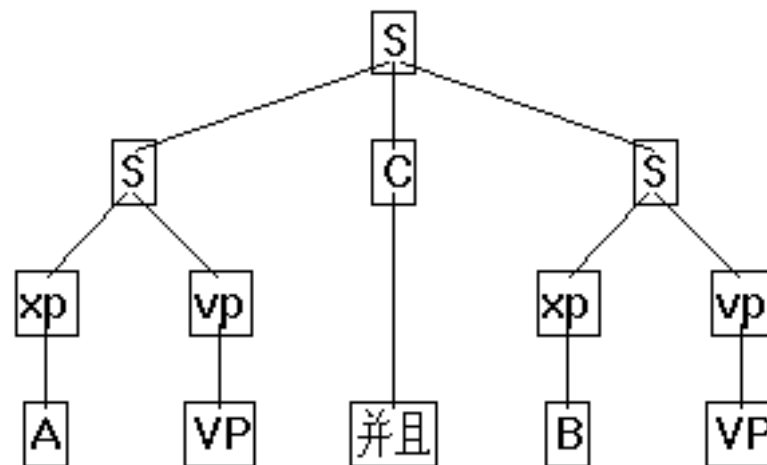
vp -> dp<怎么> vp :: \$.Tense=将来

树结构转换：语义解释的一种表达形式

- A 和 B 都 VP \rightarrow A/A VP/VP 并且 B/B VP/VP



张三和李四都结婚了



张三结婚了 并且 李四结婚了

树结构转换的条件约束

- 张三和李四送了红包
- 张三和李四送的红包

S [np1 + vp + np2]



- 张三和李四都送了红包
- ? 张三和李四都送的红包

NP [np1 + vp + 的 + np2]

- 张三和李四都喜欢那件衣服
- 张三和李四都喜欢的那件衣服

树结构转换的条件约束

- 张三和李四都有一个毛病
- 张三和李四都有的一个毛病
- 张三和李四都有一个女儿
- * 张三和李四都有的一个女儿
- 张三和李四都买得起的房子

虚词对歧义结构的制约

批评 张三 的 老师 -- 都 批评 张三 的 老师

喝 的 牛奶 -- 都 喝 的 牛奶

坐 的 飞机 -- 都 坐 的 飞机

四 虚词知识表示的困难

肉煮烂了 —— 肉被煮烂了

米卖完了 —— 米被卖完了

木头房子 —— 木头的房子

我父亲 —— 我的父亲

人群中间 —— 人群的中间

“N N” vs. “N 的 N”

- 一处别致的农村山庄吸引了大家的目光，没有现代化的设备，加上周围的森林绿地，给人以一种非常遐逸的感觉，**木头的房子**，树枝编成的装饰及高高悬挂在树上的同样是用树枝编成的鸟窝，虽然“原始”但是感觉并不很“原始”。
- 农场突然地遭到一股强龙卷风的袭击，叔叔、婶婶和其他的人们都躲进旋风地洞后，可怜的多诺茜被龙卷风连同叔叔的**木头房子**一起抛向了空中 ...

“N N” vs. “N 的 N”

- 1) 好莱坞电影能在全世界不同政见、不同文化的人群中间引起兴趣，有它的历史渊源
- 2) 我的心又一次被唤醒我站在这里想起和你曾经离别情景你站在人群中间那么孤单 ...
- 3) 国际著名传染病学专家、世界卫生组织顾问（WHO）、霍普金斯大学教授赖声汉先生从西医传染病学学的角度，对非典型性肺炎SARS在人群中间的传播、演变以及控制和治疗作了深入浅出的讲解、分析。

“N N” vs. “N 的 N”

- 孩子脾气 vs. 孩子的脾气

这孩子脾气倔

你这孩子脾气一定要改

他只是耍耍孩子脾气

? 先要摸透孩子脾气

这孩子的脾气倔

你这孩子的脾气一定要改

? 他只是耍耍孩子的脾气

先要摸透孩子的脾气

他常常发孩子的脾气

虚词语义对context的依赖

- A 张三一天就挣八十块钱，大家都很羡慕他
- B 张三一天就挣八十块钱，大家都很同情他
- C 张三一天就挣八十块钱，他这日子该怎么过啊？

参考文献

- B.Ludwig, G.Gorz, H. Niemann, 1999, A Computational Approach to the Semantics of Function Words in Dialogue Context, Proceedings of VEXTAL'99 – Venezia, San Servolo, V.I.U. – 22-24 Nov. 1999.
- 侯学超, 1998, 《现代汉语虚词词典》, 北京大学出版社。
- 陆俭明、马真, 1999, 《现代汉语虚词散论》(修订版), 语文出版社。
- 俞士汶等, 2003, 《现代汉语语法信息词典详解》(第二版), 清华大学出版社。
- 俞士汶, 2002, 《现代汉语广义虚词知识库的建设》, 新加坡国立大学“第二届肯特岗国际汉语语言学圆桌会议”, 2002年11月。

谢谢

敬请批评指正