

计算语言学

第10讲 形式语法理论I

刘群

中国科学院计算技术研究所

liuqun@ict.ac.cn

中国科学院研究生院2002~2003学年第二学期课程讲义

内容提要

- 概述
- 形式语法的分类
- 乔姆斯基语法
- 基于简单范畴的语法
- 基于复杂范畴（基于合一）的语法
- 不使用范畴（基于词）的语法
- 工程性语法

形式语法概述 1

- 形式语法理论的目的是试图用精确的数学模型（形式语言）来刻画自然语言
- 与统计语言模型不同的是，对于一个形式语法体系来说，一个句子是否属于一种语言，只存在“是”和“否”两种答案，不存在中间状态（概率）；
- 形式语言虽然在描述人工语言方面取得了很大的成功（如程序设计语言），但在描述自然语言方面还很不成功

形式语法概述 2

- 形式语法体系的目标
 - 准确：能够产生语言中所有的句子，而不产生语言中不存在的句子
 - 简洁：尽可能用比较少的规则来描述一种语言（简单就是美）
 - 高效：分析的计算复杂度越低越好

形式语法概述 3

- 从语言学角度看
 - 传统语法
 - 现代语法
 - 功能语法
 - 形式语法
- 从数学角度看
 - 代数语言学
 - 统计语言学

形式语法要回答的问题 1

- 一个字符串是不是一个合法的句子？

- ✓ 我吃了一只鸡。
- ✓ 一只鸡我吃了。
- ✓ 吃了一只鸡，我。
- ✓ 鸡我吃了一只。
- ✓ 我把一只鸡吃了。
- × 我鸡了一只吃。
- × 我了一只吃鸡。

形式语法要回答的问题 2

- 一个字符串如果不是一个合法的句子，那么是不是一个合法的语言单位？

✓ 我	× 我吃
✓ 吃了	× 吃了一
✓ 一只	× 了一只鸡
✓ 一只鸡	× 了一
✓ 我吃了	× 只鸡
✓ 吃了一只鸡	× 吃了一只

语言语法要回答的问题 3

语言单位和语言单位之间有什么不同？

- 对语言单位进行分类——范畴
 - 句法范畴：名词、名词短语、子句、句子……
 - 语义范畴：人、动物、植物、思想、程度……
- 对于分类的补充描述——特征
 - 句法特征：单复数、时态、语态……
 - 语义特征：性别、尺寸、配价数……

语言语法要回答的问题 4

- 语言单位和语言单位如何组合成更大的语言单位？
 - 语言单位之间的关系
 - 句法关系：述宾、述补、主谓、定中……
 - 语义关系：施事、受事、工具、因果……

句子结构的分析方法 1

- 中心词分析法
 - 句本位的思想
 - 每一个句子都有一个核心的动词
 - 其他所有的成分都是核心动词的附属成分
 - 汉语中典型的句子成分包括
 - 主要成分：主语、谓语、宾语
 - 附属成分：定语、状语、补语

句子结构的分析方法 2

- 层次分析法
 - 词组本位的思想
 - 任何词组都是由更小的词组组成
 - 任何词组都可以在更大的词组中充当成分
 - 每个句子通过每一个层次的成分划分可以构成一棵层次结构树
 - 句子是一种特殊的词组

句子结构的分析方法 3

- 中心词分析法
小李 远远地 看见 小王在草地上看书。
主语 | 状语 | 谓语 | 宾语
- 层次分析法
小李 远远地 看见 小王 在草地上看书。
主 | 谓
 状 | 中
 述 | 宾
 主 | 谓
 状 | 中
 介 | 宾

句子结构的分析方法 4

- 语言中的层次结构是一种客观存在的事实，不管是否使用层次分析法，这种结构都是存在的，层次分析法只是试图去揭示这种结构；
- 层次分析法不排斥寻找一个语言单位的中心成分，反而有利于寻找中心成分；
- 中心词分析法忽略了太多的语言层次信息，有很大的缺陷；

句子结构的分析方法 5

- 汉语的短语结构类型

主谓结构

述宾结构

述补结构

动结式

动趋式

动介式

得字结构

定中结构

状中结构

数量结构

联合结构

同位结构

连谓结构

介词结构

把字结构

被字结构

的字结构

地字结构

时态结构

方位结构

形式语法的分类 1

- 乔姆斯基语法体系
 - 转换生成语法
 - 标准理论
 - 管辖与约束理论
 - 最简方案
- 使用简单范畴的语法体系
 - 索引语法
 - 树粘接语法
 - 定子句语法

形式语法的分类 2

- 使用复杂范畴（特征结构）的语法体系
 - 功能合一语法
 - 词汇功能语法
 - 中心词驱动的短语结构语法
- 不使用范畴的语法体系
 - 依存语法
 - 范畴语法
 - 词汇语法
 - 链语法

形式语法的分类 3

- 工程性语法体系
 - PATRII
 - G语言
 - 分层约束语法

乔姆斯基语法体系 1

- 乔姆斯基语法体系的发展历程
 - 转换生成语法 (1957)
Syntactic Structure. The Hague: Mouton & Co.
 - 标准理论 (1965)
Aspects of the Theory of Syntax, MIT Press
 - 管辖与约束理论 (1981)
Lectures on Government and Binding. Foris, Dordrecht
 - 最简方案 (1992)
A Minimalist Program for Linguistics Theory, MIT
Occasional Papers in Linguistics, #1

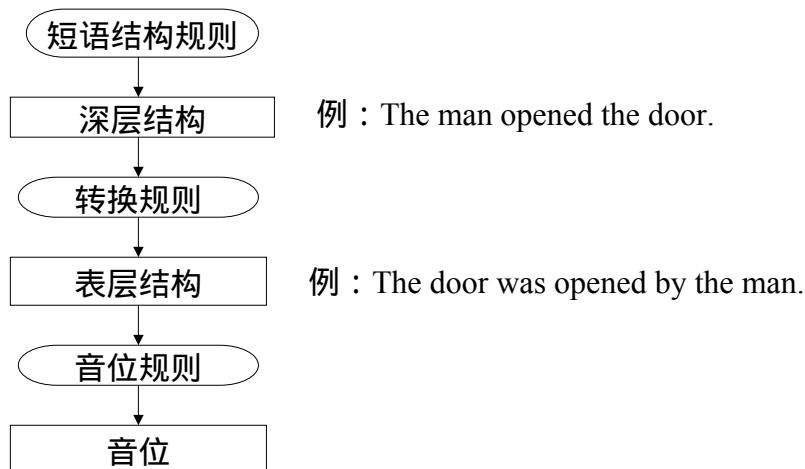
乔姆斯基语法体系 2

- 基本理念——普遍语法思想
 - 人类天生具有某种语言机制：普遍语法
 - 这种语言机制是独立于任何一种自然语言的
 - 任何一种自然语言，都是普遍语法加上一些参数的具体实现
 - 语法研究的目的是发现这种普遍语法
- 儿童语言学习的研究为普遍语法的思想提供了一定程度上的验证
- 乔姆斯基语法体系：在不断变化中发展

乔姆斯基语法体系 3

- 乔姆斯基革命
 - 传统语言学偏向于“语言运用（performance）”的研究，讲求在大规模语言事实调查的基础上对语言进行充分的描述，并总结出语言的规律（所谓“经验主义”）
 - 乔姆斯基强调“语言能力（competence）”的研究，要求语言学家通过精心设计，来诱导说话者的语言直觉，从而揭示出隐藏在语言现象后面的本质性规律（所谓“理性主义”）

转换生成语法 1



转换生成语法 2

- 引入深层结构的作用
 - 可以区分某些表层结构的歧义
表层结构：the shooting of the hunters
深层结构1：Somebody shoot the hunters.
深层结构2：The hunters shoot something.
 - 某些不同的表层结构可以对应于相同的深层结构，使得产生句子的短语结构规则更为简洁
深层结构：The man opened the door.
表层结构1：The man opened the door.
表层结构2：The door is opened by the man.
表层结构3：Did the man open the door?

转换生成语法 3

- 转换规则的形式

- 调位 (movement) : $XY \rightarrow YX$

如英语主动句变为被动句时NP与VP位置的变化

- 复写 (copy) : $X \rightarrow XX$

如英语简单句变为反意疑问句时, 要插入代词

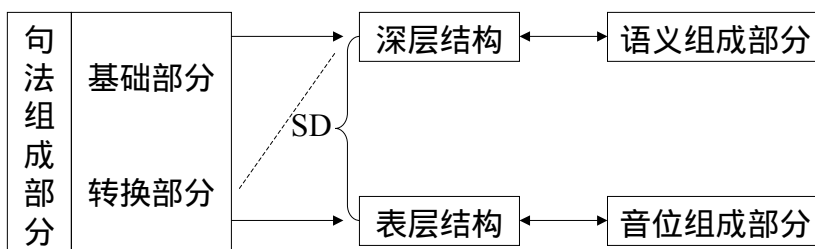
- 插入 (insertion) : $X \rightarrow XY$

如英语主动句变为被动句时, 要插入by

- 消去 (deletion) : $XY \rightarrow X$

如英语陈述句变为祈使句时, 要消去主语

标准理论 1



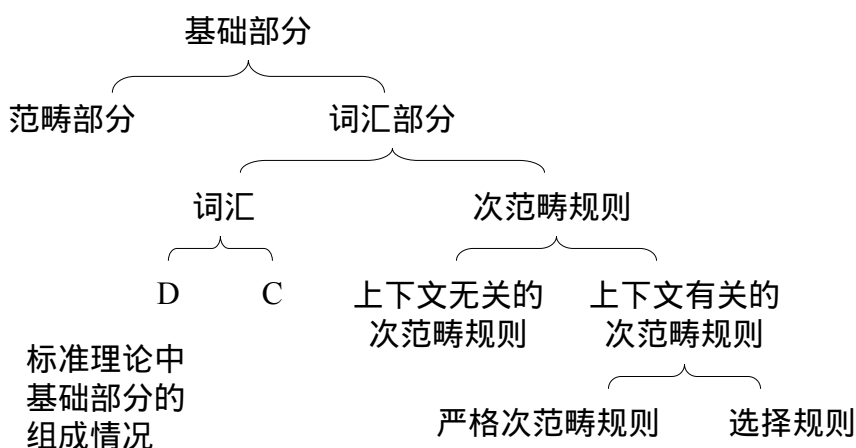
SD : 句法描写 (Syntactic Description)

标准理论 2

- 1971, “深层结构、表层结构和语义解释”, 收录于乔姆斯基“普遍语法中的语义研究”一书, pp.11 ~ 61

标准理论规定, 每一个句子都有句法结构 $\Sigma=(P_1, \dots, P_i, \dots, P_n)$ (其中 P_i 是深层结构, P_n 是表层结构), 语义表达 S , 语音表达 P 。此外, 这种理论认为, S 由 P_i 决定, P 由 P_n 决定, 它们分别受语义解释和语音解释的制约。

标准理论 3



标准理论 4

- 范畴部分：短语结构规则（上）

S → NP Predicate-Phrase

Predicate-Phrase → Aux VP (place) (time)

VP → Copula-Predicate

VP → V (NP) (Prep-Phrase) (Prep-Phrase)

VP → V (Manner)

VP → V S'

VP → V Predicate

标准理论 5

- 范畴部分：短语结构规则（下）

Predicate → Adjective

Predicate → (like) predicate-nominal

Prep-Phrase → Direction, Duration, Place, Frequency

NP → (Det) N (S)

Aux → Tense (M) (Aspect)

Det → (Pre-Article of) Article (Post-Article)

Manner → by Passive

标准理论 6

- 词汇：词项的集合，词项由（C,D）构成
 - D：语音区别特征
 - C：符合符号（Complex Symbols），由一组特定的句法特征和语义特征组成

例：boy

D: /boi/

C: /+N, +Count, +Common, +Animate, +Human/

标准理论 7

- 次范畴规则（Subcategorization Rule）：把范畴部分中出现的语法范畴再划分成小类，如动词划分为及物动词和不及物动词，名词划分为普通名词和专有名词。
- 上下文无关的次范畴规则：用于给名词规定语义特征，使用时不受上下文限制
- 上下文相关的次范畴规则：
 - 严格次范畴规则（Strict Subcategorization Rule）：给动词或名词规定上下文的语境特征
 - 选择规则（Selectional Rule）：给出动词和名词之间的搭配关系

标准理论 8

- 上下文无关的次范畴规则：

[+Det——] → [±Count]

[+Count——] → [±Animate]

[+Animate——] → [±Human]

[-Count] → [±Abstract]

Article → [±Definite]

标准理论 9

- 严格次范畴规则：

eat, [+V, + — NP]

elapse, [+V, + — #]

grow, [+V, + — NP, + — #, + — Adjective]

become, [+V, + — Adjective, + — Predicate-Nominal]

seem, [+V, + — Adjective, + — like Predicate-Nominal]

look, [+V, + — (Prepositional-Phrase) #, + — Adjective,
+ — like Predicate-Nominal]

believe, [+V, + — NP, + — that S']

persuade, [+V, + — NP (of Det N) S']

标准理论 10

- 选择规则：

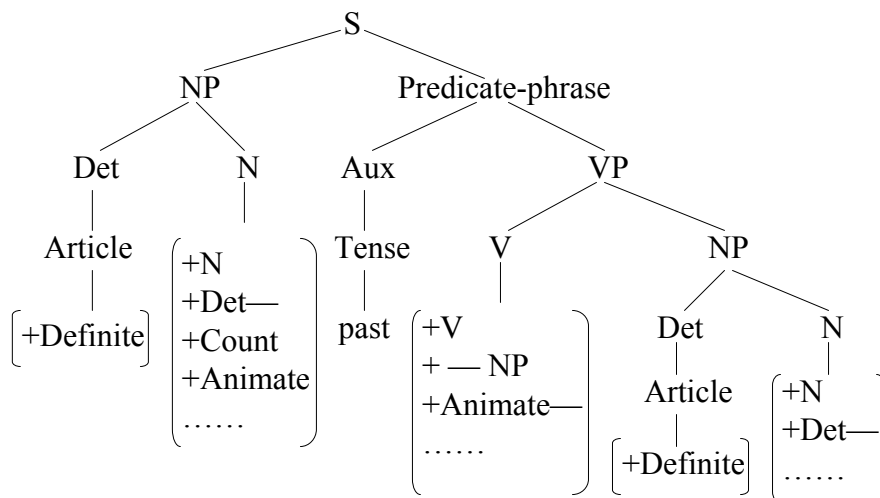
frighten → CS / $\begin{cases} [+Abstract] \text{ Aux } — \\ — \text{ Det } [+Animate] \end{cases}$

✓ Sincerity may frighten the boy.

(诚实会吓坏孩子)

✗ The boy may frighten sincerity.

标准理论 11



标准理论 12

- 触发成分 (triggering element)

- Q (疑问)
- Neg (否定)
- Imp (命令)
- Passive (被动)

例：**Imp** you write it → write it

- 早期的转换生成语法将转换操作分为强制转换和随意转换。触发成分的引入，取消了随意转换，使得转换操作更有规律

扩充式标准理论

- 组成部分

- 基础部分：(P_1, \dots, P_i) P_i 为深层结构
 - 转换部分：(P_i, \dots, P_n) P_n 为表层结构
 - 语音部分： $P_n \rightarrow$ 语音表现
 - 语义部分：(P_i, P_n) \rightarrow 语义表现
- 在扩充式标准理论中，表层结构也与语义有关。

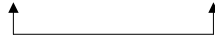
- 例：

- Not many arrows hits the target.
- Many arrows didn't hit the target.

这两个句子虽然深层结构相同，但语义显然不同。

语迹理论 1

- 所谓语迹 (trace)，就是在转换过程中，当把短语P从X位置转移到Y位置后，P在X位置留下的痕迹，记为 t_i 。
- 例子：（相同的下标表示同指）
 - Beavers built dams.
 - Dams_i are built t_i by beavers.



这里相同的下标表示两个词语指称相同的对象，即二者具有同指关系 (coreference)。

语迹理论 2

- 语迹的存在可以在语音上得到验证
 - 英语中want to经常可以连读为wanna，但是：
 - × Who do you wanna win the election?
 - ✓ Why do you wanna win the election?
 - 上面的第一句want to却不能连读，因为其中有一个Wh语迹存在：
 - Who_i do you want t_i to win the election?
- 语迹的提出可以使表层结构直接与语义解释相联系，不必再通过深层结构

管辖与约束理论 1

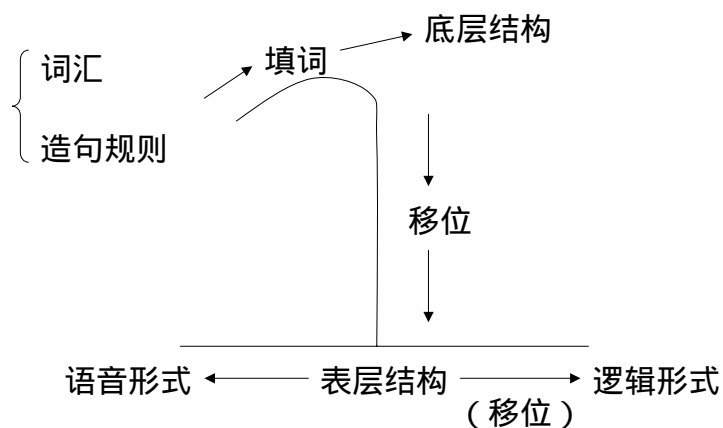
- “原则 + 参数”思想：

我们希望研究出一套结构高度严谨的普遍语法理论，它以一系列基本原则为基础，这些原则明确划定语法的可能的范围，并严格地限制其形式，但也应该含有一些参数，这些参数只能由经验决定。

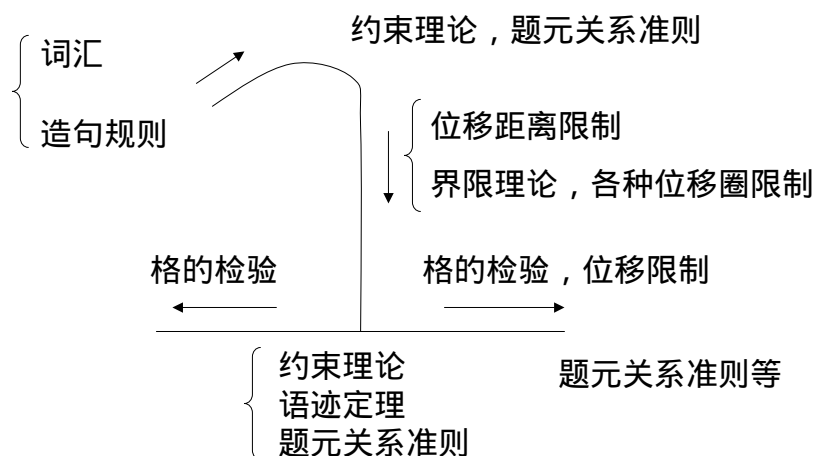
管辖与约束理论 2

- 标杆理论（ \bar{X} -Theory）
- 题元理论（ θ -Theory）
- 格理论（Case Theory）
- 管辖理论（Government Theory）
- 约束理论（Binding Theory）
- 控制理论（Control Theory）
- 界限理论（Bounding Theory）

管辖与约束理论 3



管辖与约束理论 4

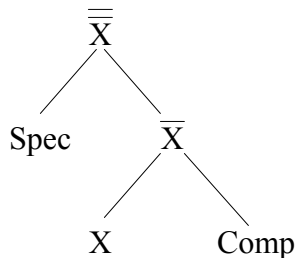


管辖与约束理论 5

- 在管辖与约束理论中，不再有专门的位移规则，所有的位移统一称为“移动 α (Move α)”
- 对于移动 α 的限制通过各种原则来体现，只要满足各种原则的移动都是合法的。

标杆理论 1

- 标杆理论： \bar{X} 理论，X-bar理论，中节理论，X标杠理论，.....



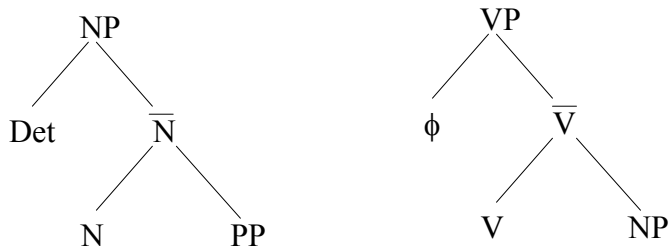
中心语： X 、 \bar{X} 、 $\bar{\bar{X}}$

补足语：Complement

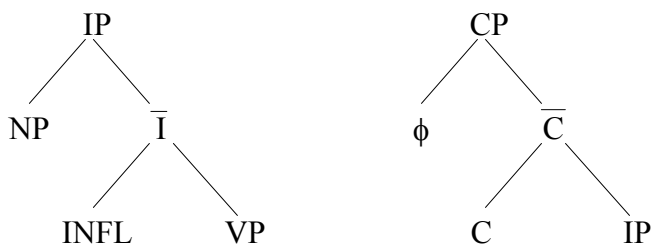
标志语：Specifier

投射： $X \rightarrow \bar{X}$ ， $\bar{X} \rightarrow \bar{\bar{X}}$

标杆理论 2



标杆理论 3



IP=S: Sentence 句子
INFL: inflection 形态变化

C: Complementizer 补语连词
CP: Complementizer Phrase 子句

标杆理论 4

1. 一切词组都是向心 (endocentric) 结构，都有一个中心词；
2. Spec和X-bar处于同一层，是XP的组成成分；Comp和X处于同一层，是X-bar的组成成分。也就是说，X与补足语的关系比X与标志语的关系更为密切；
3. 在D结构中，同一种语言中，Spec和X-bar的位置关系、Comp和X的位置关系都是固定的，不管X是什么语类；
4. 在D结构中，X上面的横线逐级递增，也就是说 X^n 的下一级必须是 X^{n-1} ，不能是 X^n ；
5. 除了中心语以外的结点上都只能出现某一语类的最大投射。

格理论

- 格理论属于句法结构的范围。它表示句中各成分之间的关系，以及这种关系实现的形式手段。确定在怎样的结构条件下和在句子生成的那个阶段，一些词（如动词、介词）可以赋格，那些词必须取得格。
- 这里的格既包括可见格，如俄语中的六个格，也包括隐性格，如汉语和英语中的格。因此，格理论研究的是一种抽象格。
- 格理论对词组位移后留下的语迹是否能取得格也做了相应的规定。

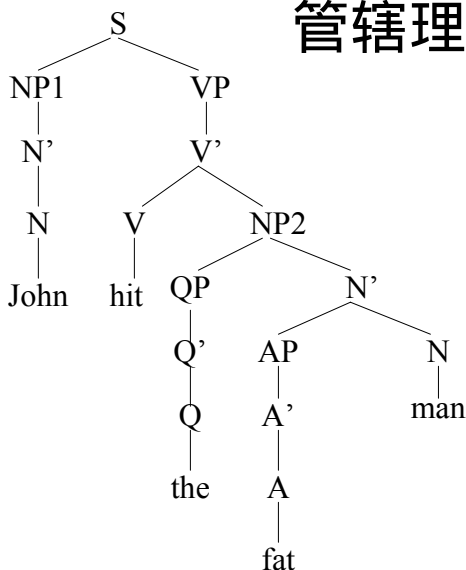
管辖理论 1

- 管辖理论试图从结构上来定义句子中哪些成分可以管辖从而可以赋格给另外一些成分。
- 可以赋格的成分称之为主管成分（如动词、介词等），受其管辖的成分称之为受管成分（如名词、形容词等），而这种关系称为管辖关系。
- 管辖理论研究成为主管、受管成分的结构条件；以及代词、空语类以及它们的先行词之间的管辖关系。

管辖理论 2

- 支配（dominate）：在句法树中，如果结点X是结点Y的祖先结点，则称结点X支配结点Y；
- 统制（c-command）：如果有两个结点X和Y，X不支配Y，但第一个支配X的词组结点（即最大投射）也支配Y，则称结点X统制结点Y；
- 管辖（government）：结点X和结点Y互相统制，且结点X为最底层的中心结点，在结点X管辖结点Y。

管辖理论 3



NP1统制VP及其子孙结点
V统制NP2及其子孙结点
NP2统制V，但不统制NP1

V管辖NP2
N man管辖QP和AP

题元理论 1

- 题元理论研究各项之间的题元关系，即传统语法中用施事、受事、工具等术语表示的概念范畴之间的关系。题元理论以动词为中心，认为每一个题元角色都是由动词分配的。
- 题元准则：
 - 每个论元（argument）必须且只许充当一个题元角色（ θ -role）；
 - 每个题元角色必须且只许由一个论元充当。

题元理论 2

- kill有“施事”和“受事”两个题元，run只有一个“施事”题元。不同的动词有不同数目的题元，这是动词的词汇特征，由词库规定。
- They go John.是不合法的句子，因为John处于“受事”题元角色的位置，而动词go却没有分配“受事”题元角色。
- 汉语动词“开了”的主语位置可以是施事，也可以是受事，而宾语位置只能是受事。而“推”的主语位置只能是施事。

约束理论 1

- 约束理论研究的是名词词组之间的指称关系，在怎样的结构范围内（辖域），代词（如你、我、他）、反身代词（如自己）、指称语（如这人、那头牛）、空语类等可以受到先行词约束。
- 照应语必须在先行语的统制范围之内。
- 管辖语域（Governing Category）：
当且仅当，结点X是S或者NP，结点Y是NP，且结点X是第一个包含结点Y的最大投射，那么称结点X是结点Y的管辖语域。

约束理论 2

三条约束原则，

- a) 照应词在管辖区域内受约束，即它的先行词必定在管辖语域内。
- b) 代名词在管辖语域内是自由的。
- c) 指称词总是自由的。

照应词：包括反身代词myself, yourself, 相互代词each other等。

代名词：指传统语法中的人称代词，如he, him等。

指称词：就是通常所说的名词短语，如the man, john等。

约束理论 3

- 例：小王说小李喜欢他。
 - “他”：有可以是“小王”，不能是“小李”
 - 代名词“他”在“小李”的管辖语域内，因此“他”不能受“小李”约束
- 例：John wants Tom to criticize himself.
 - “himself”：只可以是“Tom”，不可以是“John”
 - Tom处在himself的管辖语域内，John不是。
- 例：小王对小张说小李喜欢自己
 - “自己”：可以是“小王”、“小李”，不能是“小张”
 - “自己”受“小王”和“小李”的统制，但不受“小张”的统制。
 - 汉语的“自己”不同于英语的反身代词，不满足上述的约束原则a)

控制理论

- 控制理论研究的对象是PRO，研究PRO怎样受先行词的控制
- PRO是一个空语类，它有一个先行词，同时又具有代词的性质
- 如：老王叫儿子〔PRO去买点酒来〕。这里PRO和“儿子”同指，而不是和“老王”同指。

界限理论

- 界限理论研究在一些位移情况下，如何用界限结点（如S结点、NP结点）来对位移转换加以限制，使得语迹与移走成分之间的间隔不超过一定的界限。

复习思考题

- 试用层次分析法分析下面的短语或句子：
上街买东西 总理周恩来 走过来 洗干净
走下楼 大眼睛姑娘 高兴得跳起来 洗衣服
那天武松在客栈一口气喝了十八碗酒。
- 找出下面句子中位移产生的语迹：
我买的书 小王打算出国 咬死了猎人的狗
- 根据标杆理论画出下面句子的句法结构树：
John believe himself to be the king of samoa
- 查阅文献，了解标杆理论和约束理论在汉语中的应用，写一篇综述。