

组合范畴语法产生和发展的 动因探析*

——从句法演算的视角看

姚从军 邹崇理

【摘要】AB-演算刻画句法推演能力太弱,许多合语法的语句在AB-演算中没有合适的推演形式。斯蒂德曼在AB-演算的基础上增加组合规则,产生了原生态的组合范畴语法CCG,大大增强了范畴语法处理自然语言的能力,比如它能够处理像非成分毗连这样的非连续问题。但是对于话题句、尤其是汉语中存在的各类话题句,原生态CCG显得无能为力,因此需要对原生态CCG进行扩展。在原生态CCG中添加话题化规则和话题化范畴,可以有效地处理汉语的各类话题化语句。

【关键词】AB-演算 组合范畴语法 不连续成分 话题句

【中图分类号】B813 【文献标识码】A 【文章编号】1000-2952(2014)01-0040-05

组合范畴语法CCG是斯蒂德曼在AB-演算基础上增加与组合逻辑组合子B、T、S相对应的组合规则而产生的范畴语法,它是上下文敏感的范畴语法,比早期范畴语法的推理能力大大增强。CCG的句法和语义之间有一个非常透明、直截了当的接口,语义表达式和句法范畴一同被存放在词库的词条上,是一种完全词汇主义的处理方式。CCG表现的现象比词汇功能语法和中心语驱动语法等语法表现的现象要深,它可以更快地评价语句中的语词是否有依存关系、谓词-论元关系等,而且速度最快。从信息计算的角度看,在使用各种语法理论支持的计算机分析器中,组合范畴语法支持的分析器分析最快,准确性也很高。

一、AB-演算

AB-演算是所有形式主义范畴语法的共同起点,它是巴-希勒尔(Bar-Hillel)对爱裘凯维茨(Ajdukiewicz)的句法演算(Ajdukiewicz)进行方向性修改的产物,一般被称为AB-演算。^①AB-演算的语法对象

是范畴。

1.1 范畴

定义1.1(AB-范畴) 已知A是一个有穷的原子范畴集,范畴集C是一个满足下列条件的最小集:

- $A \subseteq C$
- $(X \setminus Y), (X/Y) \in C$, 如果 $X, Y \in C$

使用斯蒂德曼主目靠右的范畴记法,用算子 \setminus 把词汇与它的范畴连接起来形成词条(本文我们不给出语义

* 基金项目:国家社科基金重大项目“自然语言信息处理的逻辑语义学研究”(10&ZD073)成果;湖南省社科基金项目“组合范畴语法与汉语处理”(12YBA150)研究成果;中国博士后科学基金(2012M52038)资助,湖南科技学院“十二五”重点建设学科“汉语言文字学”资助;湖南省重点建设学科“思想政治教育”项目资助,项目号湘教发[2011]76号。

① Bar-Hillel, Yehoshua. *A Quasi-arithmetical Notation for Syntactic Description*. Language 29, 1953: 47-58.

表达式)。如下给出一些简单词条例子：

- (1) a. 中国 \vdash np;
b. 日本 \vdash np;
c. 打败 \vdash (s \ np) / np。

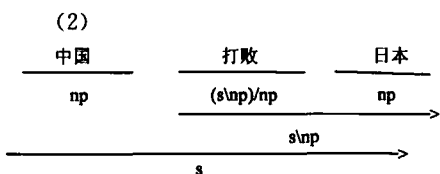
1.2 规则

在 AB-演算中，范畴通过如下两个对方向敏感的函项应用规则进行运算。

函项应用规则

- a. $X/Y \quad Y \Rightarrow X \quad (>)$
b. $Y \quad X \setminus Y \Rightarrow X \quad (<)$

在给出一个表达式的推演时，我们先把它的词汇按固有顺序依次列出，同时从词库中引入它们的范畴，然后把规则应用到毗连的成对范畴上。通过加下划线及给下划线标注代表规则的符号来表述每一步推演。不带标记的下划线表示从词库中引入一个范畴。使用 (1) 中的词条，下面给出一个语句的推演实例：



范畴语法类型驱动的本质为基本的并列组合现象提供了一个自然的解释。假定并列联结词（和、且、或……）的范畴是从两个相同类型的范畴映射到另一个同样类型的范畴的函项，可以模式化地描述为 (3)，我们可立即推出 (4) 中并列组合成分的分布。

(3) 和 \vdash (X \ X) / X

- (4) a. 中国反抗并打败日本。
b. * 中国反抗日本并打败。
c. 中国打败了韩国并赢了世界杯。
d. 刘国梁打败瓦尔德内尔并高呼。
e. 张莺带来并给我一本书。
f. 中国打败日本和俄国。

我们约定：如果一个推演的大部分步骤是明显的，仅仅需要强调一个关键步，就把与该关键步相关的词汇放在一组，并且显示该组词汇或词汇串的范畴，省略非关键的或与具体推演目的不相关的步骤。例如，语句 (4e) 的推演可以简要描述如下：

(5) 张莺 [带来]_{(s \ np) / np} 并 [给了我]_{(s \ np) / np} 一本书。
(4e) 展示了一个双宾语动词吸收了它的间接宾语后与一个及物动词并列组合。

另一方面，(4b) 被自动阻止，因为“反抗日本”具有不及物动词范畴，它不能和及物动词“打败”的范畴进行并列组合：

(6) * 中国 [反抗日本]_(s \ np) 并且 [打败]_{(s \ np) / np}。

1.3 AB-演算处理汉语的局限性

尽管使用 AB-演算可以处理大量的语言现象，但是由于它使用的规则不具有结合和交换的性质，因此无法处理汉语中模态词与及物动词毗连、副词与及物动词毗连、主谓毗连、主语和状语毗连、直接宾语和间接宾语毗连及宾语与后置状语毗连等现象，这些毗连是分析汉语的非成分并列组合结构不可缺少的中间环节，主谓毗连还是分析汉语宾语提取“的”字结构不可逾越的阶段；AB-演算也无法实现及物动词和助词的毗连推演。

考虑下面的语句：

(7) 大学生应该珍惜并且合理利用时间。

在标准分析中，模态词“应该”和副词“合理”被指派从不及物动词范畴到不及物动词范畴的函项范畴 (s \ np) / (s \ np)。然而，为了能对 (7) 实施并列组合，在及物动词“珍惜”和“利用”吸收了它们的宾语之前，需要分别与“应该”和“合理”组合。而使用 AB-演算规则无法执行这个组合，简约表述如下：

(8) 大学生 [应该]_{(s \ np) / (s \ np)} [珍惜]_{(s \ np) / np} 并且 [合理]_{(s \ np) / (s \ np)} [利用]_{(s \ np) / np} 时间。

再看下面的宾语提取表达式及其成分的范畴指派：

(9) [学生]_{np} [回答]_{(s \ np) / np} [的]_{(np / np) \ (s / np)} [问题]_{np}。

完成这个推演需要主语与谓语毗连的规则，AB-演算对此无能为力。又如下面的我们称之为汉语的右节点提升语句，完成这个推演涉及到主谓合并形成的非成分的并列组合：

(10) [小昆]_{np} [爱]_{(s \ np) / np} 而 [小东]_{np} [不爱]_{(s \ np) / np} 小华。

使用 AB-演算无法给出主语和状语毗连的推演形式：

(11) [小海]_{np} [经常]_{(s \ np) / (s \ np)} 而 [小虎]_{np} [很少]_{(s \ np) / (s \ np)} 喝酒。

使用 AB-演算无法进行直接宾语和间接宾语的毗连推演：

(12) 妈妈给了 [女儿]_{np} [一块糖]_{np} 和 [儿子]_{np} [一块饼干]_{np}。

使用 AB-演算无法进行宾语和后置状语的毗连推演：

(13) 小红喝 [红牛]_{np} [很快]_{(s \ np) \ (s \ np)} 而 [咖啡]_{np} [很慢]_{(s \ np) \ (s \ np)}。

使用 AB-演算也无法解释下面的我们称之为汉语的重型 NP-移位现象，即宾语离开了它的典型位置，解释此类现象需要实现动词与助词的毗连推演：

(14) 张三 [吃]_{(s \ np) / np} [了]_{(s \ np) \ (s \ np)} 饭。

可见，AB-演算的一个致命弱点就是生成能力太弱，其根源在于缺乏刻画自然语言结合性与交换性的规则。对 AB-演算进行扩展，就需要增加刻画结合性与交换性的规则。

二、原生态组合范畴语法 CCG

斯蒂德曼基于组合逻辑的三个组合子（复合组合子 B、类型提升组合子 T 和置换组合子 S）定义新规则，把这些新规则增加到 AB-演算中，得到了原生态组合范畴语法 CCG。^① 原生态 CCG 有效地解决了上面提到的 AB-演算无法处理的各类非成分毗连现象。

2.1 CCG 的和谐复合规则

汉语语法在一些层面具有结合性。特别地，基于复合动词的并列组合和宾语提取的分析，表明新的语法系统有能力推演非标准成分，这就需要配置刻画结合性的新规则，接下来的两条规则就是满足这一要求的基于复合组合子 B 的规则：

向前和谐复合规则 (>B)

$$X/Y \quad Y/Z \Rightarrow_B X/Z$$

有了这个规则，模态词“应该”的范畴 (s \ np) / (s \ np) 就可以和及物动词“珍惜”的范畴 (s \ np) / np 进行复合运算，得到“应该珍惜”的范畴 (s \ np) / np，同理得到“合理利用”的范畴 (s \ np) / np；再根据并列联结词的范畴指派，可得到语句 (7) 的成分“应该珍惜并合理使用”的范畴为 (s \ np) / np；进一步由向前和向后的函项应用规则，便可以完成整个语句 (7) 的句法推演：

(15) 大学生 [应该珍惜]_{(s \ np) / np} 并 [合理使用]_{(s \ np) / np} 时间。

向后和谐复合规则 (<B)

$$Y \setminus Z \quad X \setminus Y \Rightarrow_B X \setminus Z$$

在汉语中，很难提供一个使用向后和谐复合规则进行运算的简单例子，在后面直接宾语和间接宾语的毗连分析及对宾语和后置状语的毗连分析中，可以看到这个规则的使用。在考虑基于 B 诱出的刻画交换性的规则之前，我们先转向类型提升规则，它们与和谐复合规则一起进一步扩张了 CCG 的结合性。

2.2 CCG 的类型提升规则

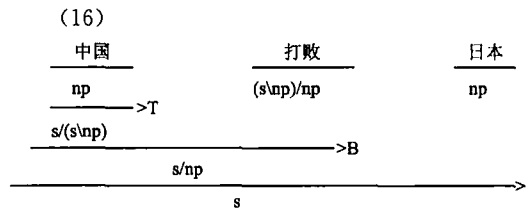
基于组合子 T 的类型提升规则把一个较简单的范畴转换成函项范畴，它包括向前类型提升规则和向后类型提升规则。

向前类型提升规则 (>)

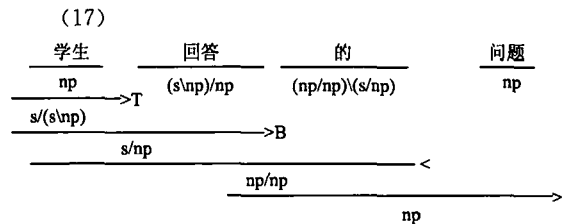
$$X \Rightarrow_T Y / (Y \setminus X)$$

这个规则可使原子范畴变成一个函项范畴，它的一

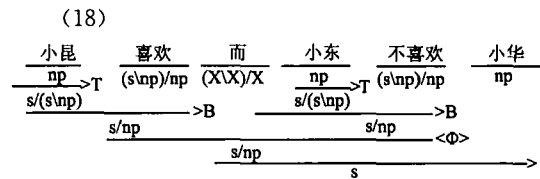
个直接作用是与向前和谐复合规则一起实现主谓毗连推演，从而诱导出汉语及物句的结合性。推演 (2) 有下面的一个择换推演：



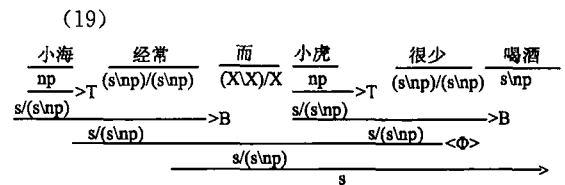
有了向前和谐复合规则和向前类型提升的规则，CCG 就有能力直接解释汉语中的宾语提取现象。例如，宾语提取表达式 (9) 的推演形式如下：



现在也可以给出 (10) 的推演：



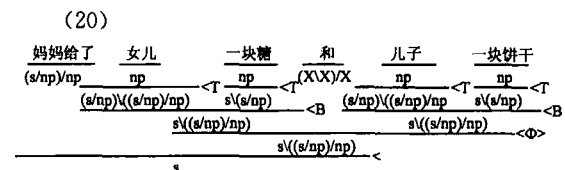
同样也可以给出主语和状语的毗连推演形式，(11) 的推演形式如下：



向后类型提升规则 (<T)

$$X \Rightarrow_T Y \setminus (Y / X)$$

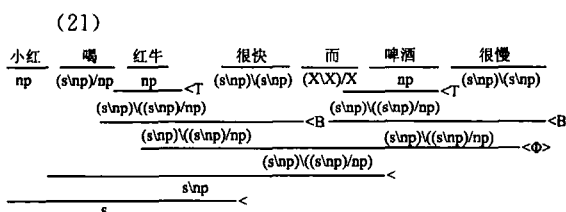
有了向后和谐复合规则和向后类型提升的规则，CCG 就有能力实现双宾语的毗连推演，(12) 的推演如下：



使用这两个规则也可以实现宾语和后置状语的毗连

^① Steedman, Mark. *Surface Structure and Interpretation*. Cambridge Mass: MIT Press. Linguistic Inquiry Monograph, 1996: 30-43.

推演，(13)的推演如下：



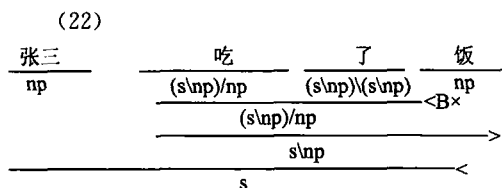
2.3 CCG 的交叉复合规则

至今所讨论的规则都是保序的。但是自然语言的某些语言成分在语句中可以离开它们的典范位置，为刻画这一现象，CCG 引入了刻画自然语言交换性的交叉复合规则。

向后交叉复合规则 ($<B_x$)

$$Y/Z \quad X \setminus Y \Rightarrow_{B_x} X/Z$$

现在可以实现及物动词和助词的毗连推演，语句(14)中的“吃”在吸收它的宾语“饭”之前先与助词“了”毗连，其推演如下：



向前的交叉复合规则 ($>B_x$)

$$X/Y \quad Y \setminus Z \Rightarrow_{B_x} X \setminus Z$$

通常认为向前交叉复合规则在英语语法中是无效的，因为它会导致不合语法的爬升词序。不过，它必须是通用语法一部分，因为它对于分析像土耳其语这样的有更大词序自由度的语言的爬升现象非常重要。目前还没找到该规则在汉语分析中的应用实例。

2.4 CCG 的置换规则

置换规则基于组合子 S，它包括和谐置换规则和交叉置换规则。

向前和谐置换规则 ($>S$)

$$(X/Y) / Z \quad Y/Z \Rightarrow_s X/Z$$

向后交叉置换规则 ($<S_x$)

$$Y/Z \quad (X \setminus Y) / Z \Rightarrow_{S_x} X/Z$$

这两个规则是处理英语的寄生缺口结构所需要的，但是汉语语法不会使用这样的规则，因为英语的寄生缺口表达式译成汉语时会改变词性或顺序。

向后和谐置换规则 ($<S$)

$$Y \setminus Z \quad (X \setminus Y) \setminus Z \Rightarrow_s X \setminus Z$$

向前交叉置换规则 ($>S_x$)

$$(X/Y) \setminus Z \quad Y \setminus Z \Rightarrow_{S_x} X \setminus Z$$

这两个规则绝不会被英语语法使用，因为缺乏用作输入的词汇范畴。然而，可以用它们分析其他语言的寄

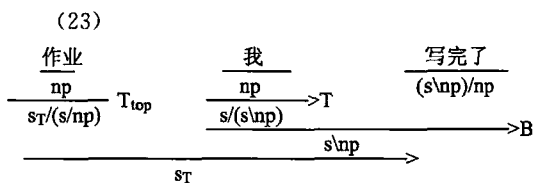
生缺口结构。^① 目前也没找到它们在汉语分析中的应用实例。

三、话题句与原生态 CCG 的扩展

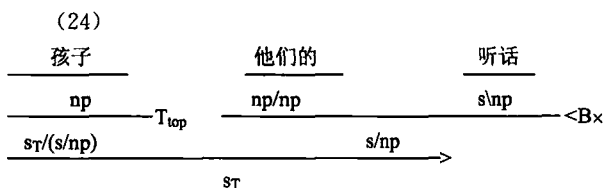
名词短语可以充当普通汉语句子的主语和宾语，也可以充当汉语话题句的话题成分。在充当汉语句子的主语和宾语成分时，我们为之指派范畴 np，必要时再利用向前和向后类型提升规则将其范畴转换成 s/ (s \ np) 和 s \ (s/ np)，便可以完成普通汉语句子的推演。但是，如果给充当话题成分的名词短语指派范畴 np，不管使用原生态 CCG 的什么规则，几乎无法给出相应语句的正确推理形式。为处理无占位代词的宾语提取话题化语句的推演，斯蒂德曼提出了一种句法操作规则，他称之为话题化规则，表达如下：

$$T_{top}: X \rightarrow S_T / (S/X).$$

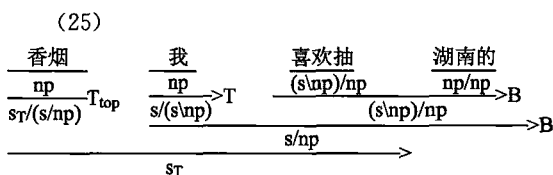
有了这个表达式，我们就可以实现如下推演：



可以检验，这条规则可以用来处理复杂主语中心成分话题化语句的推演：



这条规则可以用来处理复杂宾语中心成分话题化语句的推演：



也可以检验，这条规则还可以用来处理像“五块钱我给那个乞丐了”这样的无占位代词的直接宾语提取话题化语句的推演和像“那个乞丐我给了五块钱”这样的无占位代词的间接宾语提取话题化语句的推演，具体的

^① Baldridge, J. *Lexically Specified Derivational Control in Combinatory Categorical Grammar*. PhD. thesis. University of Edinburgh, 2002: 88.

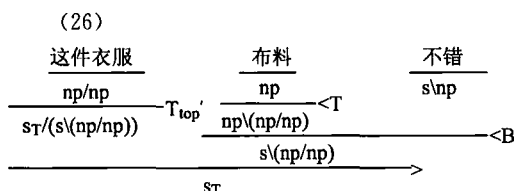
^② Steedman, Mark. *The Syntactic Process*. Cambridge Mass: The MIT Press, 2000: 215.

推演略。

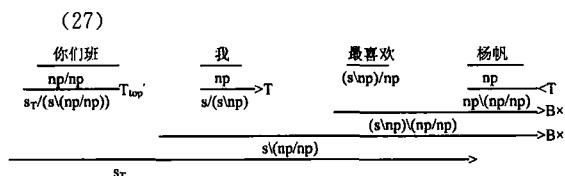
斯蒂德曼的处理方法很具有启发性，但是我们认为对汉语来说不太令人满意：因为汉语话题化语句有好几种类型，此条规则对于上面几种话题句的推演是适用的，对其他类型的话题句未必适用。汉语中还有像“这件衣服布料不错”这样的复杂主语的修饰成分话题化语句、像“你们班我最喜欢杨凡”这样的复杂宾语修饰成分话题化语句，以及像“软件我是门外汉”这样的斜格 (oblique) 成分 (除主宾语之外的其他名词短语成分) 话题化语句，对于这样的话题句，上面的话题化规则无能为力，为此我们引入上述话题化规则的变体如下：

$$T_{top}': X \rightarrow S_T / (S \setminus X).$$

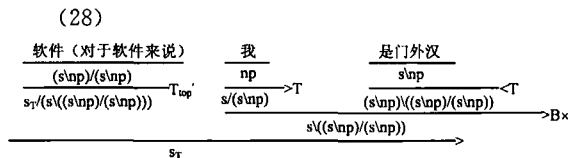
有了这条话题化规则，就可以处理复杂主语修饰成分话题化语句的推演：



可以检验，这条规则可以用来处理复杂宾语修饰成分话题化语句的推演：



这条规则还可以用来处理斜格成分话题化语句的推演：



汉语中还有像“小龙女杨过爱她”这样的有占位代词的宾语提取话题化语句，像“雷锋他将永垂不朽”这样的有占位代词的主语提取话题化语句，以及像“那个乞丐我给他五块钱”这样的有占位代词的间接宾语提取话题化语句。对于这样的话题句来说，上面的两个话题化规则都无能为力，增加新的话题化规则 T_{top}'' ： $X \rightarrow S_T / X$ 似乎很是不妥，因为话题成分在基本语句中是主宾语，给它们指派范畴 s (再根据 T_{top}'' 提升为 s_T/s) 与直观相差甚远，倒不如根据推理需要直接在词库中给话题成分指派话题范畴 s_T/s ，这里的具体推理略。

组合范畴语法及其扩展也可以用来处理汉语的“把”字句、“被”动句、焦点句、兼语句、连动句等，对于它们的处理另文再论。随着组合范畴语法广泛应用于汉语分析中，必将进一步推动组合范畴语法的完善和发展，大大提高计算机对汉语的处理能力和效率。

本文作者：姚从军是中国社会科学院哲学所博士后流动站研究人员，南开大学哲学博士，湖南科技学院副教授，湘潭大学硕士生导师；邹崇理是中国社会科学院哲学所研究员，博士生导师，中国逻辑学会会长
责任编辑：周勤勤

The Motivational Analysis about the Emergence and Development of Combinatory Categorical Grammar

—From the Perspective of Syntactic Calculus

Yao Congjun Zou Chongli

Abstract: AB-calculus is too weak, and many correct sentences which present to be appropriate in grammar rule can not be derived with it. Steedman added the combinatory rules to the AB-calculus to invent the standard: Combinatory Categorical Grammar named CCG, and CCG greatly proved its processing ability to Chinese. For example, CCG can deal with discontinuous problems such as the coordination of non-constituent. For topicalised sentences, especially for Chinese topicalised sentences, the standard CCG seems too weak to deal with them, so we need to extend it. The author adds some topicalised rules and topicalised Categories to the standard CCG, and efficiently deals with all kinds of Chinese topicalised sentences.

Keywords: AB-calculus; combinatory categorical grammar; discontinuous constituent; topical sentences