

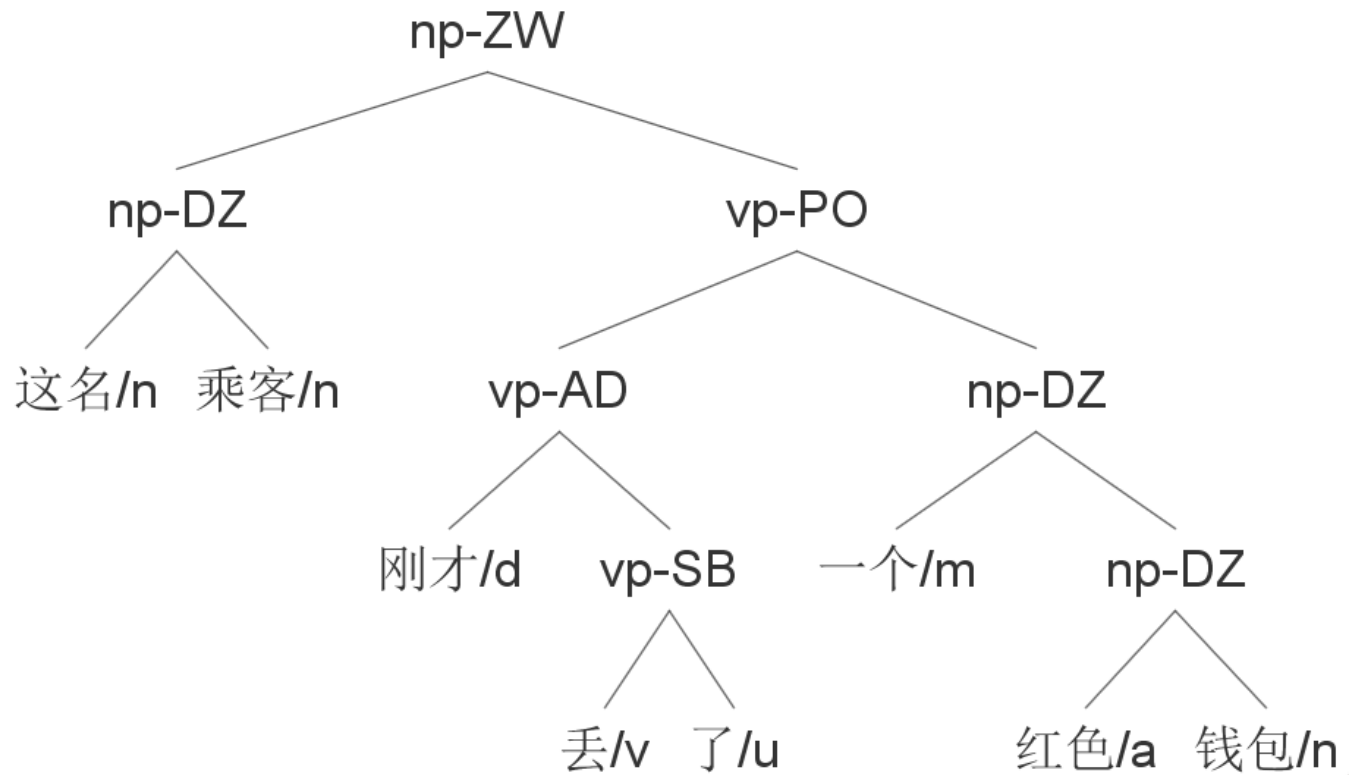
事件句式的推导树重构

邱晗

2014.5.19

工作背景(1/6)

- TCT



工作背景(2/6)

- 事件句式 (EC)
 - 汉语小句中能够完整表达该事件，动作等的最小结构
 - [S-np 乘客] [P-vp 丢] [O-np 钱包]

工作背景(3/6)

- 句法结构信息
 - [S-np 菜刀] [P-vp 切] [O-np 蔬菜]
 - [S-np 工作] [S-np 职员] [P-vp 完成]

工作背景(4/6)

- 谓词论元结构（PA）
 - EC中隐含的深层次的语义组合结构同浅层的句法的主谓宾等成分对应起来
 - 施事和受事
- 隐含谓词论元结构（IPA）
 - 句法结构信息同深层语义组合信息不一致的主要原因

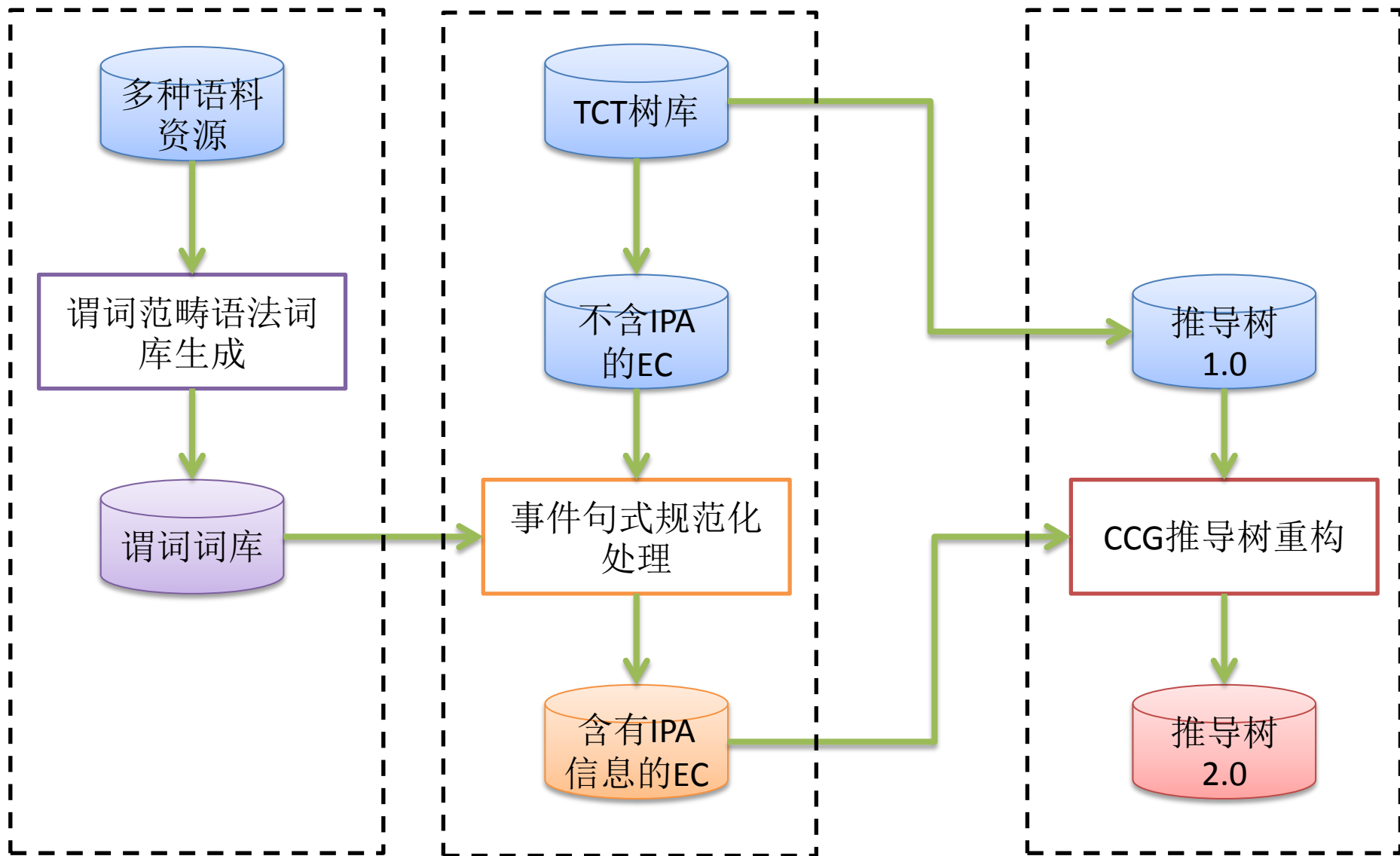
工作背景(5/6)

- 句法结构信息 → 语义组合信息
 - [S-np 菜刀] [P-vp 切] [O-np 蔬菜] →
 - [S-np 菜刀] [P-vp-Pred(2) 切] [O-np-Arg2 蔬菜]

 - [S-np 工作] [S-np 职员] [P-vp 完成] →
 - [S-np-Arg2 工作] [S-np-Arg1 职员] [P-vp-Pred(2) 完成]

工作背景(6/6)

- 理论意义
 - 提取出施事和受事内容信息
 - 建立同主谓宾等句法成分的映射关系
 - 句法过渡到语义
- 应用意义
 - 语义角色标注（Semantic Role Labeling）
 - 提取出各个论元的语义角色
 - J Choi(2011)
 - 机器翻译（Machine Translation）
 - 多义词判断，常用搭配的翻译
 - D Xiong(2012) 基于统计的翻译模型
 - 信息抽取（Information Extraction）
 - O Etzioni(2011)
 - 自动问答（Question and Answering）
 - 构建QA信息库，问题的分析



谓词CCG词库构建

EC规范化处理

CCG推导树重构

推导树重构方式(1/6)

- 原子范畴
 - S, NP, PP, SP, TP
- 复杂范畴
 - X/Y或X\Y
 - S\NP: 睡觉
 - (S\NP)/NP: 研究
 - (S\NP)\PP: 结婚

推导树重构方式(2/6)

- 规则组合

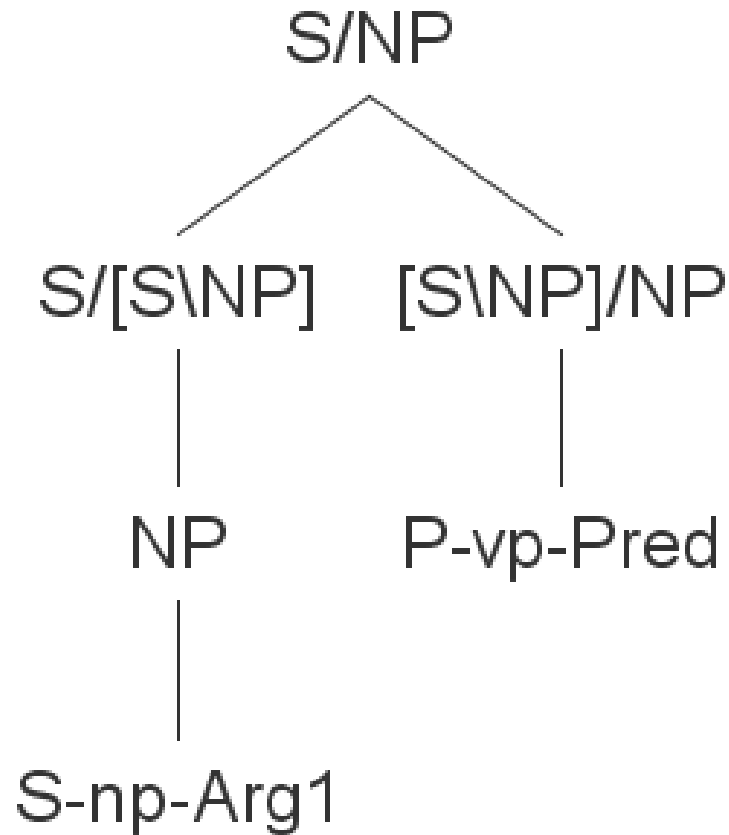
- 前向应用(*Forward Application*): $X/Y \quad Y \rightarrow X$
- 后向应用(*Backward Application*): $Y \quad X \setminus Y \rightarrow X$
- 前向组合(*Forward Composition*): $X/Y \quad Y/Z \rightarrow X/Z$
- 后向组合(*Backward Composition*): $Y \setminus Z \quad X \setminus Y \rightarrow X \setminus Z$
- 前向类型提升(*Forward Type-raising*): $X \rightarrow T/(T \setminus X)$
- 后向类型提升(*Backward Type-raising*): $X \rightarrow T \setminus (T/X)$

推导树重构方式(3/6)

- 共指信息
 - $(NP/NP) \setminus (S \setminus NP) \{CI: 2=3\}$
 - $((S \setminus NP) / ((S \setminus NP) / NP)) / NP \{CI: 3=4\}$

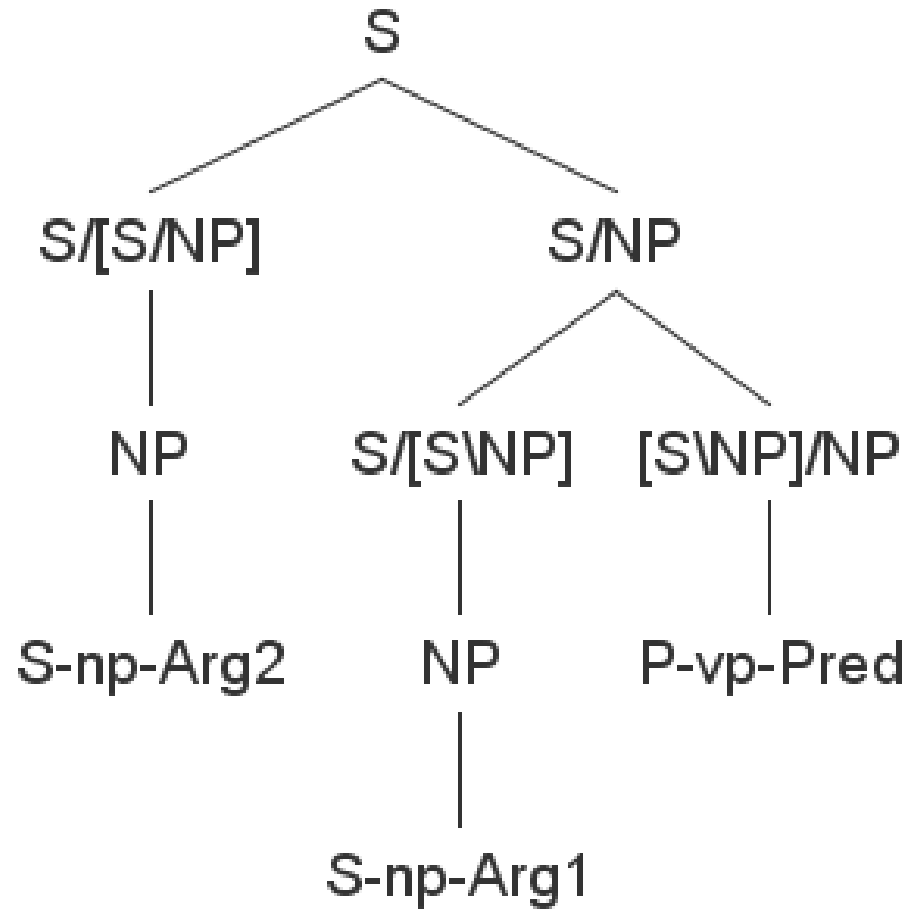
推导树重构方式(4/6)

- SP宾语省略



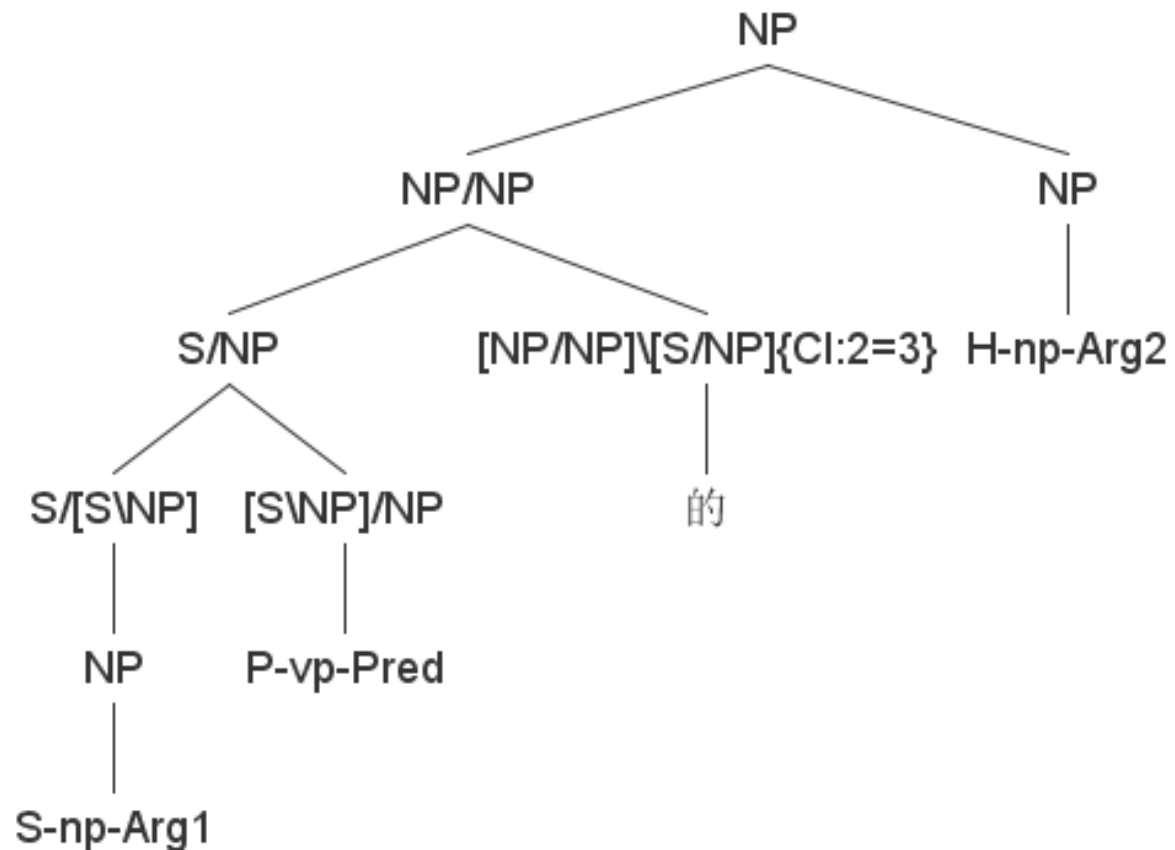
推导树重构方式(5/6)

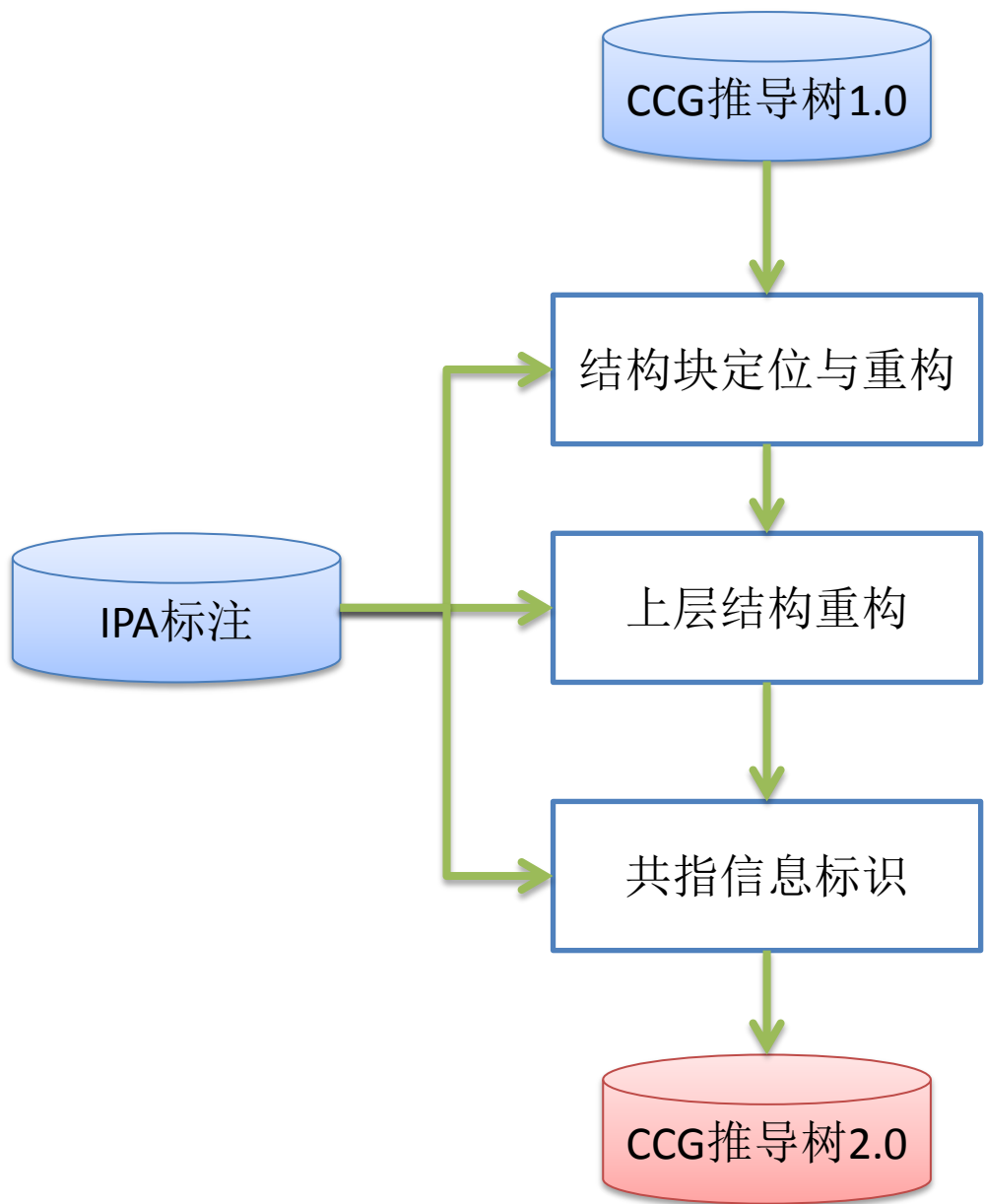
- SSP宾语话题化



推导树重构方式(6/6)

- SP_H宾语中心语



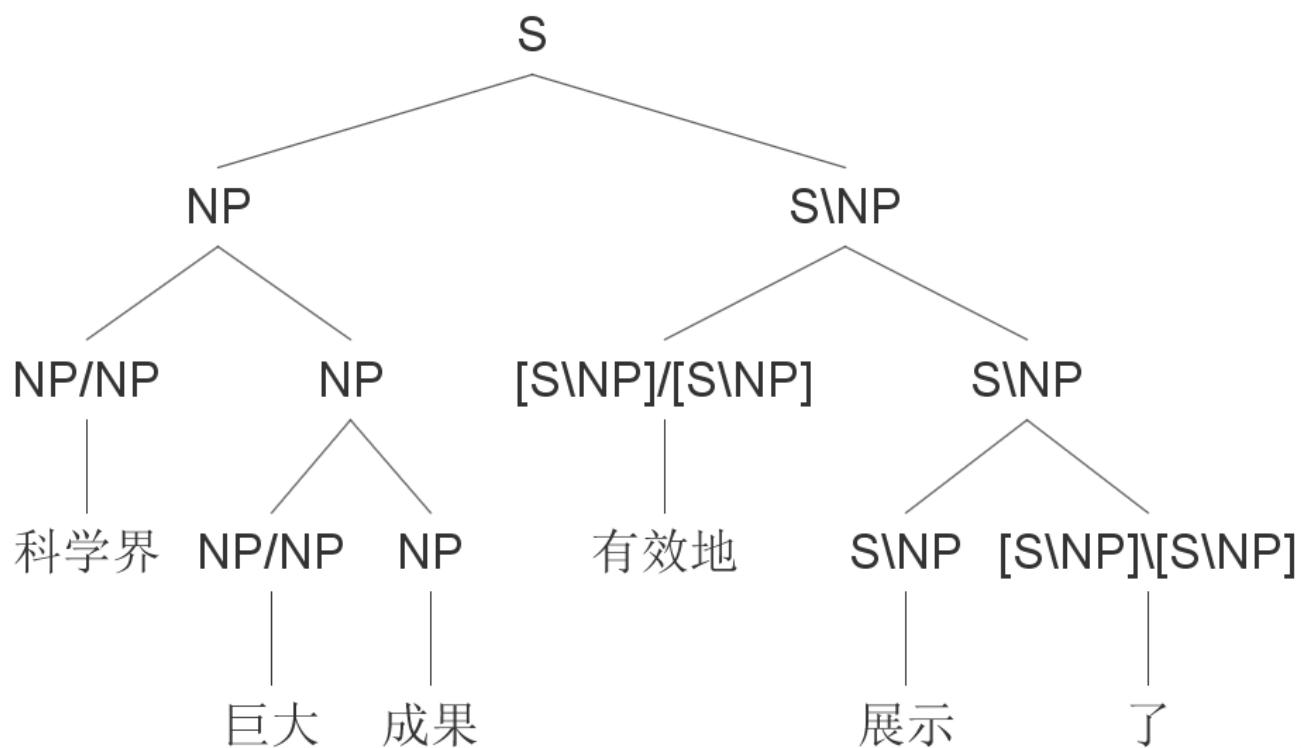


推导树重构算法(1)

- 结构块定位

- IPA标注: [S-np-Arg1 成果] [P-vp-Pred(2) 展示]

- 推导树1.0

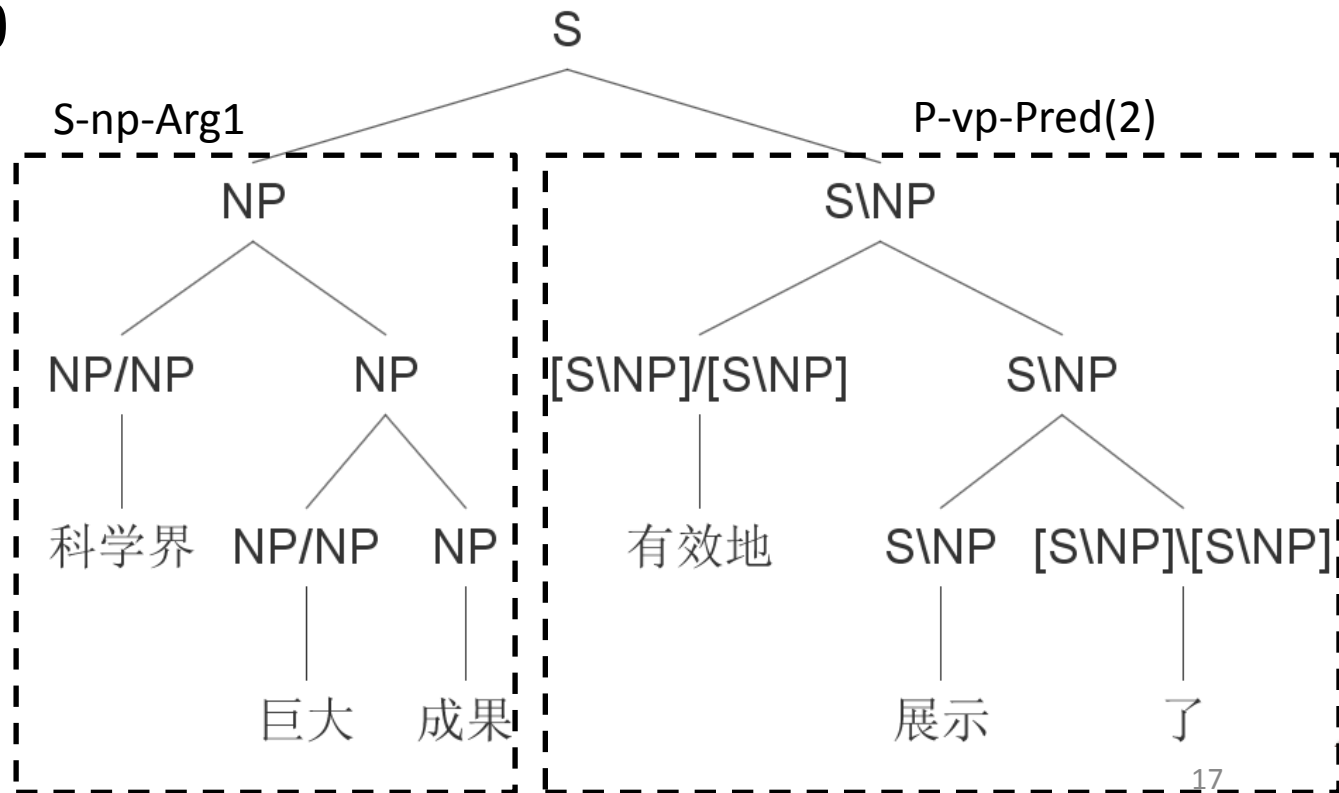


推导树重构算法(1/5)

- 结构块定位

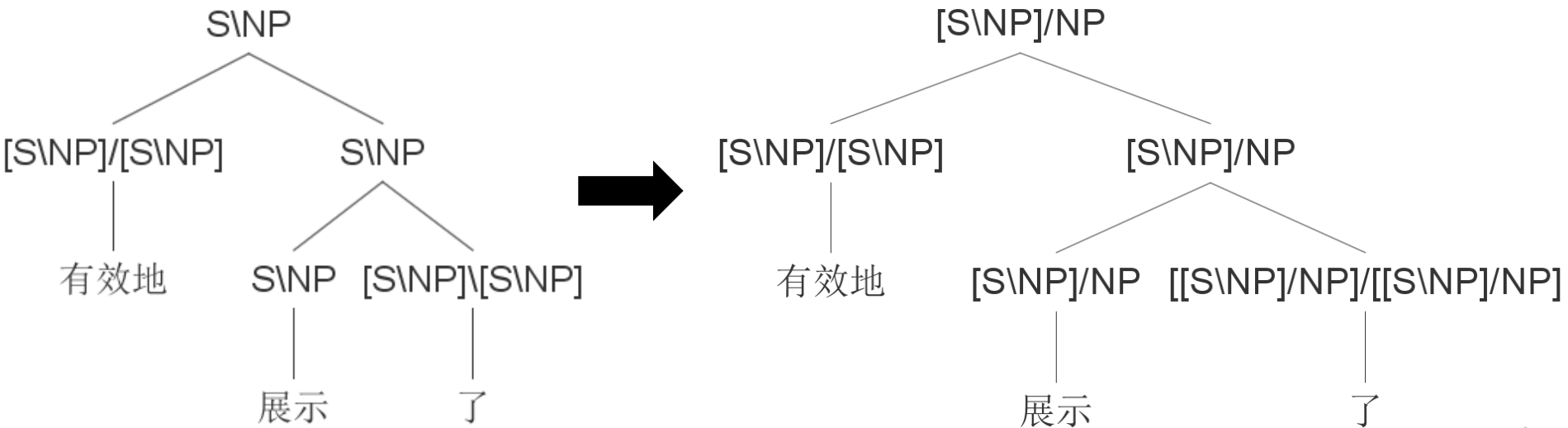
- IPA标注: [S-np-Arg1 成果] [P-vp-Pred(2) 展示]

- 推导树1.0



推导树重构算法(2/5)

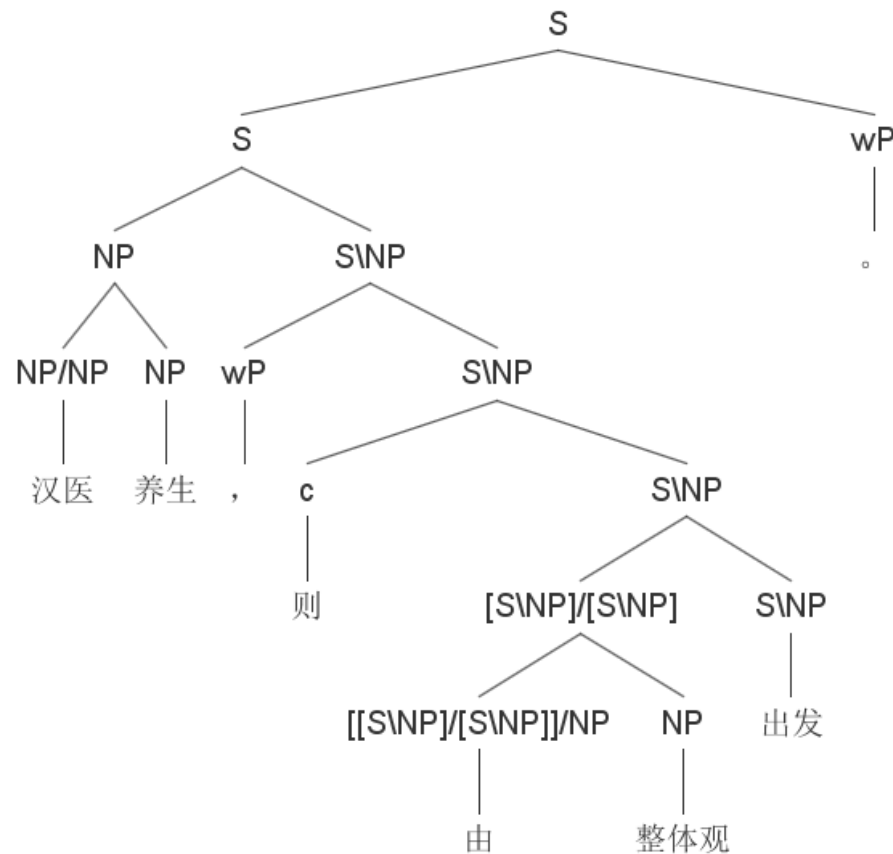
- 子结构重构



推导树重构算法(3/5)

- 子结构序列化

– SA+S+<queue>+D+P+O+<stack>

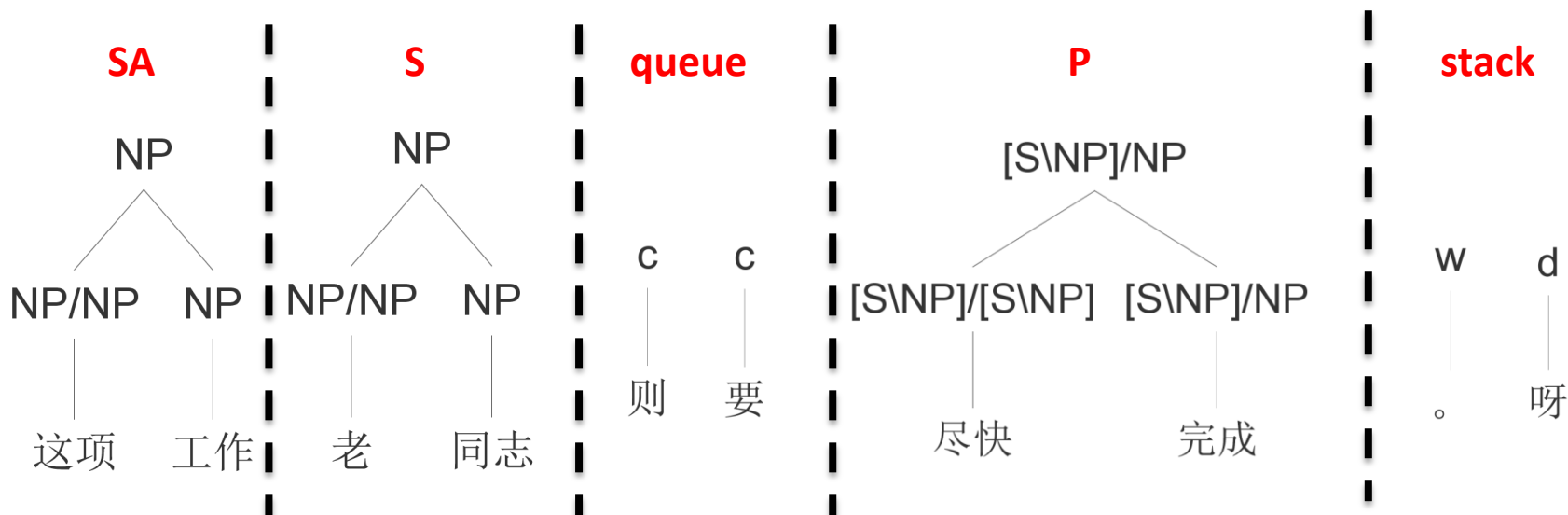


推导树重构算法(4/5)

- 上层结构重构

 - 输入

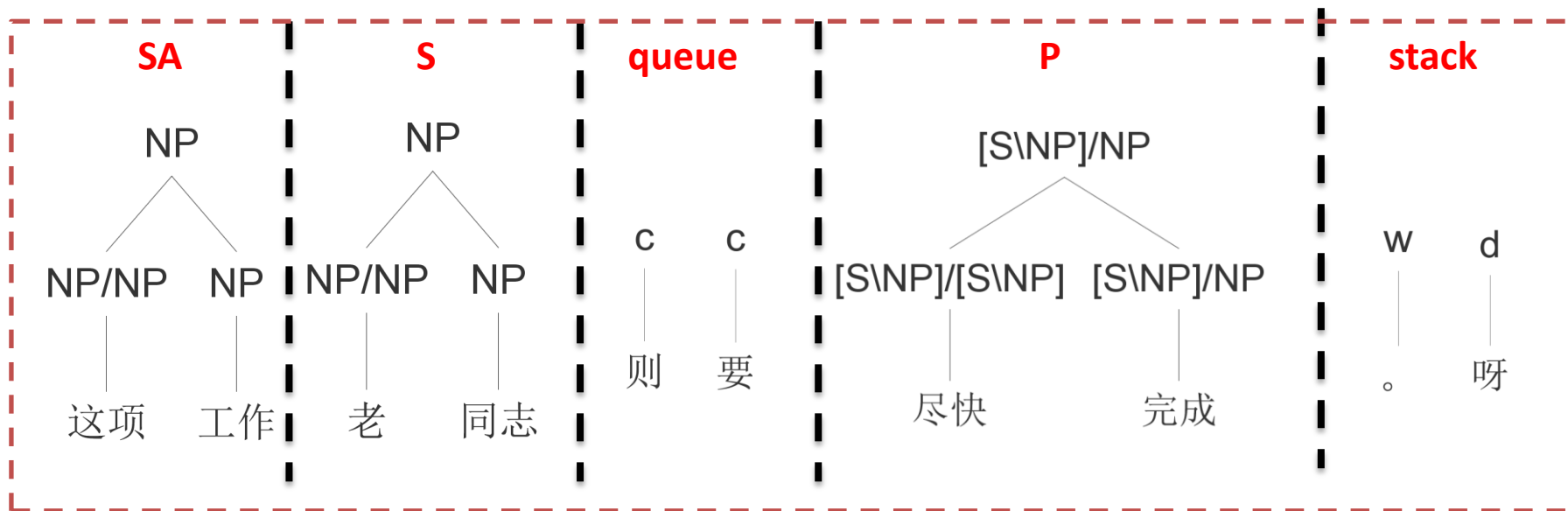
 - 期望重构类型：宾语话题化
 - 子结构序列



推导树重构算法(4/5)

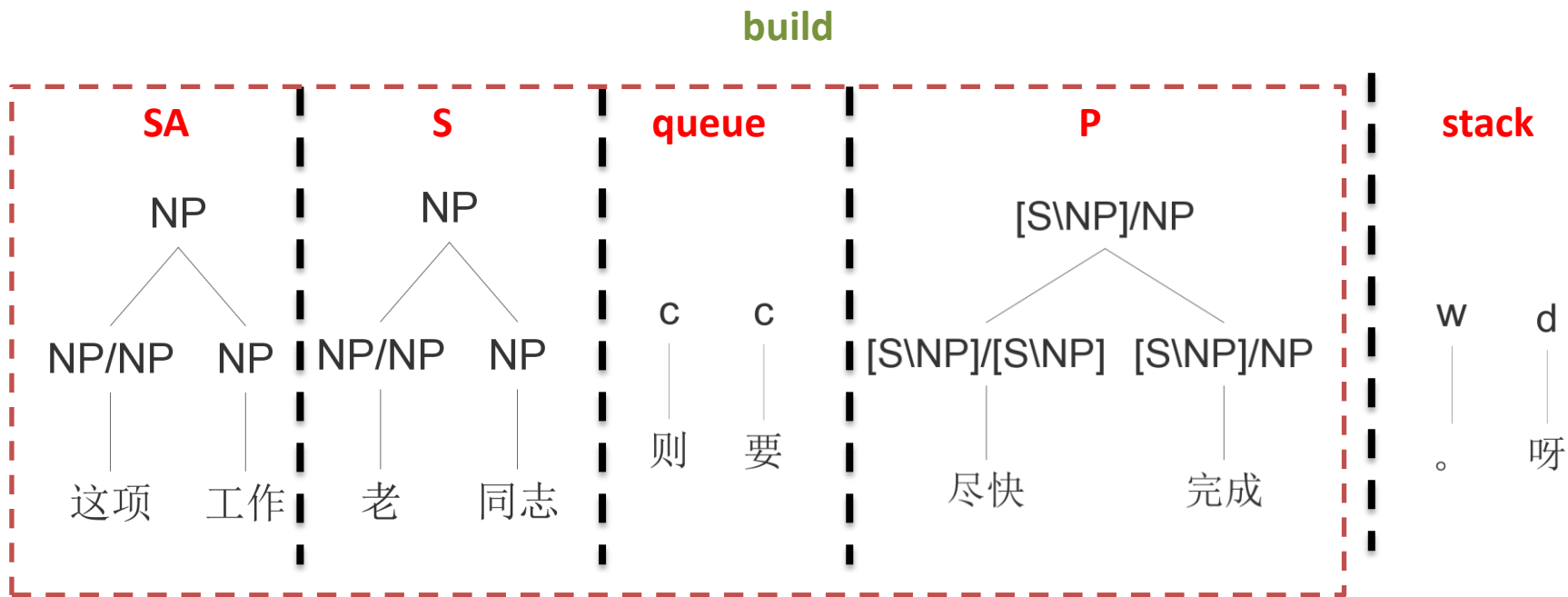
- 上层结构重构
 - 初始状态
 - 宾语话题化

build



推导树重构算法(4/5)

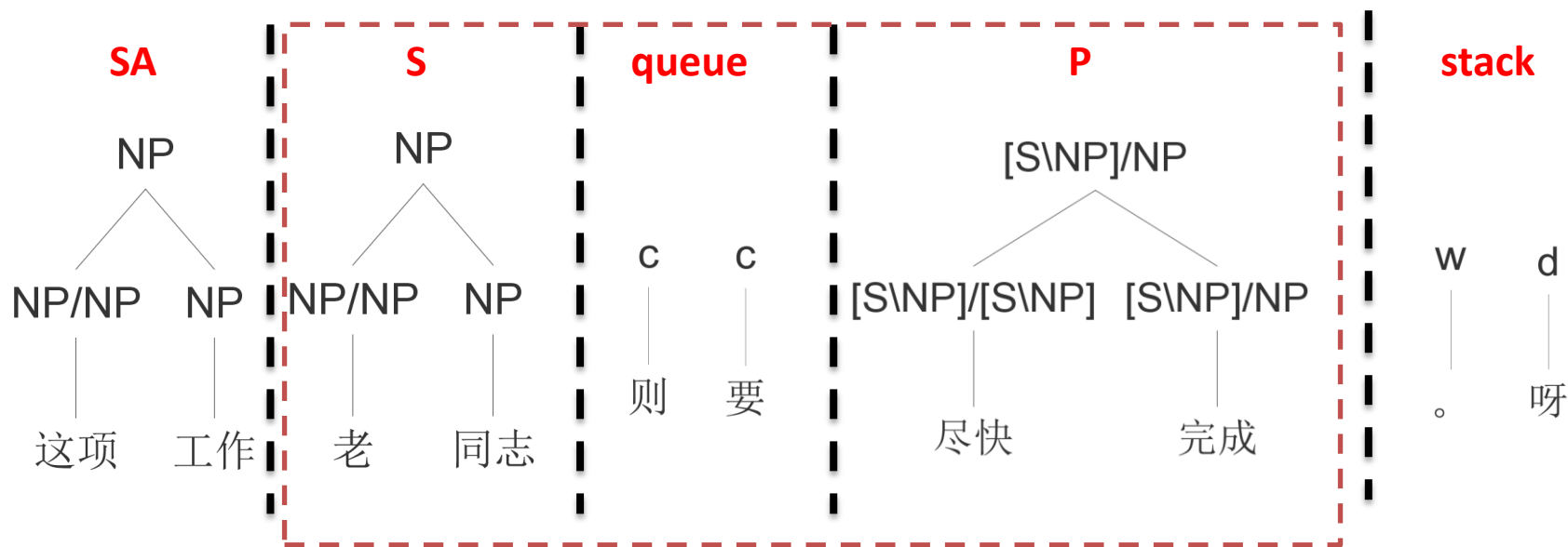
- 上层结构重构
 - 递归
 - 宾语话题化



推导树重构算法(4/5)

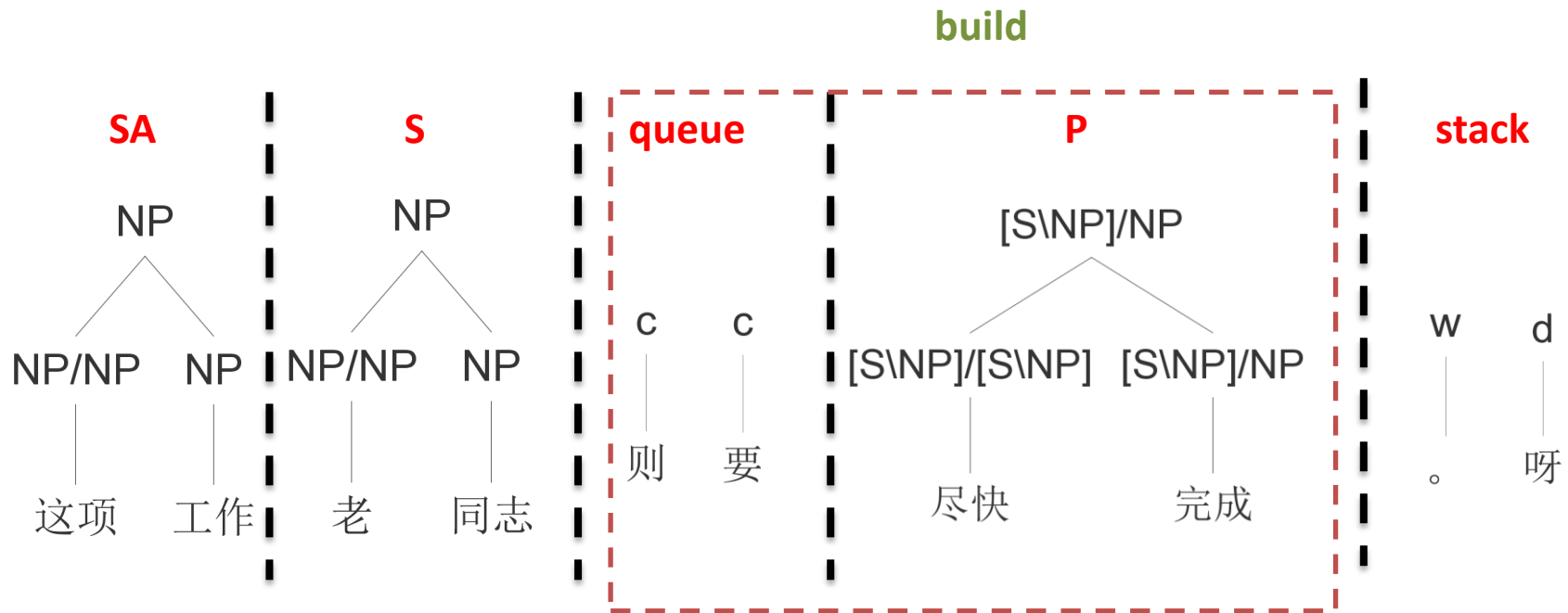
- 上层结构重构
 - 递归
 - 宾语省略

build



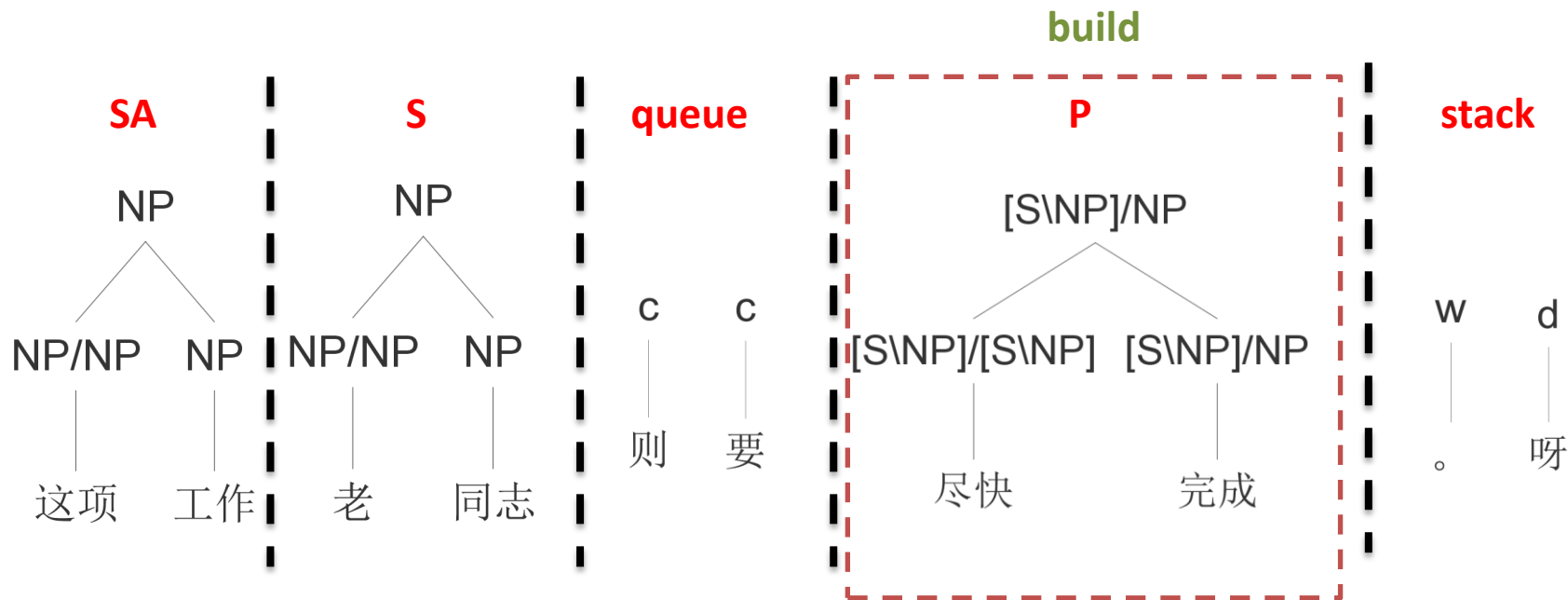
推导树重构算法(4/5)

- 上层结构重构
 - 递归
 - 主宾语省略



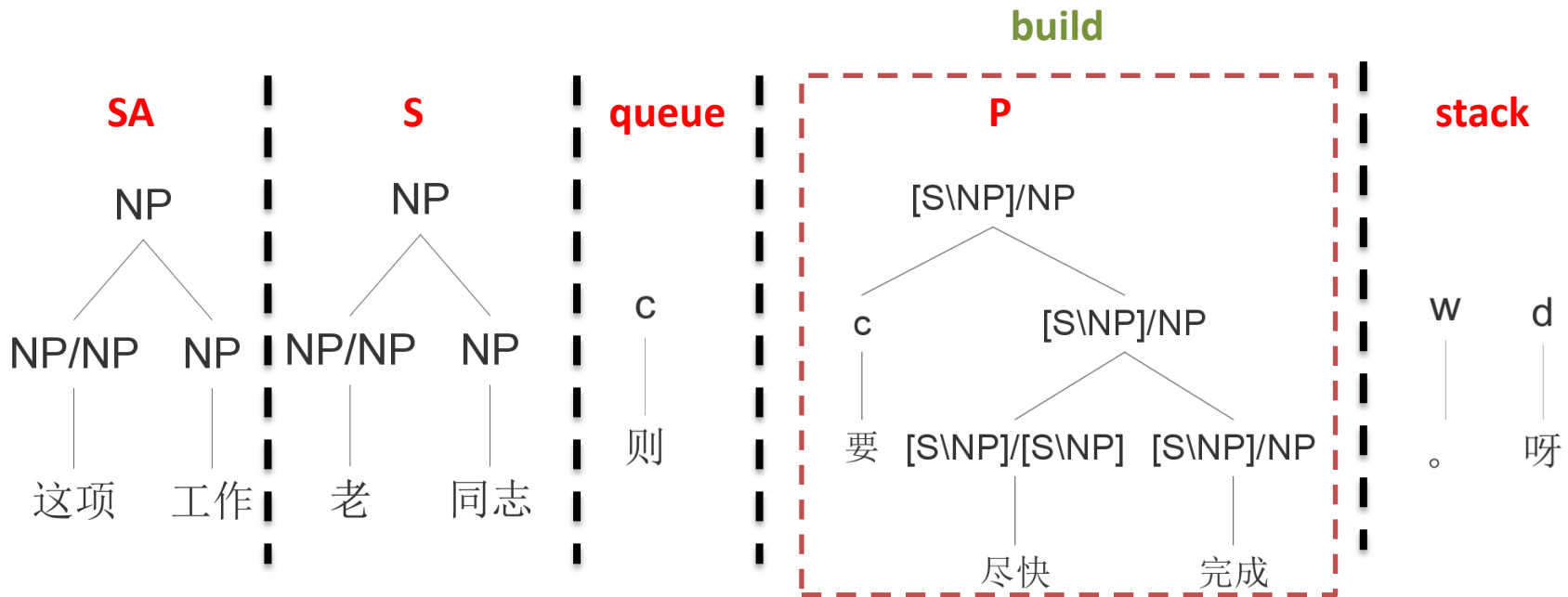
推导树重构算法(4/5)

- 上层结构重构
 - 递归
 - 主宾语省略



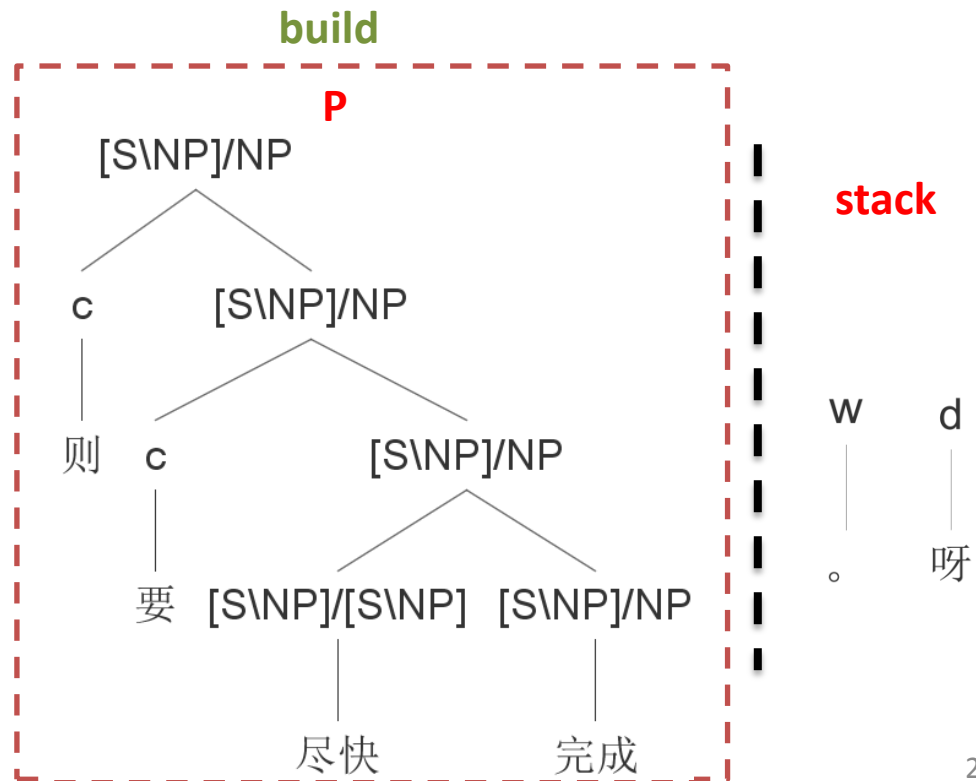
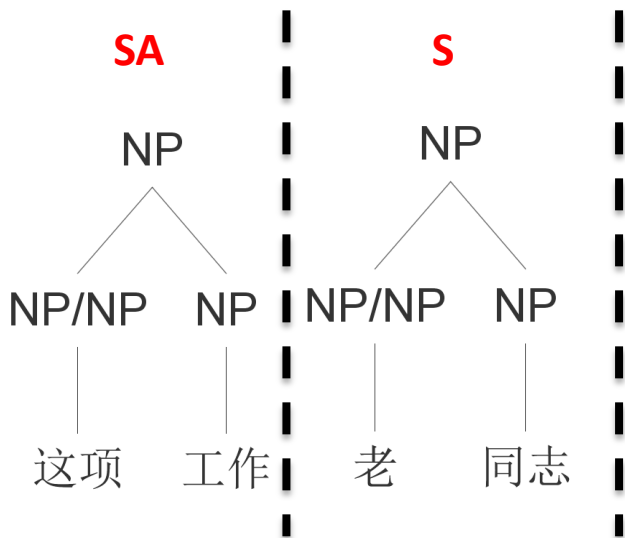
推导树重构算法(4/5)

- 上层结构重构
 - 合并
 - 主宾语省略



推导树重构算法(4/5)

- 上层结构重构
 - 合并
 - 主宾语省略



stack

w
|
。
d
|
呀

推导树重构算法(4/5)

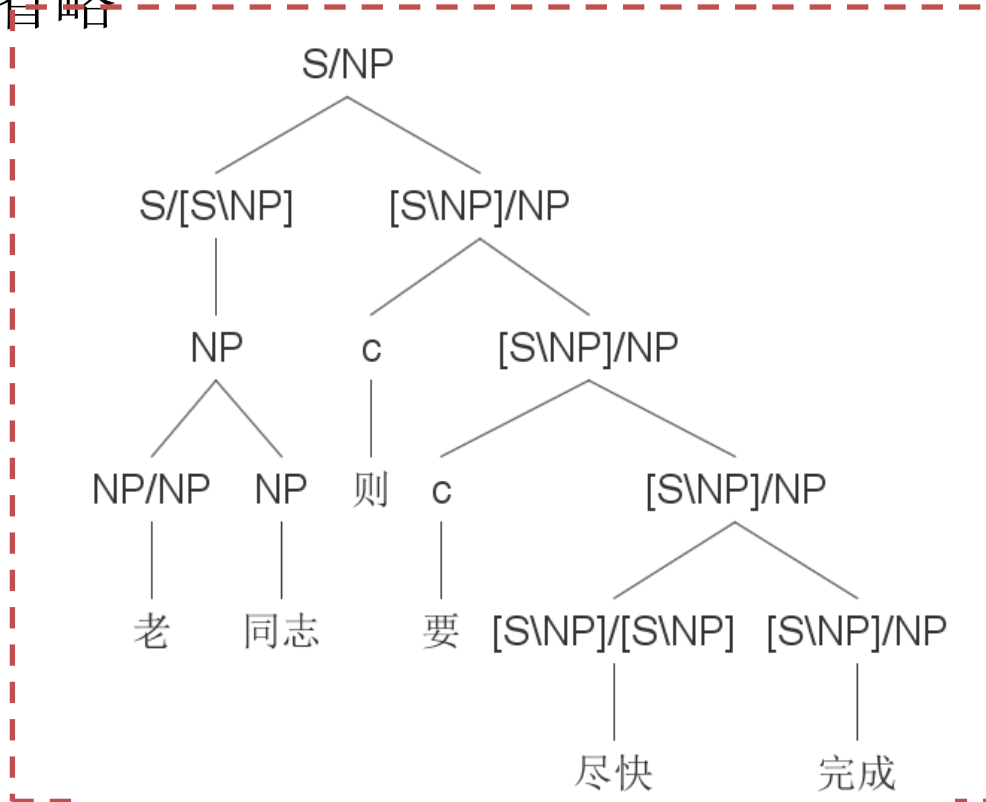
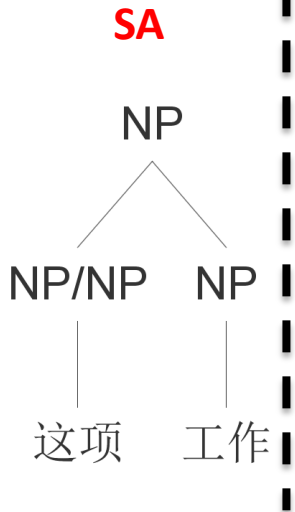
- 上层结构重构

- 合并

- 宾语省略

build

P



stack

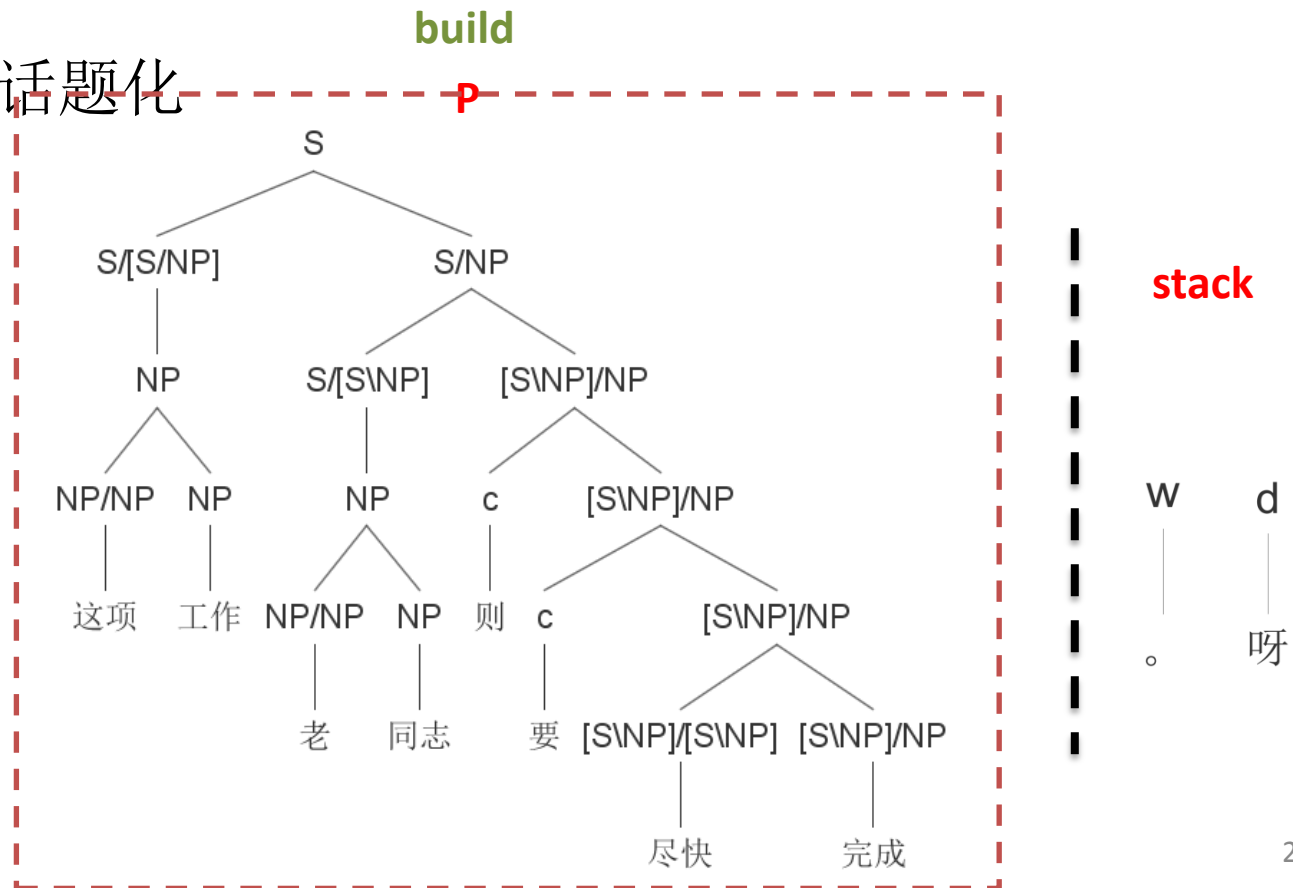
w 。
d 呀

推导树重构算法(4/5)

- 上层结构重构

 - 合并

 - 宾语话题化

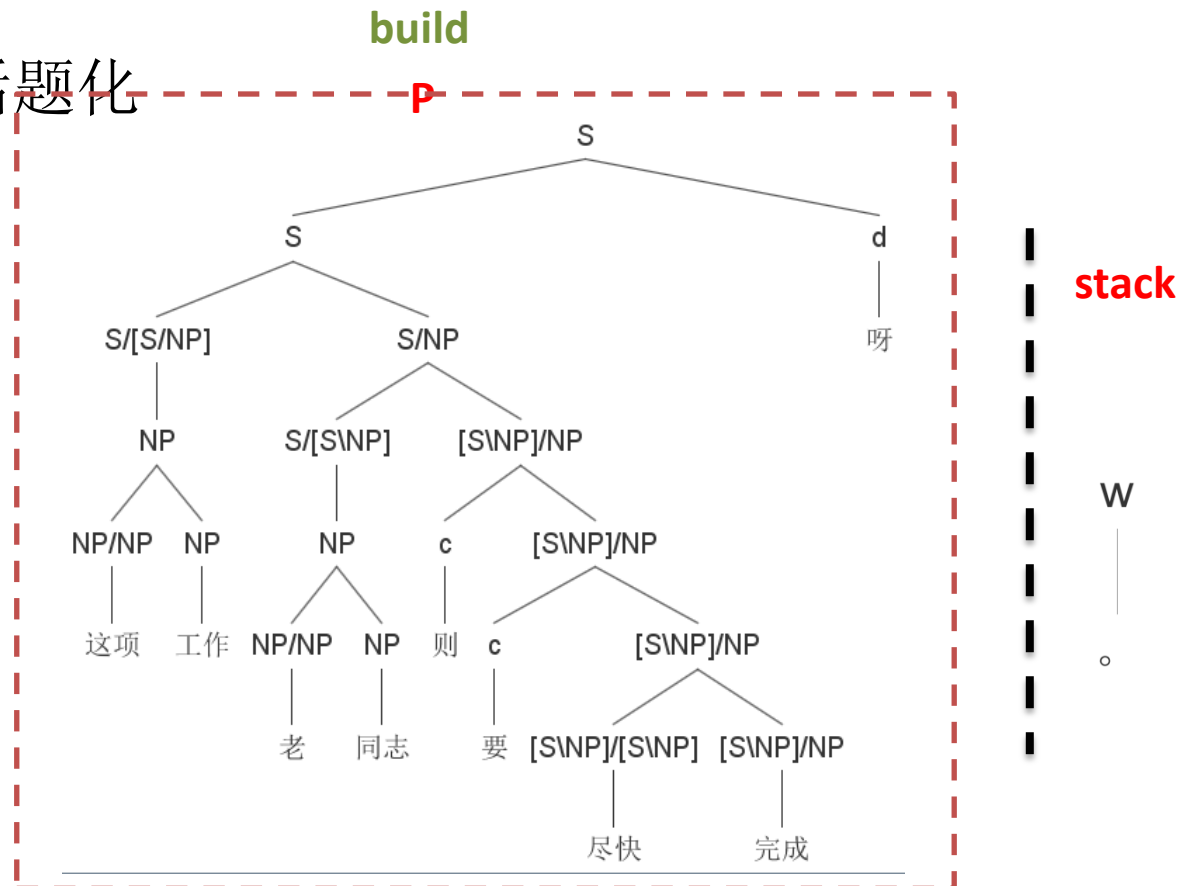


推导树重构算法(4/5)

- 上层结构重构

 - 合并

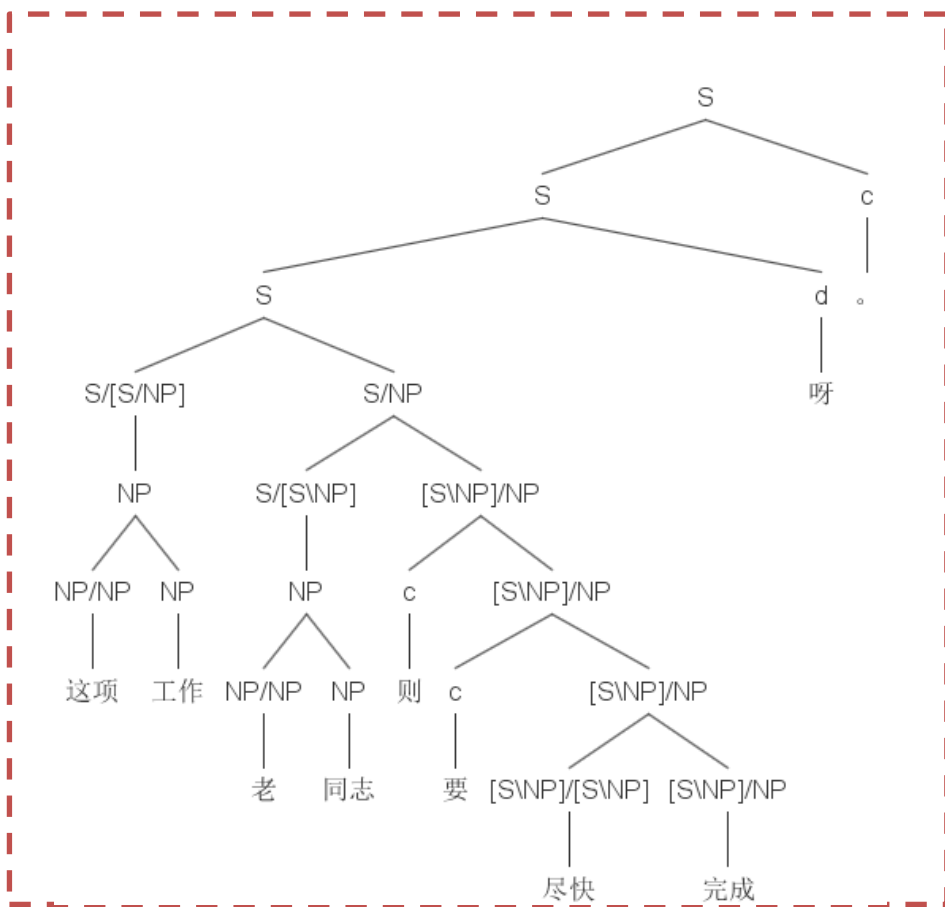
 - 宾语话题化



推导树重构算法(4/5)

- 上层结构重构
 - 合并
 - 宾语话题化

build



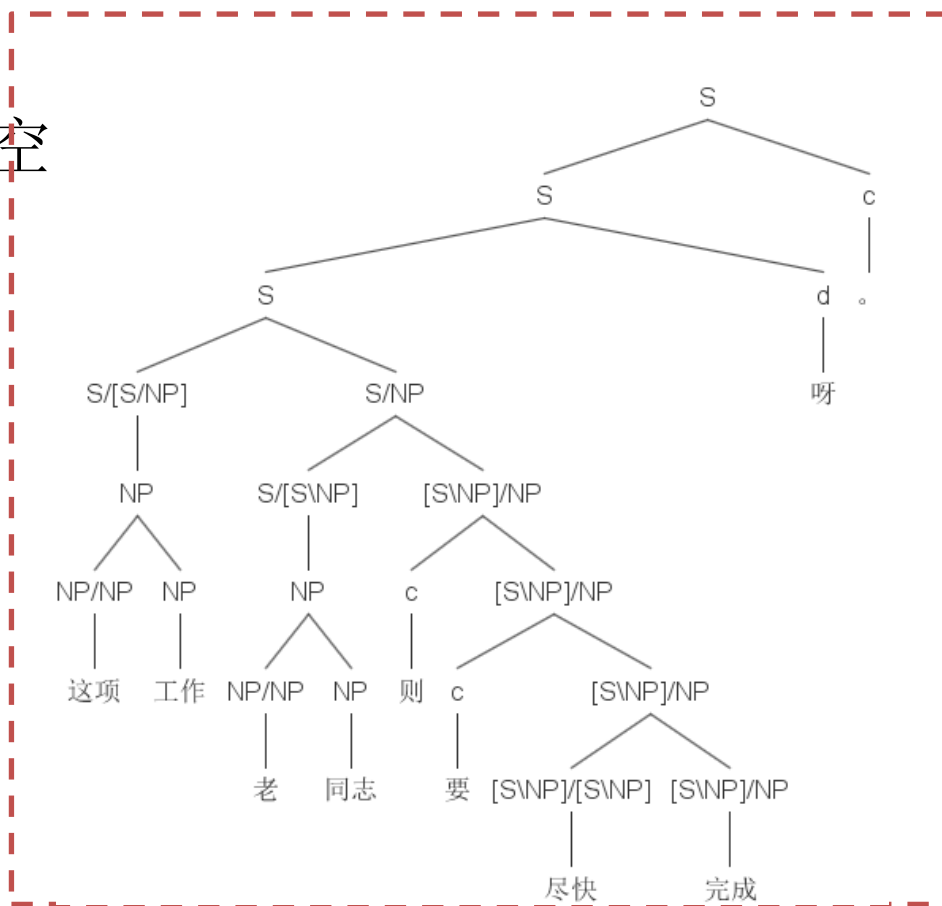
推导树重构算法(4/5)

- 上层结构重构

build

- 终止

- 子结构序列为空



推导树重构算法(5/5)

- 共指信息标识
 - 定语从句，谓词被动化

算法评价(1/2)

- 自动评价
 - 事件句式→IPA信息
 - 重构推导树→词汇依存关系

EC类型	PA搭配一致率
BEC	98.2%
DEC (论元省略)	96.3%
DEC (论元话题化)	96.1%
DEC (论元介词化)	93.6%
TEC (中心语非论元)	99.5%
TEC (中心语为论元)	96.4%
总计	97.3%

算法评价(2/2)

- 人工评价
 - 各类型随机选取20个小句提供推导树的叶子节点序列
 - 人工重构推导树

EC类型	PA搭配一致率
BEC	95%
DEC (论元省略)	100%
DEC (论元话题化)	90%
DEC (论元介词化)	95%
TEC (中心语非论元)	85%
TEC (中心语为论元)	80%
总计	84%

谢谢！