支持中文句法结构套叠的组合范畴语法

王庆江1 张琳2

- (1. 华北水利水电大学 信息工程学院,河南 郑州 450046;
- 2. 河南财经政法大学 现代教育技术中心, 河南 郑州 450046)

摘要: 词组入句时发生语法性质转变,词组的句法范畴就应该转换,这种转换依赖词组作为整体进入的更大词组的结构需要。要解释这种现象,组合范畴语法应该增加相应的转换规则。从集合论角度看,汉语有名词-动词-形容词依次包含,由此可归纳基本句法结构之间的转类套叠。在类有定职的前提下,得出转类套叠对应的范畴转换规则,形成有范畴转换机制的组合范畴语法 C²-CCG。示例说明这种基于词组本位的形式语法有希望能解释中文句法中各种转类套叠。

关键词:组合范畴语法;结构套叠;范畴转换

中图分类号: TP391 文献标识码: A

Combinatory Categorial Grammar for Chinese Syntactic Structure Overlapping

WANG Qingjiang¹, ZHANG Lin²

- (1. School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou, Henan 450046, China;
 - 2. Modern Educational Technology Center, Henan University of Economics and Law, Zhengzhou, Henan 450046, China)

Abstract: When a phrase enters into a sentence, its syntactic category should change with the change of its grammatical nature, which depends on the structural requirements of bigger phrases the phrase joins into. To explain the phenomenon, Combinatory Categorial Grammar should be augmented with the corresponding conversional rules. In view of set theory, Chinese has noun-verb-adjective successive inclusion, from which various grammatical nature-changed overlapping between basic syntactic structures could be induced. Under the premise of grammatical natures having their definite syntactic functions, category-conversional rules corresponding to nature-changed overlapping were obtained, and the new Combinatory Categorial Grammar C²-CCG with category-conversional mechanism was formed. Examples show this phrase-based formal grammar is promising in explaining various grammatical nature-changed overlapping in Chinese syntax.

Key words: combinatory categorial grammar; structural overlapping; categorial conversion

0 引言

组合范畴语法^{[1][2]}(Combinatory Categorial Grammar,缩写 CCG)早期的设计背景是英语和荷兰语,对范畴语法的显著改进是增加类型提升和函数组合,以解释宾语提取(Object Extraction)、状语(Adverbial Modifier)和补语(Complement)。后来引入斜线模态,将范畴结合的精确控制由规则一侧转到词汇一侧,使范畴结合规则尽可能与所施用的语言无

关,让语言的语法特殊性体现在词范畴设计上。CCG 句法分析伴随 λ-语义演算[3][4][5],在产生句法树的同时产生谓词-论元结构。CCG 正成为自然语言句法、语义形式化的理想工具之一。

汉语较英语的主要区别是转类时词没有曲折变化,形成基于词组本位的句法。CCG用于汉语时应建立于词组本位思想之上,词组入句时语法性质发生变化,词组的句法范畴就应该转换,这种转换依赖词组作为整体进入的更大词组的结构需要。

收稿日期: 2019-07-31; 定稿日期: 2019-08-14

基金项目:河南省高等学校重点科研项目计划(17A520042)

赵元任先生(1948)《国语入门》[6]阐述了汉语造 句的基本结构,即主谓、并列、向心、动宾和连动, 丁声树等(1961)把句法结构扩展为六种[7],朱德熙 先生(1982)《语法讲义》[8]明确阐述了词组本位思想, 即可以把各类词组(主谓、述宾、述补、偏正、联合、 连动)作为抽象的句法格式来描写它们的内部结构以 及每一类词组作为一个整体在更大的词组里的分布状 况,这种思想在陆俭明先生(1990)论著^[9]中也被表 述为结构包孕(或套叠),即结构的基本类型虽然很有 限,但每一种结构都可以包孕与它自身同类型或不同 类型的结构,这些被包孕的结构又可以包孕与它自己 同类型或不同类型的结构,如结构套叠<1>、<2>。刘 乃仲老师对同构套叠(见下文定义 2)中句法成分的 相对位置关系进行了分析[10]。句法上的这种递归性, 无疑产生于汉语缺乏系统的形态标记这一基本事实 [11]。

自清末学者马建忠(1898)参照拉丁语法建立汉语词类系统,词类的基本格局至今未变,而划分词类的标准始终在争议中。汉语语法学界以前认为的,词类根据词的意义和词的语法特点来划分,造成名词做谓语、动词做主语等"类无定职"现象,即一类词可充当多种句法成分[12]。随着结构主义语法理论深入人心,现在的主流共识是,词类的划分只依赖分布特点或语法功能,即"类有定职"。缺乏曲折变化的汉语较英语有更多的兼类词[13],其本质就是类有定职情况下词根据句法需要发生了转类。

既然词按句法功能归类,而且"类有定职",就可以建立"词类-句法成分-范畴"的简单映射。例如对充当述语的动词指派范畴 s\np/np,为充当宾语的名词指派范畴 np,动词和名词拼接时发生函数 s\np/np 应用于 np 的运算,形成述宾结构,其范畴是 s\np,语法性质是可充当谓语的动词性词组。词组的语法性质与该词组作为一个整体在更大的词组里充当的句法成分相对应,例如结构套叠<2>中"一篇文章"的范畴是

np,是名词性词组,正好可做"写"的宾语,这样的套叠可以被组合范畴语法解释[14]。词组需要转变语法性质(相当于词的转类,下文统称转类),才能作为整体在更大的词组里充当所需的句法成分,意味着该词组的句法范畴发生了改变,例如结构套叠<1>中"发表文章"的范畴是 s\np,是动词性词组,需要转名词性词组,把范畴转为 np,才能做"同意"的宾语,这种转类之后的套叠不能被组合范畴语法解释。本文分析结构套叠中的词组转类,通过增加相应的范畴转换规则,使扩展后的组合范畴语法能解释各种经由转类的结构套叠。

1 汉语句法基本结构的套叠

一般认为,句法成分有主(Subject)、谓 (Predicate)、宾(Object)、定(Attribute)、状 (Adverbial)、补 (Complement)、述 (Verb) 等, 句 法基本结构有主谓(Subject-Predicate, SP)、定中 (Attribute-Headword , AH_n) 、 状中 (aDverbial-Headword , DH_v 或 DH_a)、中补 (Headword-Complement, H_vC 或 H_aC)、述宾 (Verb-Object, VO)、 连谓(Coordinate-Predicate, PP)、联合(Coordination, XX) 等,中心成分 H 的语法性质可以是名词 (n)、 动词(v)和形容词(a),用下标来区分。联合结构要 求成分间使用连词或顿号, 用连词或顿号的句法范畴 (X*X)/*X 解释联合[1], 使基本结构中可去掉联合。连 谓中的两个谓语不能颠倒顺序,之间没有连词或顿号, 不能归入联合,但可以归入状中,即第1谓语修饰第 2 谓语,用转类解释连谓。一个具体的句子是由这些 抽象的基本结构通过套叠构造出来的。假设 Γ ={S, P, O, A, D, C, V}, Ω ={SP, $AH_n, DH_v, VO, H_vC, H_aC$ }, 变量 $X, Y, W, Z \in \Gamma$ 且 $XY, WZ \in \Omega$, 有下面的定义。

定义 1. 左套叠和右套叠

基本结构 XY 中的左边成分 X 本身就是一个基本结构 WZ,可表示为产生式 $XY \to WZY$,这种情况称为左套叠。相应地, $XY \to XWZ$ 称为右套叠。

定义 2. 同构套叠和异构套叠

 $XY \rightarrow WZY$ 或 $XY \rightarrow XWZ$, 其中 WZ=XY 时称

为同构套叠,而 WZ≠XY 时称为异构套叠。

一种结构(如 XY)左或右套叠某种结构(如 WZ),其实是其左(X)或右(Y)成分重写为那种结构(WZ),这与词类、中心扩展规约^[12](Head Feature Convention)有关。

在"类有定职"前提下,词类(名词 n、动词 v、形容词 a 和副词 d)与句法成分有简单的对应关系如图 1 上半部分。补语由"得"字短语、介词短语、数量短语、趋向动词等充当,没有哪类词或短语它的典型功能是做补语,暂用 y 标记补语对应的语法性质。按中心扩展规约考虑基本结构的语法性质,即以一个成分为中心加以扩展,扩展后的结构的语法性质与中心成分的语法性质一致,得到图 1 下半部分。主谓结构是两心相照,语法性质记为 s(注意不要与范畴符号 s 混淆),作为整体充当名词性成分时,按转类考虑。副词一般都不受其他词类修饰,很难被扩展。

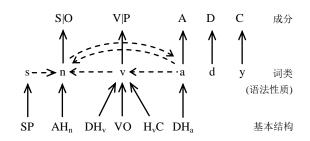


图 1 汉语的类有定职与中心扩展规约

按沈家煊先生"名-动-形"依次包含的思想^[12],形容词都可转动词或名词,动词都可转名词,如图 1中虚线箭头。名词都可以省略后接的助词"的"直接做修饰语,即可转形容词,这样虚线箭头构成一个环,这不意味着名词都可转动词、动词都可转形容词,因为违背了"名-动-形"依次包含,即不存在经过两个虚线箭头实现的转类。必须加助词"的"、"地"、"得"才能改变语法性质的,属于曲折变化,不考虑为转类。一个句法成分沿箭头逆向可达的终结点,就是它可重写为的基本结构,在该路径上允许出现至多一个虚线箭头,用 x/y 表示虚线箭头完成的 y 到 x 的转类,得到如下重写规则。

$$S \mid O \to n/s \to SP \tag{1}$$

$$S \mid O \rightarrow AH_n$$
 (2)

$$S \mid O \to n/v \to DH_v \mid VO \mid H_v C \tag{3}$$

$$S \mid O \rightarrow n/a \rightarrow DH_a$$
 (4)

$$V \mid P \to DH_v \mid VO \mid H_vC \tag{5}$$

$$V \mid P \to v/a \to DH_a \tag{6}$$

$$A \to DH_a \mid H_a C \tag{7}$$

$$A \to a/n \to AH_n \tag{8}$$

定义 3. 同类套叠和转类套叠

经由语法性质转变的句法成分到基本句法结构的 重写,称为转类重写,反之称为同类重写,相应的套 叠称为转类套叠和同类套叠。

由转类重写(1)、(3)、(4)、(6)和(8),得到基本结构之间的左套叠和右套叠,如表1和2。状语、补语的重写一般需要用助词完成语法性质的显式转变,范畴语法能解释这种经由助词的套叠,只需给助词指派反映语法性质转变的范畴符号^[15]。

表 1 结构 s1 转类后被结构 s0 左套叠

s1 s0	SP	AH_n	$\mathrm{DH_{v}}$	VO	H_vC	DH_{a}
SP	SPP					
AH_n		AH_nH_n				
DH_v	DH_vP					
VO	VOP					
H_vC	H _v CP					
DH _a	DH _a P		*	DH _a O	DH _a C	

^{*:} s1 须加"地"才能被 s0 套叠。

表 2 结构 s1 转类后被结构 s0 右套叠

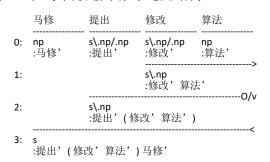
s1 s0	SP	AH_n	$\mathrm{DH_{v}}$	VO	H _v C	DHa
SP		ASP		VSP	*	
AH_n						DAH_n
DH_v		ADH_v		VDH_v	*	
VO		AVO		VVO	*	
H_vC		AH_vC		VH_vC		
DHa	SDH_a	ADH_a	DDH_a	VDH _a	*	

^{*:} s1 须加"得"才能被 s0 套叠。

2 句法成分的范畴指派

用范畴结合解释句法的基本结构,在两个成分输入、一个结构输出对应的三个范畴中,给定其中两个范畴可得出第三个范畴。先给定句子范畴 s 和主宾语范畴 np,由后向应用(<)、前向应用(>)、前向组合(>B)、后向组合(<B)可得到谓语范畴 s\.np、述语范畴 s\.np/、或语范畴 s\.np/、状语范畴 s\.np/。(s\.np)、补语范畴 s\.np\/ s\.np)、状语范畴中斜线模

态使用"◊"或"ҳ"是考虑中心成分为述语的情况。 用范畴类型逻辑^[3](CTL)的斜线引入和交换规则也 能推出上述成分的范畴。这种仅考虑结构内在得出的 范畴,可称为成分的典型范畴。若句中结构套叠都是 同类套叠,即每个基本结构的结果范畴正好是其参与 外层结构所需的范畴,各成分的范畴是固定不变的。 若存在转类套叠,即内部结构的结果范畴不是外部结 构所需的范畴,可以有两种做法:一是转变内部结构 的结果范畴为外部所需的范畴,这时各成分的范畴保 持不变,如图 2(a)中述语"提出"的范畴仍是 s\.np/.np, 述宾结构对应的范畴 s\.np 转为宾语对应的 np, 再参 与外部计算(此规则记为 O/v); 二是传统做法, 不考 虑转类和相应的范畴转换, 述语增加非典型范畴 s\.np/.(s\.np), 如图 2(b)。要列出述语或其他任一成分 的可能范畴是很难完全的, 因为要考虑它所有可能的 结合对象。图 2 中左侧数字是规则跨度,即线上范畴 数减一,范畴下方是其谓词-论元结构^[14]。



(a) 引入范畴转换后述语范畴保持不变

	马修	提出	修改	算法
0:	np :马修'	s\.np/.(s\.np) :提出'	s\.np/.np :修改'	np :算法'
1:			s\.np :修改'算法	> E'
2:		s\.np :提出'(修改'	算法')	
٦.	_			

:提出'(修改'算法')马修'

(b) 拒绝范畴转换时述语被迫采用非典型范畴 图 2 述宾结构右套叠述宾结构的两种句法分析

3 基于结构上下文的范畴转换

考虑第 1 节经转类的重写规则,将规则(1)逆置得到主谓结构转类充当主、宾语的规则(9)。

$$SP \to n/s \to S \mid O \tag{9}$$

进而得到相应的范畴转换规则(10)和(11),其中规则标记"X/x"表示语法性质 x 的结构充当句法成分 X。

$$s \rightarrow_{S/s} np$$
 (10)

$$s \rightarrow_{O/s} np$$
 (11)

将规则(3)逆置得到动词性结构转类充当主、宾 语的规则(12)。

$$DH_{v} \mid VO \mid H_{v}C \rightarrow n/v \rightarrow S \mid O$$
 (12)

在各成分只采用其典型范畴前提下,得到依赖上下文 SP 和 VO 的范畴转换规则(13)和(14),其中"\$1"为空或 np 时分别对应谓语和述语,斜线模态"."允许词范畴中斜线使用任意模态^[1]。

$$s\.np/.np s\.np/.$1 \rightarrow_{O/v} s\.np/.np np$$
 (14)

同理, 重写规则(4)应用于上下文SP和VO, 得范畴转换规则(15)和(16)。

$$np/.np \quad s \cdot .np/.\$1 \rightarrow_{S/a} np \quad s \cdot .np/.\$1$$
 (15)

$$s\.np/.np \quad np/.np \rightarrow_{O/a} s\.np/.np \quad np$$
 (16)

重写式(6)应用于上下文 SP(罕见于 VO),得 范畴转换规则(17)。

$$np \quad np/.np \rightarrow_{P/a} np \quad s \backslash .np \tag{17}$$

重写式(8)应用于上下文AHn,得转换规则(18)。

$$np np \rightarrow_{A/n} np/.np np$$
 (18)

对范畴转换(13)至(18)的右边,施用函数应 用或类型提升连用函数组合,得范畴结合规则(19) 至(24)。

$$s\.np/.$1 s\.np \rightarrow_{S/v} s$$
 (19)

$$s\.np/.np \quad s\.np/.$1 \rightarrow_{O/v} s\.np$$
 (20)

$$np/.np$$
 $s \cdot np/.$1 \rightarrow_{S/a} s/.1 (21)

$$s\.np/.np \quad np/.np \rightarrow_{O/a} s\.np$$
 (22)

$$np np/.np \rightarrow_{P/a} s (23)$$

$$np np \to_{A/n} np (24)$$

在"类有定职"、各成分只采用典型范畴的前提下,转类伴随着范畴转换。动词、形容词都不对应范畴 np, 在主宾语位置转名词才有对应的范畴 np; 形容词不对应范畴 s\.np, 在谓语位置转动词才有对应范畴 s\.np; 名词不对应范畴 np/.np, 在定语位置转形容词才有对应范畴 np/.np。

CCG 原有的函数应用、函数组合和类型提升是用

范畴变量表示的对各语言通用的范畴运算规则,规则 (19) 至(24) 是用范畴常量表示的反映汉语词组本 位的运算规则,显然这些范畴常量的结合规则不是通 用规则的实例,而是对通用规则的补充。这种允许范 畴按结构需要而转换的 CCG 可称为范畴转换式 (Category Conversional) CCG, 缩写 C²-CCG。

4 范畴转换的解释能力

给定两个词范畴, 应先尝试通用规则, 再尝试补 充规则,以尽量限制 C^2 -CCG 的解释能力。函数应用、 函数组合的六条规则有互斥性,都不可用时尝试类型 提升[15]。遇到范畴 s*np/*np 和 s*np 毗连, 规则 S/v、 O/v 都可用, 按句法歧义处理。遇到结果范畴 s, 句法 树尚未形成,使用规则 S/s 或 O/s。规则选择算法如图 3 所示。

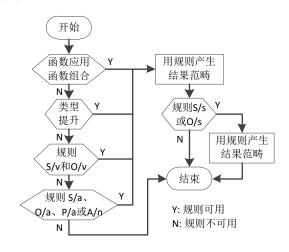


图 3 C2-CCG 中的规则选择算法

C²-CCG 和 CCG 对例句 1 "增加规则解释结构套 叠"产生相同结构的分析树,如图 4,区别是前者比 后者多用了规则 A/n、S/v 蕴含的范畴转换, 使前者的 动词"解释"和名词"结构"保持采用其典型范畴, 而后者的"解释"用非典型范畴 s\.(s\.np)/.np,"结构" 用非典型范畴 np/.np, 才能支持述宾结构做主语、名 词做定语。

	增加	规则	解释	结构	套叠				
0:			s\.np/.np :解释'	np :结构'	np :套叠' A/n				
1:	s\.np :增加'规则	IJ'	->	np :结构'套叠'					
2:									
4:									
	(a) C ² -CCG 的分析								
	增加	规则	解释	结构	套叠				
0:	s\.np/.np :增加'	np :规则'	s\.(s\.np)/.np :解释'	np/.np :结构'	np :套叠'				
1:	s\.np :增加'规则	,	•	np :结构'套叠'					
2:									
4:									
	(b) CCG 的分析								
	图 4 例句 1 的分析对比								

图 4 例句 1 的分析对比

例句 2"谦虚使人进步"中述宾结构右套叠主谓 结构。 C^2 -CCG的分析如图 5(a),内部的主谓结构先由 规则 "<"形成范畴 s, 后完成 "O/s" 蕴含的主谓结 构到名词性成分的范畴转换,跨度3形容词做主语也 发生了规则 "S/a" 蕴含的范畴转换。CCG 分析例句 2 时,述语被赋予了非典型范畴 s\.(np/*np)/.s,满足形容 词做主语、句子做宾语,如图 5(b)。

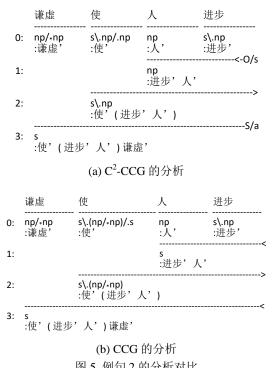
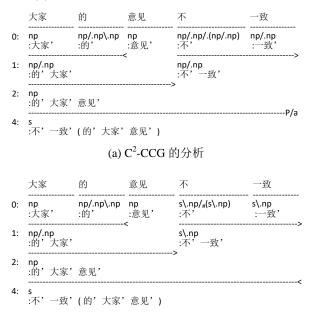


图 5 例句 2 的分析对比

例句 3 "大家的意见不一致"在 C^2 -CCG 看来是主谓结构右套叠状中结构(DH_a), DH_a 经规则"P/a" 把范畴 np/.np 转为 s/.np,再参与构成主谓结构,如图 6(a)。而在 CCG 看来,此句是主谓结构右套叠状中结构(DH_v),形容词"一致"被指派非典型范畴 s/.np,如图 6(b)。



(b) CCG 的分析 图 6 例句 3 的分析对比

综上,在运用 CCG 分析汉语句法时,采用类无定职,动词除了做述谓语,还可做主宾语,形容词除了做定语,还可做谓语、主宾语,名词除了做主宾语,还可做定语,使同一类词的范畴有许多个,确定词类所有可能的范畴并从中选择合适的一个往往比较困难。C²-CCG 采用类有定职,各类词有明确范畴,按结构要求完成范畴转换,也能生成同样结构的分析树。

5 结语

汉语句法基于词组本位,句法结构可以自由套叠。词组需要转类,才能作为整体在更大的词组里充当所需的句法成分,要求组合范畴语法增加相应的范畴转换规则。在通用规则都不可用情况下,按名-动-形依次包含允许词(组)转类,改变其句法范畴,再按通用规则尝试范畴结合,这种用范畴转换解释转类套叠的组合范畴语法(C²-CCG)暗示了没有形态曲折的转类的客观存在,使句法成分的范畴指派只考虑词组的内

部结构,这符合词组本位思想。作为一个语法系统,C²-CCG 解释转类套叠的范畴转换规则是否简约和完备?需要在一定规模的真实汉语语料下进行验证。CCG 通用规则支持范畴类型透明性,即每个句法范畴与一个逻辑式相关联,逻辑式的语义类型完全由句法范畴决定,与句法范畴的衍生过程无关,这种透明性在 C²-CCG 里是否还存在?也需要进一步研究。

参考文献

- [1] STEEDMAN M, BALDRIDGE J. Combinatory Category Grammar [M] //D BORSLEY R, BORJARS K. Non-Transformational Syntax: Formal and Explicit Models of Grammar. New Jersey: Blackwell, 2011, 181-224.
- [2] 邹崇理. 关于组合范畴语法 CCG [J]. 重庆理工大学学报 (社会科学), 2011, 25(8):1-5.
- [3] 姚从军.组合范畴语法与范畴类型逻辑的分析与比较[J]. 云南师范大学学报(哲学社会科学版),2015,47(1):75-81.
- [4] 陈鹏. 组合范畴语法(CCG)的计算语言学价值[J]. 重庆理工大学学报(社会科学), 2016, 30(8):5-11.
- [5] JAYANT K, M MITCHELL T. Joint syntactic and semantic parsing with combinatory categorial grammar[C] // Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, Pennsylvania: ACL, 2014, 1188-1198.
- [6] 赵元任. Mandarin Primer, An Intensive Course in Spoken Chinese[M]. Cambridge, Massachusetts: Harvard University Press, 1948.
- [7] 丁声树, 吕叔湘, 李荣等. 现代汉语语法讲话[M]. 北京: 商务印书馆, 1961.
- [8] 朱德熙. 语法讲义[M]. 北京: 商务印书馆, 1982.
- [9] 陆俭明. 汉语句法成分特有的套叠现象[J]. 中国语文, 1990, (2):81-90.
- [10] 刘乃仲. 汉语句法成分的"错叠"与"套叠"[J]. 沈阳师 范学院学报(社会科学版),1997,21(2):46-49.
- [11] 张伯江. 汉语的句法结构和语用结构[J]. 汉语学习, 2011, (2):3-12.
- [12] 沈家煊. 我看汉语的词类[J]. 语言科学, 2009, 8(1):1-12.
- [13] 张斌. 现代汉语语法十讲[M]. 上海:复旦大学出版社, 2005.
- [14] WANG Q, WANG Z, ZHANG L. Syntax-semantic synchronous parsing of Combinatory Categorial Grammar[J]. ICIC Express Letters Part B: Applications, 2015, 6(6):1499-1504.

[15] WANG Q, ZHANG L. A novel CCG parser based on operating properties of combinatory rules[J]. ICIC Express Letters Part B: Applications, 2014, 5(5):831-836.



王庆江(1968一),通信作者,博士,副教授,主要研究领域为自然语言处理,特别是中文句法和语义的形式化。

E-mail: wangqingjiang@ncwu.edu.cn



张琳(1969一),硕士,副教授,主要研究领域为中文 信息处理、问题-回答系统、机器翻译等。

E-mail: zhanglin@huel.edu.cn