



第五讲 语篇知识及其应用

詹卫东

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



提纲

1. 从句子到篇章
2. CT (Centering Theory)
3. DRT (Discourse Representation Theory)
SDRT (Segmented Discourse Representation Theory)
4. CDT (Concept Dependency Theory)
5. RST (Rhetorical Structure Theory)
6. 小结



1 从句子到篇章

- 篇章的连贯性以及连贯的篇章中各组成部分之间的语义关系
- 篇章中指称成分的所指对象



1.1 篇章连贯性

1. 张三送了一套茶具给李四。李四很喜欢。
2. * 李四经常迟到。张三送了一套茶具给王五。
3. * 张三送了一套茶具给李四。李四很喜欢一套茶具。
4. ? 张三送了一套茶具给李四。李四很喜欢茶具。

连贯的句子 = 合法的篇章

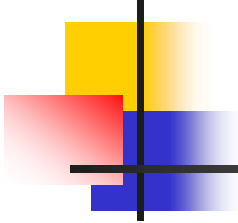
不连贯的句子 = 不合法的篇章

Qs: 是什么在制约句子之间的连贯性 (coherence) ?



篇章连贯性的实现方式

回指	阿Q以前就认识 吴妈 ，那时 她 还没有嫁人。
省略	阿Q给吴妈买了个 戒指 ，她很喜欢 ____。
关联成分	阿Q 起初 是喜欢吴妈的。 后来 进城打工就渐渐淡忘了。
结构平行关联	有的人 活着，他已经死了； 有的人 死了，他还活着；
语义联想关联	吃一口 黄连 吃一口 糖 ，王贵娶了李香香



篇章连贯性：语义关联

1a. 张三明天坐飞机到北京。

1b. 李四会去机场见他。

1a'. 张三坐飞机到北京。

1b'. 李四去机场见了了他。

2a. 张三明天坐飞机到北京。

2b. 李四会去操场见他。

2a'. 张三坐飞机到北京。

2b'. 李四去操场见了了他。

3a. 张三看上去好像没有女朋友。

3b. 但他最近老往健身房跑。

4a. 张三看上去好像没有女朋友。

4b. 但他最近老去食堂吃饭。



篇章连贯性：事件的时间关系

1a. 张三摔倒了。

1b. 是李四把他扶起来的。

1a 先于 1b

2a. 张三摔倒了。

2b. 是李四推倒他的。

2b 先于 2a

3a. 张三学习很好。

3b. 他每门课都是优。

3a ? 3b



篇章连贯性：汉英表达对比的例子

- 1 One has to wonder whether proper controls have been exercised to stop favoured factions in the Vatican from simply pushing people forward unhindered.

人们不由得想知道，是否已运用适当的控制手段去制止梵蒂冈偏袒的某些小派系毫无阻碍地把人推举上去。

- 2 In order to determine its breaking point, the simulated bridge deck has been tested for durability and strength by the expert group.

为了确定这个模拟桥面的断裂点，专家小组测试了它的耐力与强度。



从局部连贯到全局连贯

- 局部连贯 —— 语义知识/语篇知识（交界地带）
小句 —— 复句 —— 句群
- 全局连贯 —— 语篇知识
复句 —— 句群 —— 语篇



全局连贯

(论证体的篇章结构——引自廖秋忠1988)

《“索嘎”并未用错》

S₁

一月九日晚报刊登了《“索嘎”用错了》一文。其实，在《麦子王传奇》中，“索嘎”并未用错。

S₂

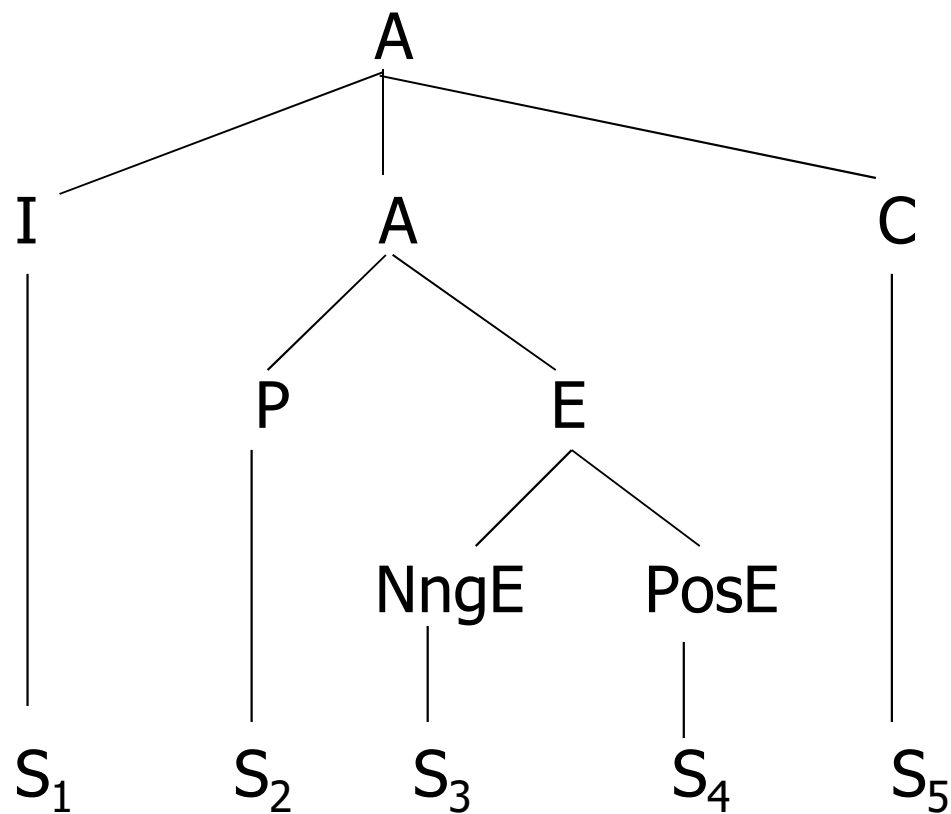
诚如该文作者王为同志所说，“索嘎”有“是吗？”或“是那样吗？”的含义，但这只是疑问口气时的含义。至于以自言自语的口气，亦即以“降调”说出这句话时，则是“原来如此啊”的意思，表示说话者已搞清了事实的真相。因而，在竹下发怒而打王庆财时，口说“索嘎”是恰如其分的。（《北京晚报》1987年1月21日3版）

S₃

S₄

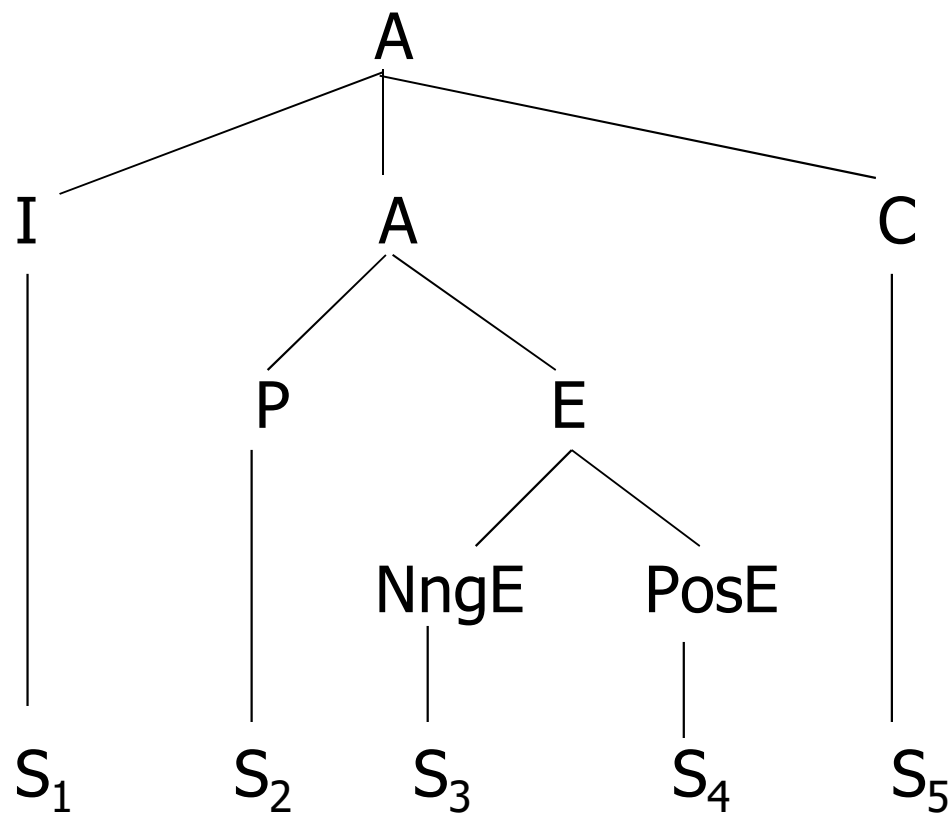
S₅

汉语论证体的篇章结构树形图



A 论证结构
I 引言
C 结尾（结论）
P 论点
E 论据
NngE 反面论据
PosE 正面论据

汉语论证体的篇章结构树形图



A -> I	A	C
A -> P	E	
E -> NngE	PosE	



1.2 篇章中指称成分的所指对象

- 篇章中有哪些指称成分？
- 指称成分有哪些指称（指代）类型（现象）？
- 指称的规则是什么？

胡建华、潘海华（2002）NP显著性的计算与汉语反身代词“自己”的指称，《当代语言学》2002年第1期。

胡建华、石定栩（2006）约束B原则与代词的句内指称，《中国语文》2006年第1期。

沈阳、董红源（2004）“直接统制”与“他”的句内所指规则，《中国语文》2004年第1期。

徐赳赳（2003）《现代汉语篇章回指研究》中国社会科学出版社2003年版。

徐烈炯（1988）《生成语法理论》6.5节，上海外语教育出版社1988年版。

篇章中指称成分的所指对象

1. 那个小偷说他**他**没有偷过东西。

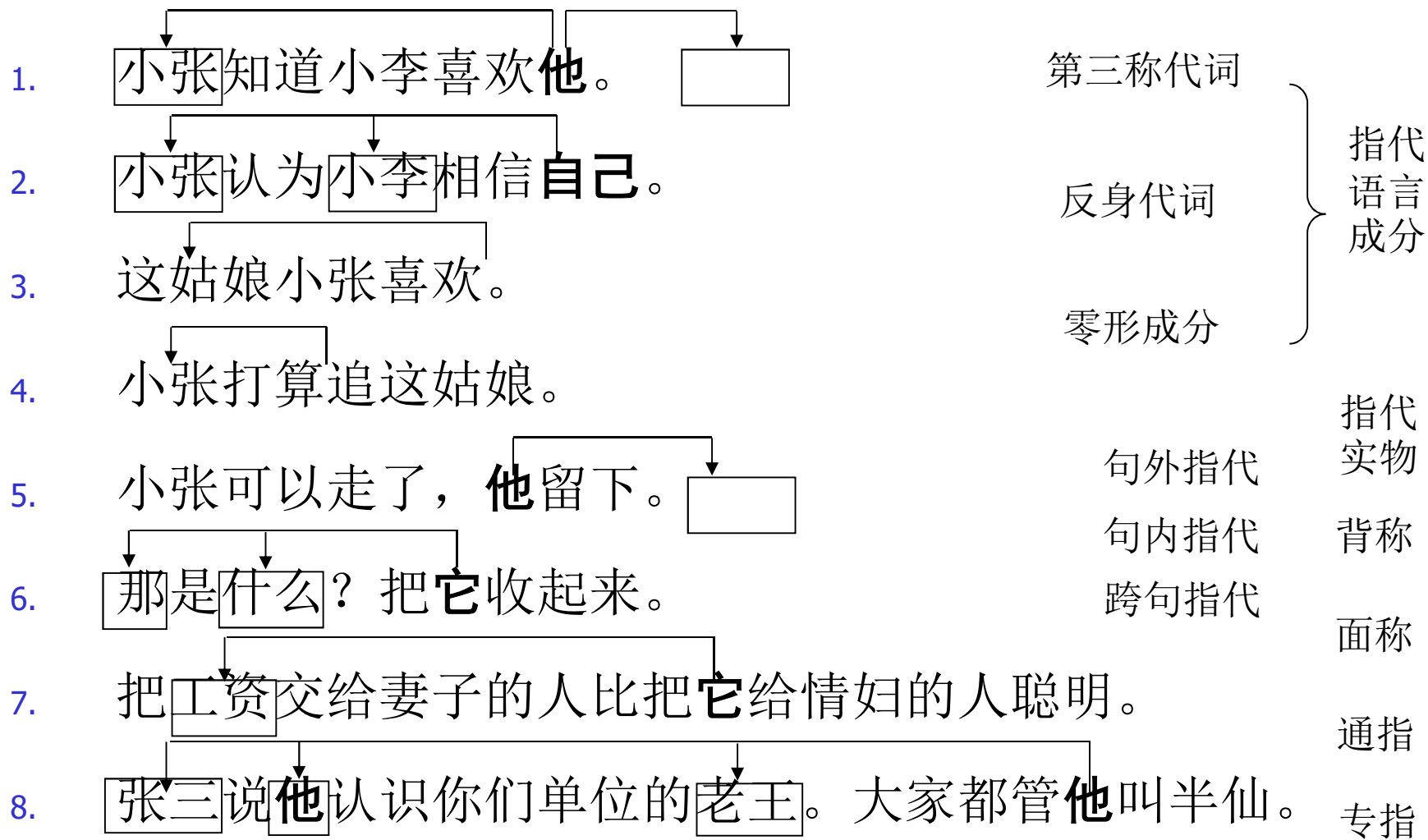
2. 那个小偷说**自己**没有偷过东西。

3. 如果**他**在这儿，**李四**一定会这么做。

4. 想讨好**她**父亲的人争先恐后地向**小芳**献殷勤。

5. **这些公司**的所有权已经通过**各种各样的方式**发生了变化，**其中**包括破产、上市、拍卖和出售给海内外的商人。

6. 以哥伦比亚大学为例，**该校55%的MBA毕业生**受雇于**华尔街的公司**，**其中**半数以上被派驻海外。





三种成分的指代

	句内/跨句	指实物/指语言成分	被指成分是通指/专指
第三人称代词	均可	均可	均可
零形成分	均可	指语言成分	均可
反身代词	均可	指语言成分	均可

1. 王老师说能批评**自己**是一种美德
2. 王老师说**自己**做错了事情
3. 王老师说**自己**做错了事情**自己**要承担责任



指代的约束原则 (Binding Principle)

指称词 (referential expression)	John, 张三, The man, 那个人, 这些书, 这些书的封皮,
代名词 (pronominal)	我, 你, 他, 他们,
照应词 (anaphor)	自己, 他自己,

约束原则:

- A: 照应词必须在**管辖语域**内受约束
- B: 代名词不得在**管辖语域**内受约束
- C: 指称词是自由的

必须指什么 不能指什么

管辖语域 (governing category)

x 是 名词性成分 y 的 管辖语域，当且仅当：

- (1) x 的子孙节点中包含 y，以及**管辖**y的节点；
- (2) x 是距离 y 最近的那个 **IP(S)** 或 **NP**；
- (3) x中应包含**统制** y 的节点，且该节点充当与y 相关的谓词的逻辑主语

符号这些条件的x语类 (category) 的是一个“最小的有完整命题义的结构体”

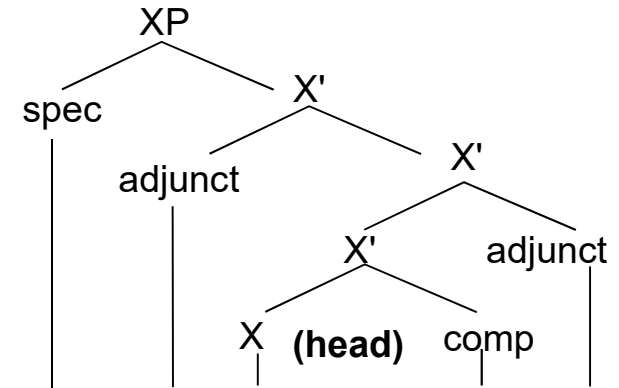
那么，什么是“**管辖**” (government)，什么又是“**统制**” (command) 呢？

管辖与统制

- 管辖（government）是一种句法结构关系，指一个结构中主管成分（governor）与被管成分（governee）之间的关系。

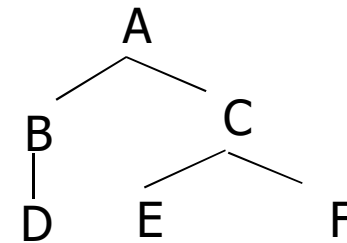
主管成分是结构的中心语（head）；

被管成分是依存于中心语的成分，包括补足语、
附接语、指示语等



- 统制（command）是一种句法结构关系。简言之，一个节点可以统制树结构上它的兄弟节点以及兄弟节点的子孙节点。

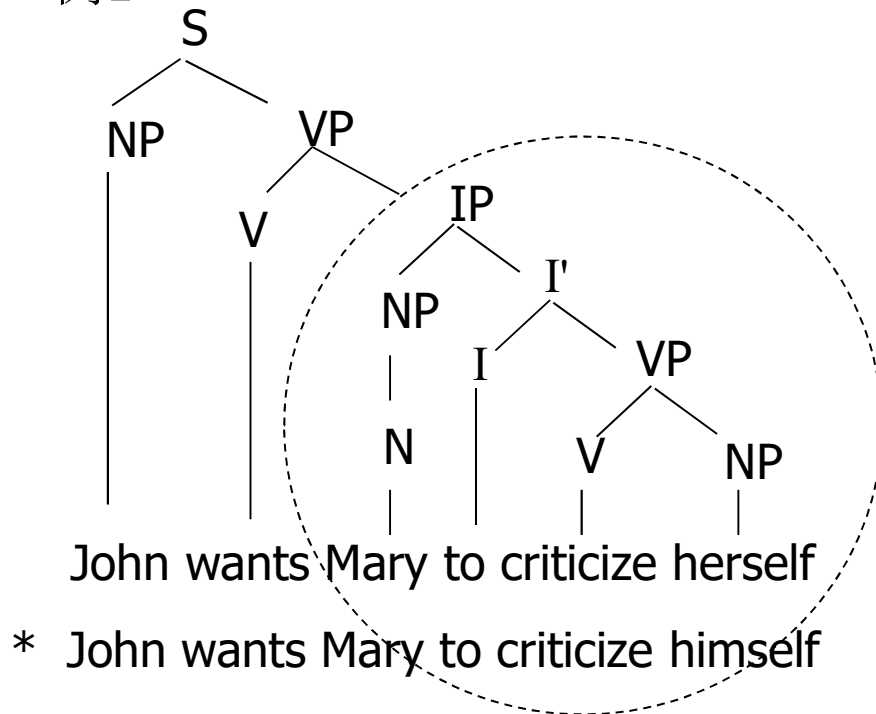
X统制Y，当且仅当Y从属于X所属的第一个有分叉的节点，且X、Y互不从属（无父子/爷孙关系）。



B 统制 C, E, F
C 统制 B, D
D 统制 C, E, F
E 统制 F
F 统制 E

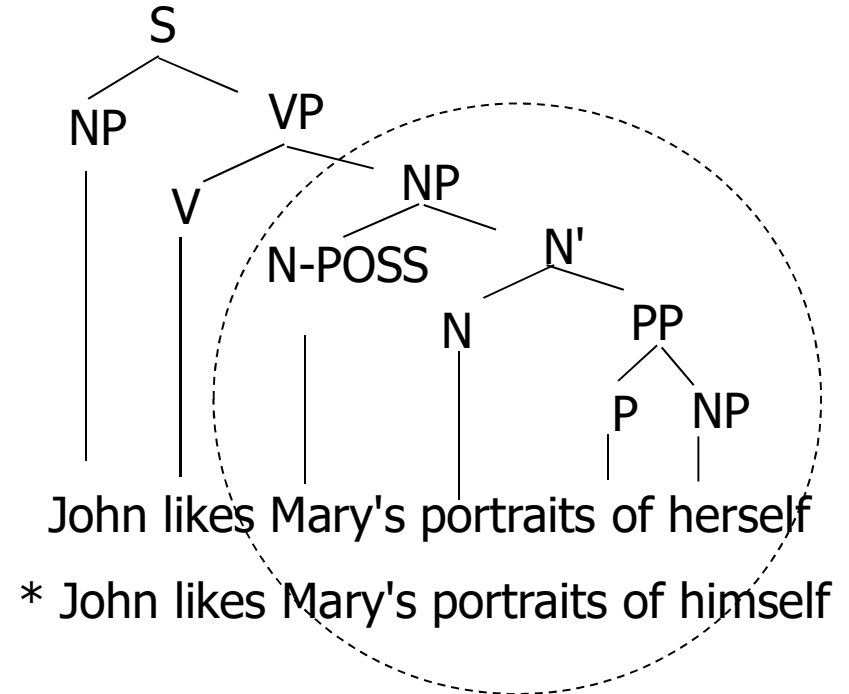
管辖语域 (governing category)

例1



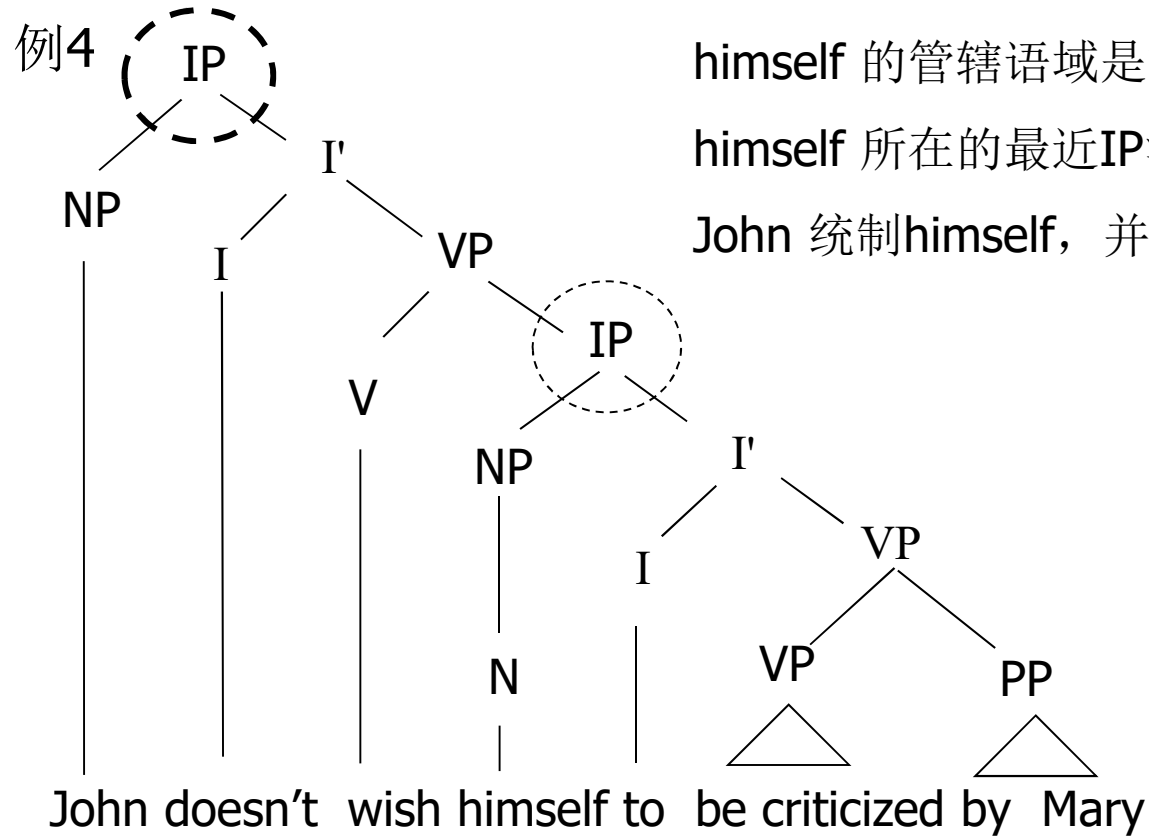
IP 是 herself 的管辖语域
Mary 是批评herself 的逻辑主语

例2



NP 是 herself 的管辖语域
Mary是画像herself的逻辑主语

管辖语域 (governing category)



himself 的管辖语域是根节点 IP;

himself 所在的最近IP没有统制它的逻辑主语节点;

John 统制himself, 并且是 wish himself 的逻辑主语

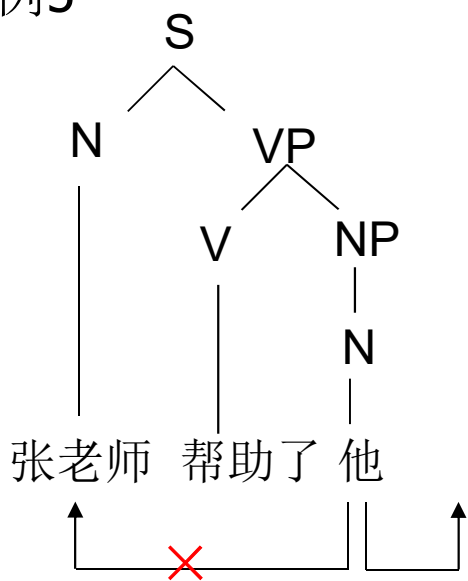
* John doesn't wish himself is criticized by Mary

时态句的主语均为主格，主格照应语必约束在主格成分所在小句内部 —— 主格禁区条件 (NIC)

Nominative Island Condition

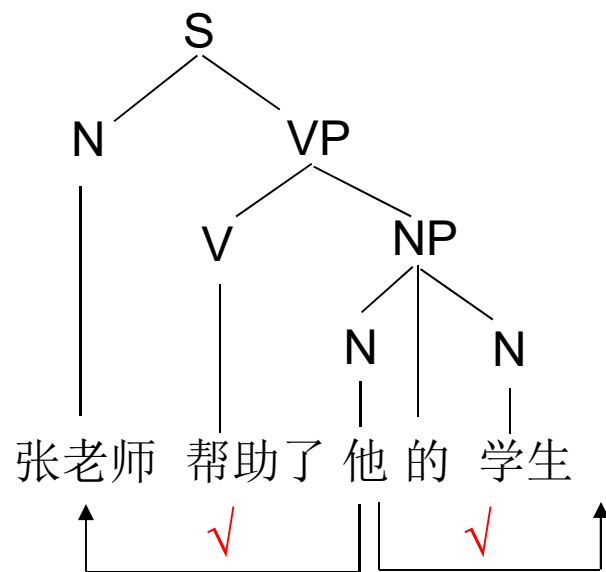
影响代词指代的因素：句法结构

例5



符合约束原则B

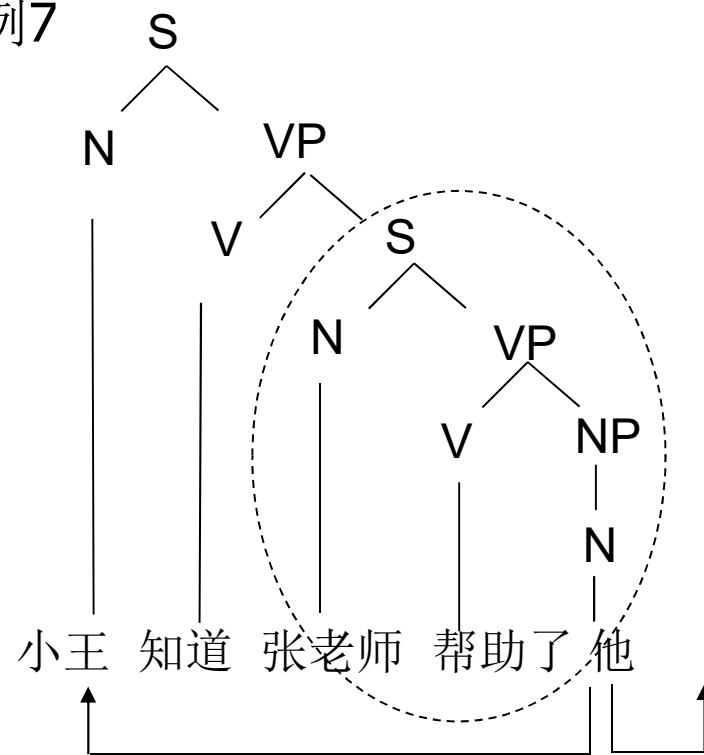
例6



不符合约束原则B

约束原则B未涉及?

例7



约束原则B未涉及



影响代词指代的因素：论元结构与词汇义

1. 小王的父亲感谢老李找到了他的车子。 他 = 父亲 or X
2. 小王的父亲听说老李找到了他的车子。 他 = 父亲 or 老李 or X
3. 孙国仁不顾他的阻挡，强行摘下他的帽子。 他 = X
4. 孙国仁不承认他的失败，一如既往地走进了考场。 他 = 孙国仁 or X
5. 小明打他的老师。 他 = 小明 or X
6. 小明是他的老师。 他 = X
7. 小明可以解雇他的医生。 他 = 小明 or X
8. 小明可以当他的医生。 他 = X



影响代词指代的因素：论元结构与词汇义

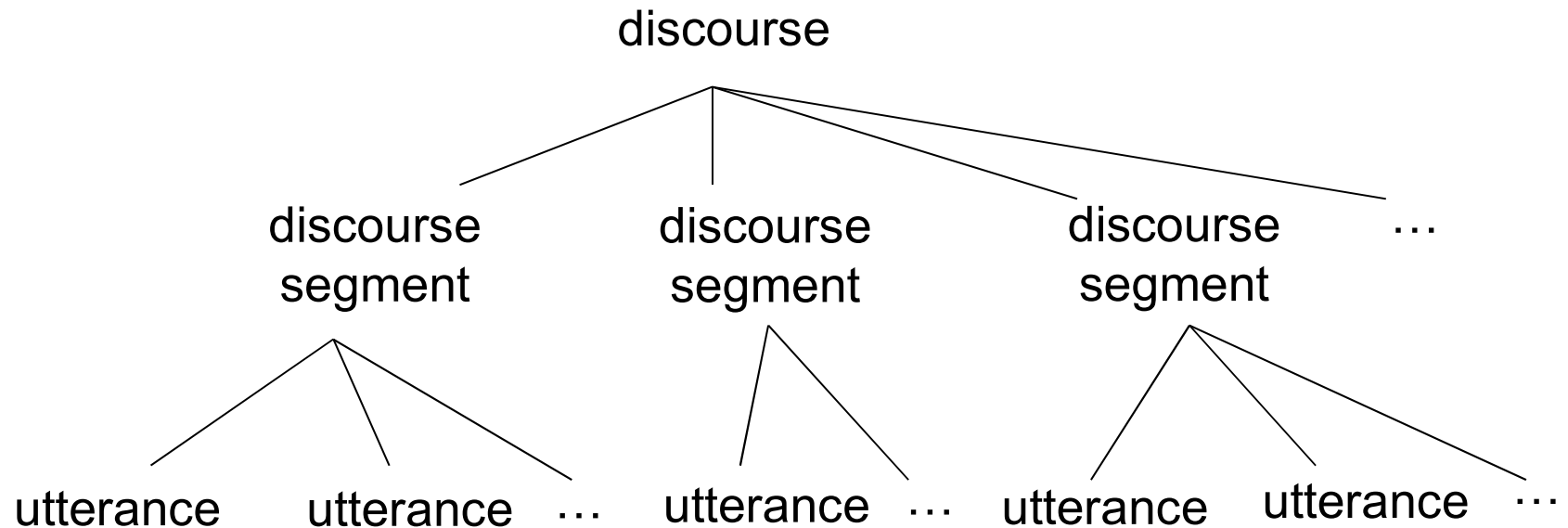
- 代词所指向的**NP**不能跟代词是同一个动词的题元；
- 代词倾向于指向离它更近的**NP**；
- 代词倾向于指向更显著的**NP**；

主语比修饰语显著
施事比非施事显著
有生**NP**比无生**NP**显著

.....



2 CT (语段中心成分理论)



Grosz, Joshi & Weinstein, 1995, Centering : A Framework for Modelling the Local Coherence of Discourse, in Computational Linguistics Vol.21, No.2. (Harvard University)



语段中心成分（center）

- 语段中心成分是指在篇章片段中，联系不同语段的那些实体。
 1. 语段中心成分是一种篇章结构的概念，同一形式的语段在不同的篇章中将会有不同的语段中心成分。
 2. 语段中心成分是语义实体，而不是单纯的词、词组或一些句法形式。



center的类型

- 语段潜在中心 (**forward-looking center**, 简写为**C_f**)
一个语段可能存在的会话焦点
- 语段现实中心(**backward-looking center**, 简写为**C_b**)
一个语段的当前会话焦点



C_f 和 C_b 的性质

1. 一个篇章片段中的每一语段都可以有一组潜在中心，但现实中心最多只能有一个。
2. 每一语段的潜在中心在下一语段中起码应该部分实现。
3. $C_b(U_n, DS)$ 是 $C_f(U_{n-1}, DS)$ 中的一个，而且是 U_n 所实现的 $C_f(U_{n-1}, DS)$ 中阶 (**rank**) 最高的那个 C_f



center的rank（阶）

- 语段中心成分的“阶”是指语段中心成分在语段中的显著度。

主语 > 直接宾语 > 间接宾语 > 补足语 > 附接语

在一个语段潜在中心集合里，阶最高的那个成员为优选中心 (Preferred Center, 简称C_p)

center: 示例1

- 小明和妈妈去逛街。他看中了一件衣服。但小明太胖了，所以妈妈没给小明买那件衣服。

语段		语段中心成分
U ₁ :	小明和妈妈去逛街。	C _f : 小明、妈妈、街 C _p : 小明 C _b : undefined
U ₂ :	他看中了一件衣服。	C _f : 他(=小明)、一件衣服 C _p : 他(=小明) C _b : 实现的C _f (U ₁): 他(=小明)
U ₃ :	但小明太胖了，	C _f : 小明 C _p : 小明 C _b : 实现的C _f (U ₂): 小明
U ₄ :	所以妈妈没给小明买那件衣服。	C _f : 妈妈、那件衣服、小明 C _p : 妈妈 C _b : 实现的C _f (U ₃): 小明

center: 示例2

- 小明和妈妈去逛街。他看中了一件衣服。但那件衣服太艳了，所以妈妈没给小明买那件衣服。

语段		语段中心成分
U ₁ :	小明和妈妈去逛街。	C _f : 小明、妈妈、街 C _p : 小明 C _b : undefined
U ₂ :	他看中了一件衣服。	C _f : 他(=小明)、一件衣服 C _p : 他(=小明) C _b : 实现的C _f (U ₁): 他(=小明)
U ₃ :	但那件 衣服 太艳了，	C _f : 那件衣服 C _p : 那件衣服 C _b : 实现的C _f (U ₂): 那件衣服
U ₄ :	所以 妈妈 没给 小明 买 那件衣服 。	C _f : 妈妈、那件衣服、小明 C _p : 妈妈 C _b : 实现的C _f (U ₃): 那件衣服



center的跳转类型

■ Centering Transition State

	$C_b(U_n)=C_b(U_{n-1})$ or $C_b(U_{n-1})$ undefined	$C_b(U_n)\neq C_b(U_{n-1})$
$C_b(U_n)=C_p(U_n)$	延续(Continue)	顺转(Smooth Shift)
$C_b(U_n)\neq C_p(U_n)$	保持(Retain)	硬转(Rough Shift)

$$n \geq 1$$

如果 $C_b(U_n)=\text{undefined}$ ，则跳转类型为 Φ ，这表示新的篇章片段的开始



跳转类型： 示例

列车员检票时发现，一个苏格兰成人用的是儿童票。但苏格兰人坚决不肯补票。于是检票员拿起旅客的衣箱就往车外扔。此时，火车正在过桥。“您疯啦！”苏格兰人狂喊，“您跟我的票过不去，又淹死了我的弟弟！”

跳转类型： 示例

语段		语段中心成分	跳转类型
U ₁ :	列车员检票时发现，一个苏格兰成人用的是儿童票。	C _f : 列车员、苏格兰成人、儿童票 C _b : undefined C _p : 列车员	C _t : Φ
U ₂ :	但苏格兰人坚决不肯补票。	C _f : 苏格兰人、票 C _b : 苏格兰人 C _p : 苏格兰人	C _t : 延续(continue)
U ₃ :	于是检票员拿起旅客的衣箱就往车外扔。	C _f : 检票员、衣箱、苏格兰人(旅客)、车 C _b : 苏格兰人 C _p : 检票员	C _t : 保持(retain)
U ₄ :	此时，火车正在过桥。	C _f : 车、桥 C _b : 车 C _p : 车	C _t : 顺转(smooth-shift)
U ₅ :	“您疯啦！” 苏格兰人狂喊，“您跟我的票过不去，又淹死了我的弟弟！”	C _f : 您（检票员）、苏格兰人、票、弟弟 C _b : undefined C _p : 苏格兰人	C _t : Φ



center的跳转类型与篇章的连贯性

- 延续 > 保持 > 顺转 > 硬转

Continue	主题不变，主题显著
Retain	主题不变，主题不显著
Smooth Shift	主题变化，主题显著
Rough Shift	主题变化，主题不显著

跳转类型与连贯性： 示例1

小明今天休息，所以他和妈妈去逛街了。他看中了一件衣服。

语段		语段中心成分	跳转类型
U ₁ :	小明今天休息，	C _f : 小明 C _p : 小明 C _b : undefined	C _t : Φ
U ₂ :	所以他和妈妈去逛街了。	C _f : 他(=小明)、妈妈、街 C _p : 他(=小明) C _b : 小明	C _t : 延续
U ₃ :	他看中了一件衣服。	C _f : 他(=小明)、一件衣服 C _p : 他(=小明) C _b : 小明	C _t : 延续

跳转类型与连贯性： 示例2

小明今天休息，所以他和妈妈去逛街了。一件衣服被他看中了。

语段		语段中心成分	跳转类型
U ₁ :	小明今天休息，	C _f : 小明 C _p : 小明 C _b : undefined	C _t : Φ
U ₂ :	所以他和妈妈去逛街了。	C _f : 他(=小明)、妈妈、街 C _p : 他(=小明) C _b : 小明	C _t : 延续
U ₃ :	一件衣服被他看中了。	C _f : 一件衣服、他(=小明) C _p : 一件衣服 C _b : 小明	C _t : 保持

跳转类型与连贯性： 示例3

小明今天休息，所以他和妈妈去逛街了。妈妈走累了。

语段		语段中心成分	跳转类型
U ₁ :	小明今天休息，	C _f : 小明 C _p : 小明 C _b : undefined	C _t : Φ
U ₂ :	所以他和妈妈去逛街了。	C _f : 他(=小明)、妈妈、街 C _p : 他(=小明) C _b : 小明	C _t : 延续
U ₃ :	妈妈走累了。	C _f : 妈妈 C _p : 妈妈 C _b : 妈妈	C _t : 顺转

跳转类型与连贯性： 示例4

小明今天休息，所以他和妈妈去逛街了。一个人撞了妈妈一下。

语段		语段中心成分	跳转类型
U ₁ :	小明今天休息，	C _f : 小明 C _p : 小明 C _b : undefined	C _t : Φ
U ₂ :	所以他和妈妈去逛街了。	C _f : 他(=小明)、妈妈、街 C _p : 他(=小明) C _b : 小明	C _t : 延续
U ₃ :	一个人撞了妈妈一下。	C _f : 一个人、妈妈 C _p : 一个人 C _b : 妈妈	C _t : 硬转



利用CT理论做指代消解

■ Anaphora Resolution基本方法:

- step1: 为每个语段中的实体生成可能的 C_b - C_f 组;
- step2: 通过各种约束条件来过滤 (比如: 句法位置约束、语义选择限制, 等等);
- step3: 通过连贯性来评级: 如果一个代词R的指代成分为A所得到的篇章连贯性高于指代成分为B时得到的篇章连贯性, 则将R的指代成分确定为A。

指代消解示例1 (续)

	语段	中心成分	跳转类型
U1	John saw a beautiful Acura Integra at the dealership.	Cf(U1): {John, Integra, dealership} Cp(U1): John Cb(U1): undefined	Φ
U2	He showed it to Bob.	Cf(U2): {John=he, Integra=it, Bob} Cp(U2): John Cb(U2): John	continue
U3	He bought it.	Cf(U3): {John=he, Integra=it} Cp(U3): John ? Cb(U3): John	continue

指代消解示例1 (续)

	语段	中心成分	跳转类型
U1	John saw a beautiful Acura Integra at the dealership.	Cf(U1): {John, Integra, dealership} Cp(U1): John Cb(U1): undefined	Φ
U2	He showed it to Bob.	Cf(U2): {John=he, dealership=it, Bob} Cp(U2): John Cb(U2): John	continue
U3	He bought it.	Cf(U3): {Bob=he, dealership=it} Cp(U3): Bob ? Cb(U3): Bob	smooth

指代消解示例1（续）

	中心成分	跳转类型	中心成分	跳转类型
U1	Cf(U1): {John, Integra, dealership} Cp(U1): John Cb(U1): undefined	Φ	Cf(U1) {John, Integra, dealership} Cp(U1) John Cb(U1) undefined	Φ
U2	Cf(U2): {John=he, Integra=it, Bob} Cp(U2): John Cb(U2): John	continue	Cf(U2) {John=he, dealership=it, Bob} Cp(U2) John Cb(U2) John	continue
U3	Cf(U3): {John=he, Integra=it} Cp(U3): John Cb(U3): John	continue	Cf(U3) {Bob=he, dealership=it} Cp(U3) Bob Cb(U3) Bob	smooth

跳转类型有助判断U3中he指John，但对U2中it的指称，则无法确定。



指代消解示例2

1. Terry really goofs sometimes.
2. Yesterday was a beautiful day and he was excited about trying out his new sailboat.
3. He wanted Tony to join him on a sailing expedition.
4. He called him at 6am.
5. He was sick and furious at being woken up so early.
6. He told Terry to get lost and hung up.
7. Of course, Terry hadn't intended to upset Tony.

he = Terry

he = Terry

he=? him=?

he=?

指代消解示例2

	语段	语段中心成分	跳转类型
U ₁	Terry really goofs sometimes.	Cf={Terry}, Cb=undefined	Φ
U ₂	Yesterday was a beautiful day and he was excited about trying out his new sailboat.	Cf={he=Terry, sailboat}, Cp=Terry, Cb=Terry,	continue
U ₃	He wanted Tony to join him on a sailing expedition.	Cf={Terry, Tony, expedition}, Cp=Terry, Cb=Terry	continue
U ₄	He called him at 6am.	Cf={he=Terry, him=Tony}, Cp=Terry, Cb=Terry	continue
U ₅	He was sick and furious at being woken up so early.	Cf={he=Tony}, Cp=Tony, Cb=Tony	smooth
U ₆	He told Terry to get lost and hung up.	Cf={he=Tony, Terry}, Cp=Tony, Cb=Tony	continue
U ₇	Of course, Terry hadn't intended to upset Tony.	Cf={Terry, Tony}, Cp=Terry, Cb=Tony	retain

指代消解示例2

	语段	语段中心成分	跳转类型
U ₁	Terry really goofs sometimes.	Cf={Terry}, Cb=undefined	Φ
U ₂	Yesterday was a beautiful day and he was excited about trying out his new sailboat.	Cf={he=Terry, sailboat}, Cp=Terry, Cb=Terry,	continue
U ₃	He wanted Tony to join him on a sailing expedition.	Cf={Terry, Tony, expedition}, Cp=Terry, Cb=Terry	continue
U ₄	He called him at 6am.	Cf={he=Tony, him=Terry}, Cp=Tony, Cb=Terry	retain
U ₅	He was sick and furious at being woken up so early.	Cf={he=Terry}, Cp=Terry, Cb=Terry	continue
U ₆	He told Terry to get lost and hung up.	Cf={he=Tony, Terry}, Cp=Tony, Cb=Terry	retain
U ₇	Of course, Terry hadn't intended to upset Tony.	Cf={Terry, Tony}, Cp=Terry, Cb=Tony	rough

指代消解示例2

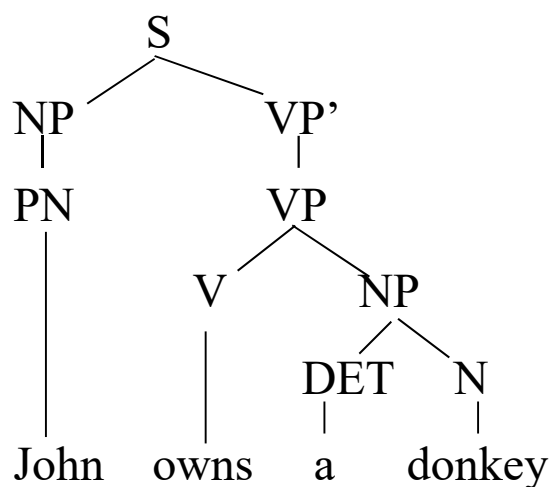
	语段中心成分	跳转类型	语段中心成分	跳转类型
U ₁	C _f ={Terry}, C _b =undef	Φ	C _f ={Terry}, C _b =undef	Φ
U ₂	C _f ={Terry,sailboat}, C _p =Terry,C _b =Terry,	continue	C _f ={Terry,sailboat}, C _p =Terry,C _b =Terry,	continue
U ₃	C _f ={Terry,Tony,expedition}, C _p =Terry, C _b =Terry	continue	C _f ={Terry,Tony,expedition}, C _p =Terry, C _b =Terry	continue
U ₄	C _f ={he=Terry,him=Tony}, C _p =Terry, C _b =Terry	continue	C _f ={he=Tony,him=Terry}, C _p =Tony, C _b =Terry	retain
U ₅	C _f ={he=Tony}, C _p =Tony,C _b =Tony	smooth	C _f ={he=Terry}, C _p =Terry,C _b =Terry	continue
U ₆	C _f ={he=Tony,Terry}, C _p =Tony,C _b =Tony	continue	C _f ={he=Tony,Terry}, C _p =Tony,C _b =Terry	retain
U ₇	C _f ={Terry,Tony}, C _p =Terry,C _b =Tony	retain	C _f ={Terry,Tony}, C _p =Terry,C _b =Tony	rough



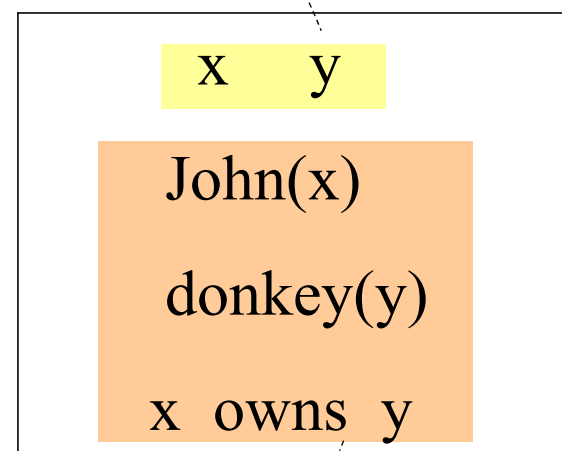
3 DRT理论（篇章表示理论）

- Hans Kamp, 1981, *A theory of truth and semantic representation*, In J. Groenendijk, T. Janssen & M. Stokhof, eds., *Truth, Interpretation and Information*, Dordrecht: Foris. pp.1-41.
- 潘海华, 1996, 《篇章表述理论概述》, 载《国外语言学》1996年第3期。
- **特色:** 动态处理代词所指、无定名词词组的语义解释
- **组成:**
 1. 篇章语义构造规则与算法: 句子（句群）→ DRS
 2. 篇章语义的正确性验证: 给出模型对DRS进行解释

从句法树到DRS (篇章表示结构)



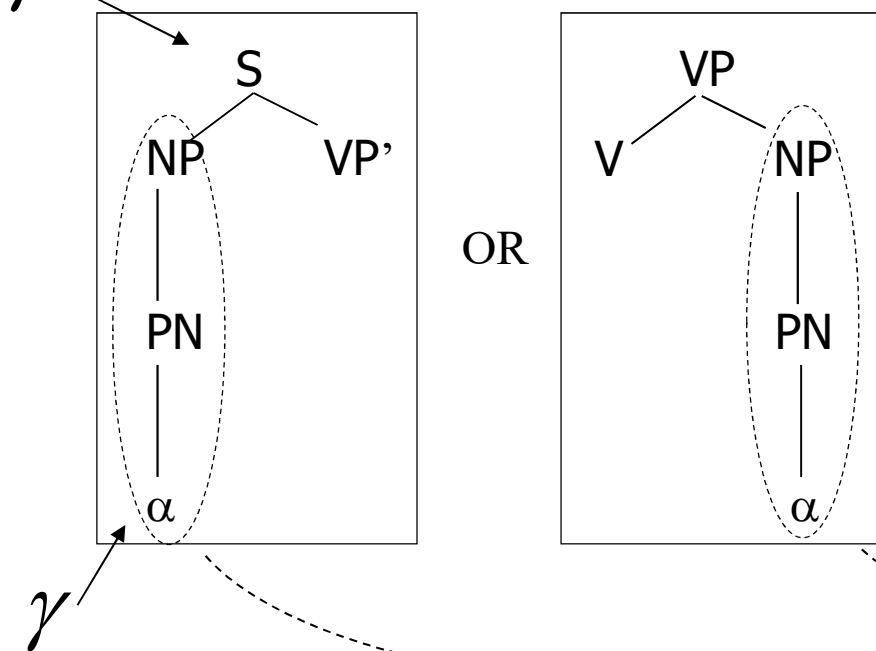
话语所指集合, 记做 U_k



DRS状况集合, 记做 Con_k

DRS构造规则 (专名的DRS构造规则)

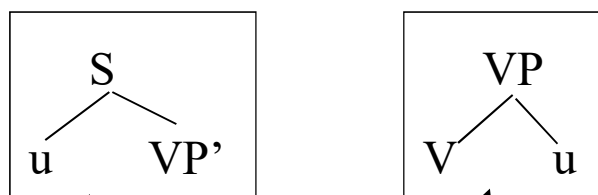
γ' 起始句法树与DRS格局



专名的DRS构造规则

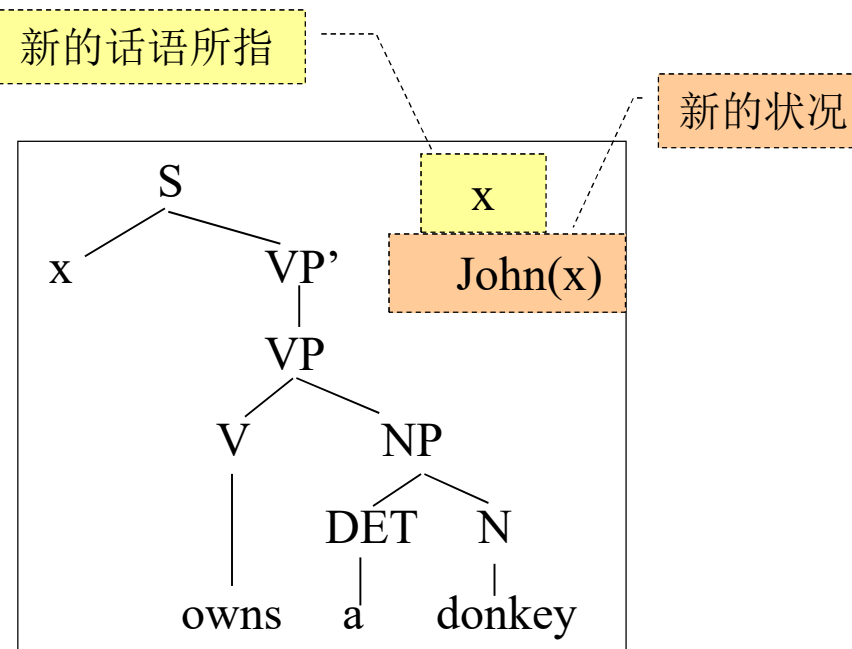
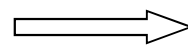
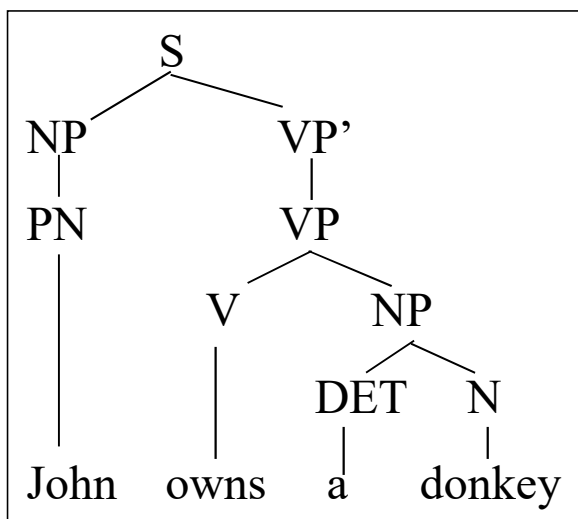
1. 在 U_k 中加入新的话语所指 u
2. 在 Con_k 中将加入新的状况 $\alpha(u)$
3. 在 γ' 中用 u 替换 γ

句法树格局变为: 输出DRS格局



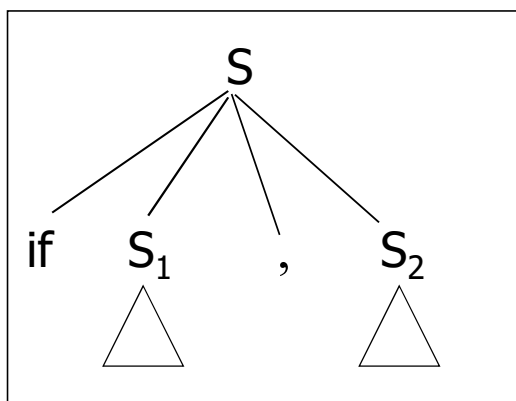
γ 是 γ' 的子树

专名的DRS构造规则应用示例

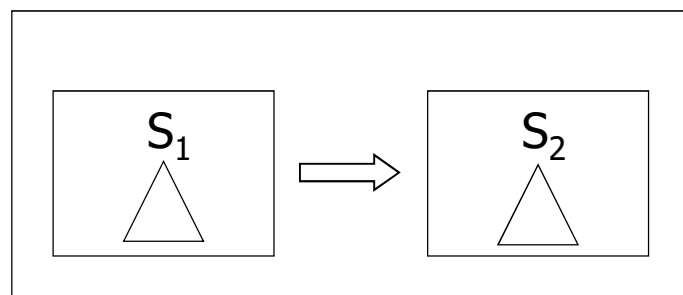


条件句的DRS构造规则

起始句法树与DRS格局

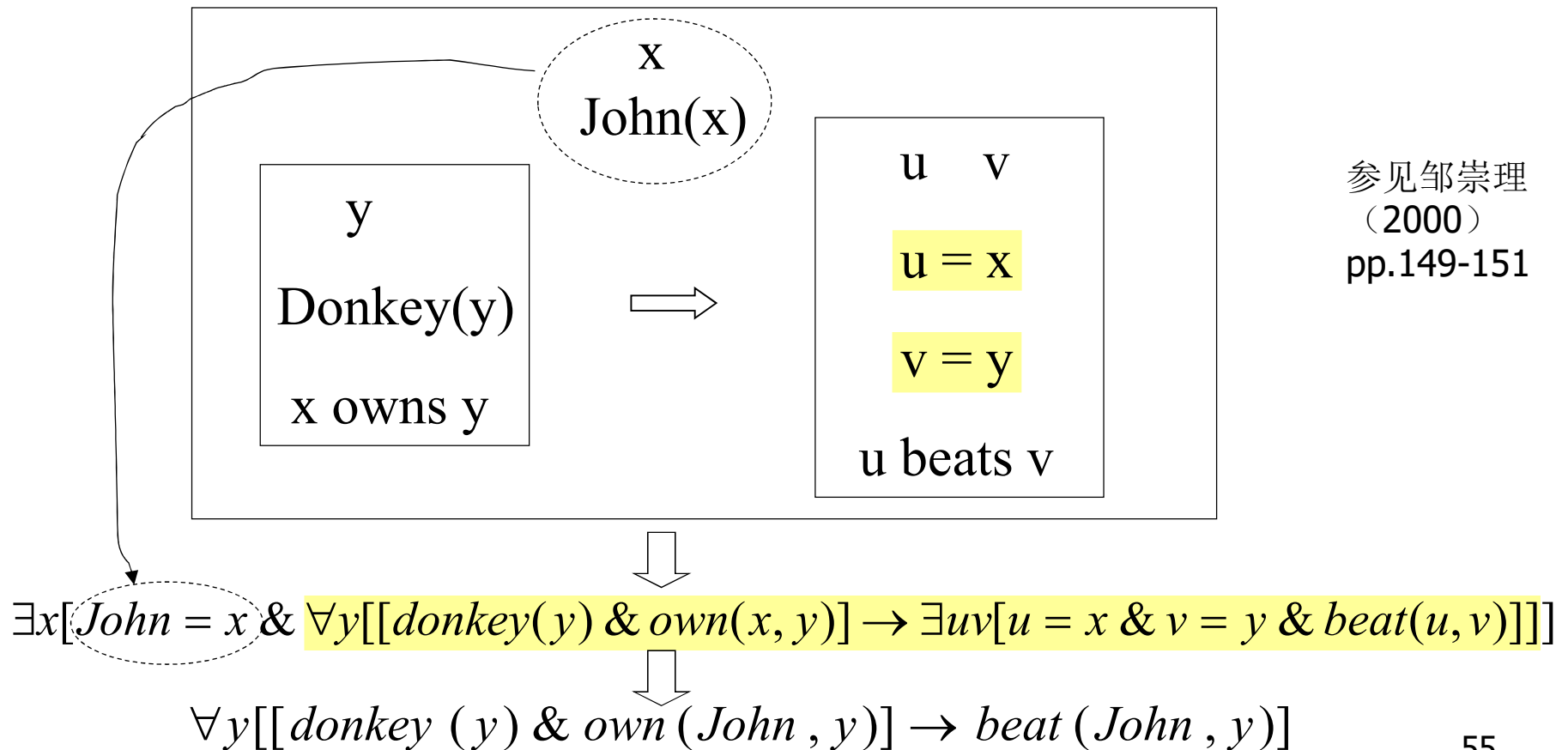


输出DRS格局



条件句的DRS (从DRS到逻辑表达式LF)

If John owns a donkey, he beats it.



参见邹崇理
(2000)
pp.149-151



SDRT

- Nicholas Asher Univ. of Texas at Austin
- Asher, N. 1993, *Reference to Abstract Objects in Discourse*. Kluwer Academic Publishers.
- Asher, N. and Lascarides, N., 2003, *Logics of Conversation*. Cambridge University Press.
- 毛翊、周北海，2007，分段式语篇表示理论——基于语篇结构的自然语言语义学，《语言学论丛》35辑，pp114 -141。商务印书馆。



4 CDT（概念依存理论）

- Schank, R.C. ed., 1975, *Conceptual Information Processing*, Amsterdam: North-Holland.
- Schank, R.C. & Abelson, R., 1977, *Scripts, Plans, Goals and Understanding*, Hillsdale, NJ: Lawrence Erlbaum.
- CD理论三个层次
 - 概念依存图（Concept Dependency Graph）
 - 脚本、故事（Script）
 - 规划、目标（Plan、Goal）



概念基元

- **理想：**用若干种表示动作的概念原语 (Primitive) 和它们的组合就可以表达自然语言的各种语义。
- **实践：**

GRASP

MOVE

PTRANS

ATRANS

PROPEL

INGEST

HIT

MTRANS

CONCEPTUALIZE

MBUILD

SMELL

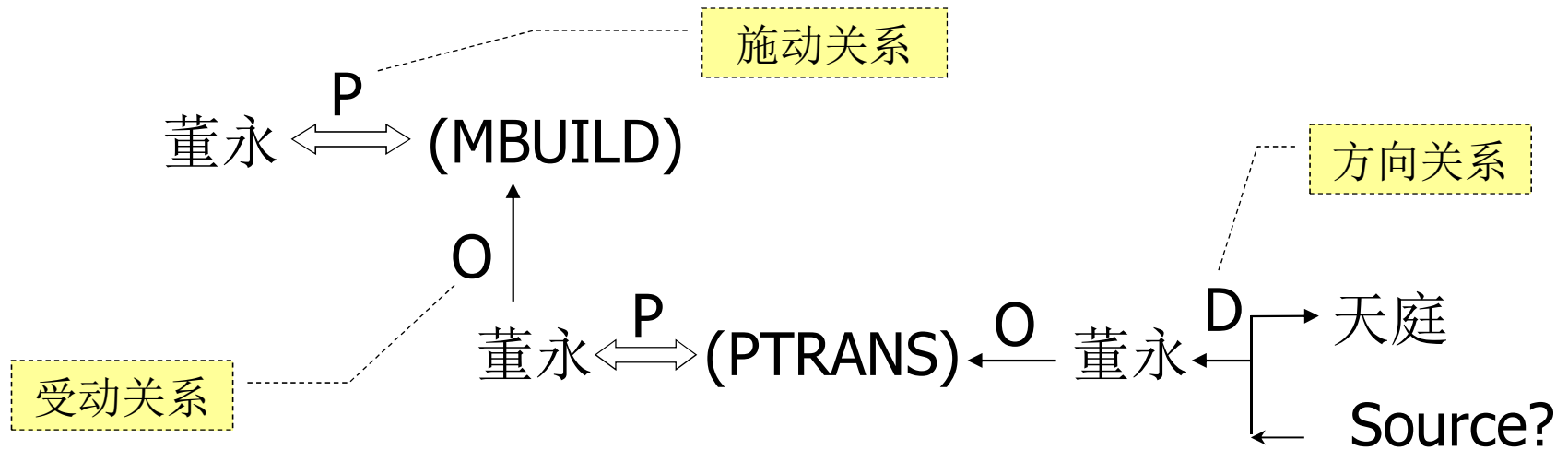
LOOK-AT

LISTEN-TO

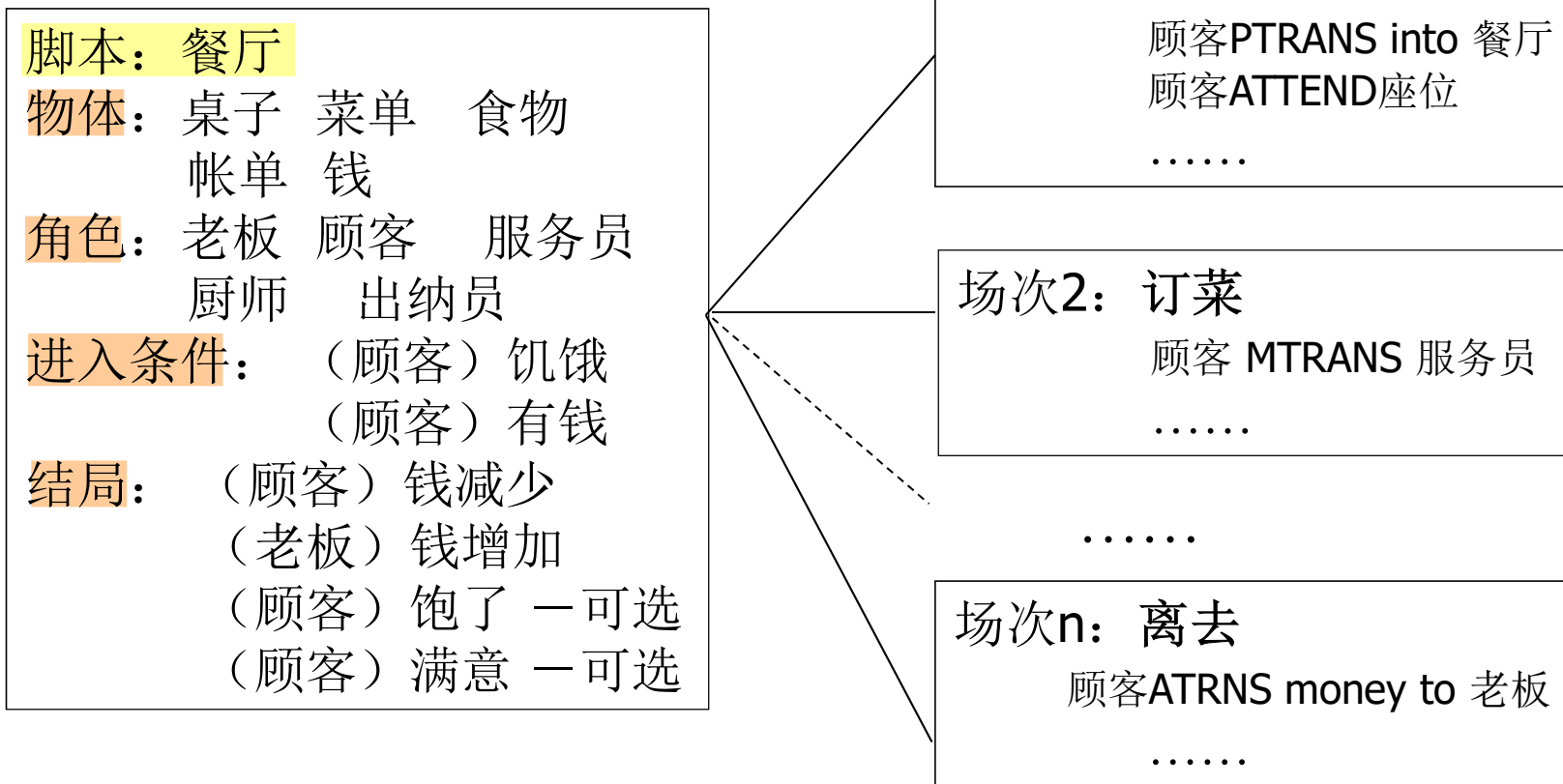
SPEAK

概念依存图

- 董永想去天庭



脚本（叙事体篇章的结构描述）





规划（在特定脚本基础上组织更大的篇章）

- 表述规划知识的规则类型：
 - 1 初始规则：主题 → 目标
e.g. 饿了(x) → 不饿(x)
 - 2 目标规则：目标 → 规划
e.g. 不饿(x) → 去餐馆吃饭(x)
 - 3 子目标规则：规划 → 规划的先决条件(目标)
e.g. 去餐馆吃饭(x) → 在餐馆(x)
 - 4 行为规则：规划 → 行为
e.g. 走到餐馆(x) → x PTRANS to 餐馆



叙事体篇章分析示例

John was hungry. He grabbed the Michelin guide. He went to his car.

theme(John was hungry)
→ goal (John be not hungry)
→ plan(John do eat_at_restaurant plan)
→ goal(John be at restaurant)
→ goal(John know where restaurant is)
→ plan(John read restaurant guide)
→ goal(John possess restaurant guide)
→ plan(John take restaurant guide)
→ ACT(John GRASP restaurant guide)

.....
→ goal(John be at restaurant)
→ plan(John use car)
→ goal(John be in car)
→ plan(John walk to car)
→ ACT(John PTRANS to car)

刘开瑛，郭炳炎（1991）《自然语言处理》，科学出版社，pp155-158



5 RST（修辞结构理论）

- William C. Mann & Sandra A. Thompson, 1988, *Rhetorical Structure Theory: Toward a functional theory of text organization*, *Text*, Vol.8, No.3, pp243-281.
- <http://www.sil.org/~mannb/rst/index.htm>
- 1980s, Studies of computer_based text generation, Information Sciences Institute, **Univ. of Southern California**（RST开始研究时的背景）



RST的两级篇章结构

- 上层： 整篇文本的结构框架(*Schema*)
- 基层： 局部段落中句与句之间的连贯关系(*Coherence relation*)

在RST理论中，篇章整体关系就是由这两个层次搭建起来的：

- (1) 句子与句子形成具有基本连贯关系和复杂连贯关系的单元；
- (2) 这些单元再组织为一定的结构框架，从而叠床架屋地形成一个完整的篇章。



RST的基层篇章结构

- 篇章基本单位：span（片段）
- 篇章知识：
 - （1）有关联的两个span通常是相邻的；
 - （2）若A和B是两个有关联的span，A通常扮演着B的某个角色，或者B扮演着A的某个角色；
- 在span的基础上，可以定义两个span之间的关联（*Relation*）
- 一般来说，满足一定关系的两个span，一个居于核心地位（*Nucleus*），一个居于从属地位（*Satellite*）。



RST关联类型示例

关系类型	核心片段	从属片段
让步CONCESSION	作者确认的情况A	作者确认的情况B，B与A明显不一致
条件CONDITION	在给定条件B下出现情况A	条件B
证明EVIDENCE	结论A	信息B，用来增强读者对A的相信程度

让步 尽管他很有道理（B），但还是没有说服他的母亲（A）。

条件 只要他有道理（B），他就可以说服他的母亲（A）。

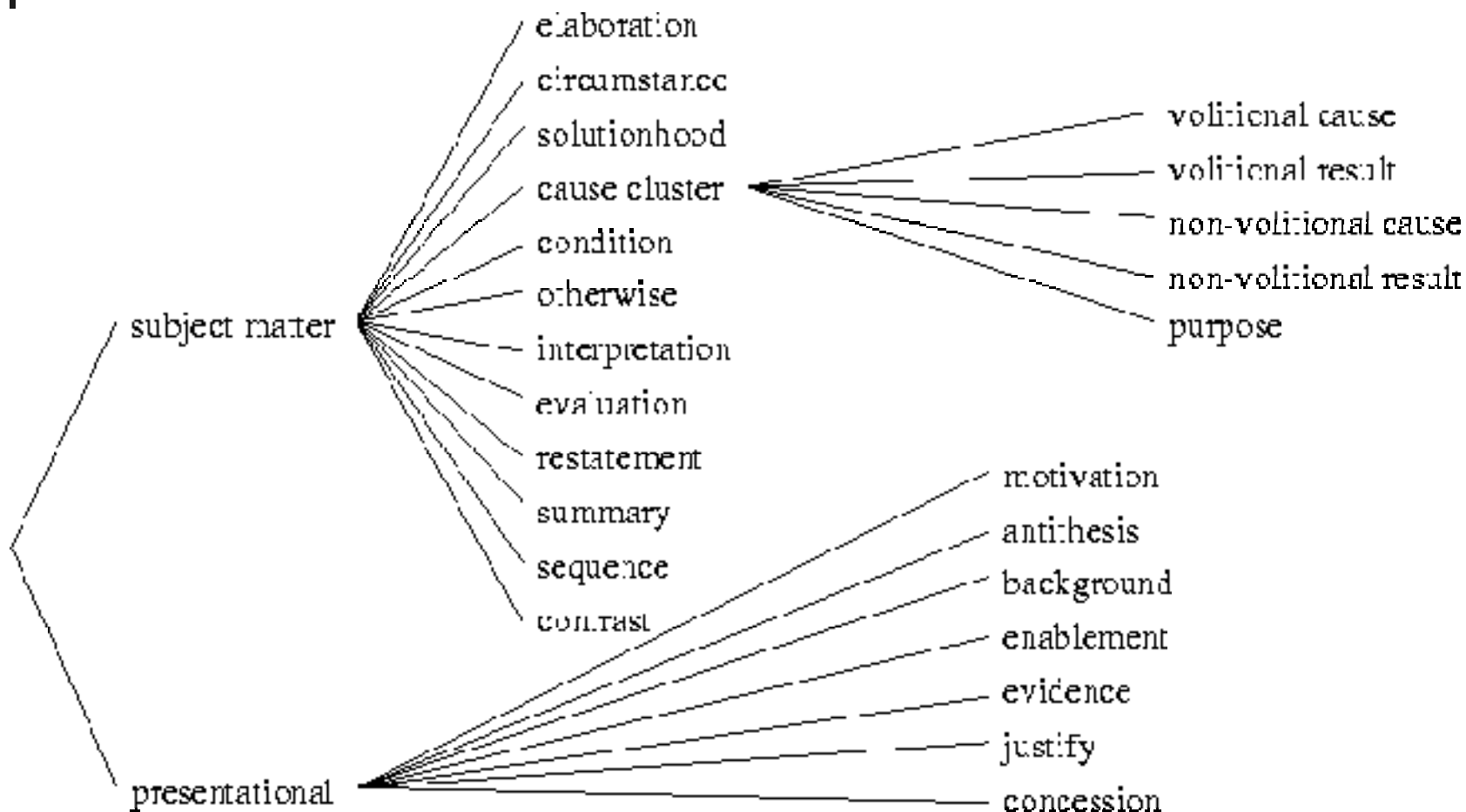
证明 因为他说的有道理（B），所以他母亲相信了（A）。



多核心关联关系 (multinuclear)

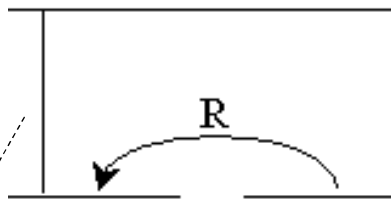
- 对比关系 (contrast)
Unlike Bill, Bob is hungry.
- 罗列关系 (list)
Bob has three sisters: Tess, Jess and Bess.
- 序列关系 (sequence)
Remove the white plastic tray and unpack the cable
- 其他关系 (joint)
John is 21 years old. He studies engineering.

Mann & Thompson给出的关系类型

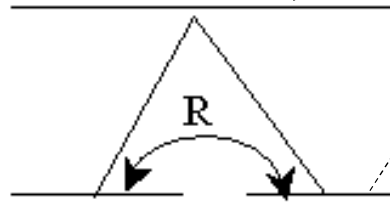


RST的上层结构图示(schema)

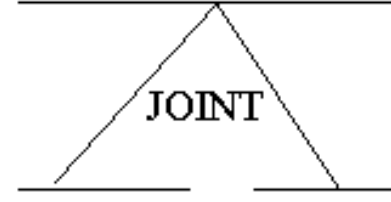
水平线表示span



a

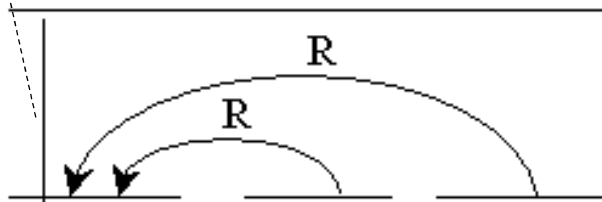


b 对比关系

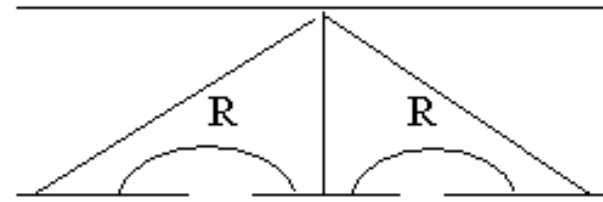


c 其他关系

垂直线表示核心成分



d



e 序列关系



RST篇章分析示例

1) Darwin as a Geologist

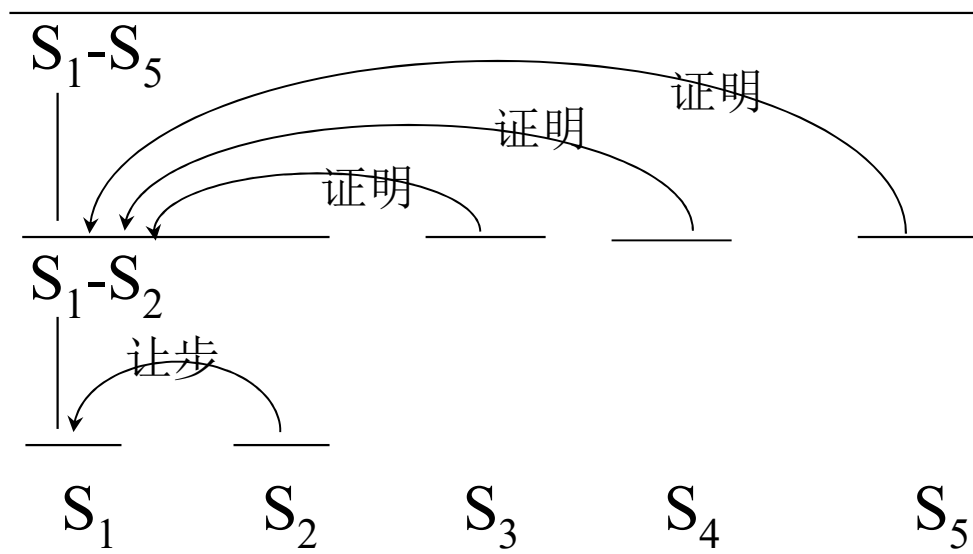
2) He tends to be viewed now as a biologist, 3) but in his five years on the Beagle his main work was geology, 4) and he saw himself as a geologist. 5) His work contributed significantly to the field.

1. 作为地理学家的达尔文

2. 他一般被人们看作是生物学家, 3. 但是在他加入贝戈尔号航海旅行的五年中, 他的主要工作是地理学研究, 4. 他把自己看作是个地理学家。5. 他的研究工作对这个领域做出了重大贡献。



RST篇章分析结果示意图





6 小结

- 指代 —— 同指、异指的约束条件：句法、语义？
- 省略 —— 如何补回？句法填充？论元结构？
- 篇章单位 —— 如何定义一个片段？
- 篇章关联 —— 片段间有多少种关系？

自然语言单位

音素 — 音节 — 语素 — 词 — 词组 — 句子 — 篇章

评价尺度： 对 —— 错； 好 —— 差（不好）

符号形式 —— 符号功能（意义）对应的不确定性？ ↑ ↓



进一步阅读文献

- 廖秋忠, 1986, 《现代汉语篇章中的连接成分》, 载《中国语文》1986年第6期。
- 廖秋忠, 1988, 《篇章中的论证结构》, 载《语言教学与研究》1988年第1期。
- 胡壮麟, 1994, 语篇的衔接与连贯, 上海外语教育出版社。
- 黄国文, 1988, 语篇分析概要, 湖南教育出版社。
- 邹崇理, 2000, 自然语言逻辑研究, 北京大学出版社, 第3章。
- 蒋严、潘海华, 1988, 形式语义学引论, 中国社会科学出版社, 第14章。
- 姚天顺等, 1995, 自然语言理解, 清华大学出版社, 第5, 6章。
- 毛翊、周北海, 2007, 分段式语篇表示理论——基于语篇结构的自然语言语义学, 《语言学论丛》35辑, pp114 -141。商务印书馆。
- 冯志伟、孙乐译, 《自然语言处理综论》(Daniel Jurafsky & James H.Martin原著, 电子工业出版社2005年版) pp.428 – 430
- Grosz, Joshi, and Weinstein, 1995, Centering : A Framework for Modelling the Local Coherence of Discourse, in Computational Linguistics 21(2),pp.203-225.
- James Allen, 1995, *Natural Language Understanding*, Second Edition, The Benjamin/Cummings Publishing Company, Inc., Chapter 14, 16.
- Asher, N. 1993, Reference to Abstract Objects in Discourse. Kluwer Academic Publishers.
- Asher, N. and Lascarides, N., 2003, *Logics of Conversation*. Cambridge University Press.



复习思考题

1 脑筋急转弯 ☺

三个鸡蛋放在两个盘子里，但一个盘子只能放一个鸡蛋，如何放法？

试从语言学的角度对上题进行分析。

2 一个男人临死前对siri说，**tell my wife, I love her.**

于是，他妻子的手机收到了一条来自丈夫手机的短信：**I love her.**

试从语言学角度对上面这个笑话进行分析。

3 试考察并归纳汉语说明文体的篇章结构模式

