

doi:10.13582/j.cnki.1672-7835.2022.01.006

语言、逻辑与计算互动视角下 汉语直接被动句的 MMCCG 处理

姚从军, 俎孟晨

(湘潭大学 碧泉书院, 湖南 湘潭 411105)

摘要:汉语被动句有直接和间接之分,也有长短之别,如何处理汉语被动句历来是计算语言学家需要攻克的难题。汉语直接被动句的显著特征是主谓语的受事指向主句主语,传统分析方法使用移位、删除等句法操作分析汉语被动句,否定了被动句的原生成性,也出现了许多难以解释的现象。多模态组合范畴语法 MMCCG 抛弃了移位和删除等句法操作,以表层组合的方式生成各种被动句的句法-语义推演树,避免了传统方法分析被动句时出现的所有问题,并从计算语言学视角证明了被动句的原生成性。

关键词:被动句;多模态组合范畴语法;广义斯科伦项

中图分类号:B81-06 **文献标志码:**A **文章编号:**1672-7835(2022)01-0042-09

一 汉语被动句的界定及来源

对于汉语被动句的定义素有争议,洪心衡认为,带被动标记词“被”的句子才是被动句,而没有标记词“被”的句子仅仅表述论断事实,并没有表示被动的语义^①。然而,王力认为,无被动标记词“被”的句子也可以是被动句,只要它有被动的含义;从语义角度来看,只要主语在句子中表现为受事,这个句子就是被动句,哪怕“被”“让”等被动标记词不出现在句子中(比如“花盆摔碎了”)^②。从语义出发界定被动句,无论句子是否包含被动标记词,只要包含被动语义,就可断定为被动句。这样看来,汉语被动句既包括带被动标记词“被”的句子,比如“我被人打了”;也包括带“让”“给”等被动标记词的句子,比如“我让人打了”和“我给人打了”等等;还包括一些不带被动标记词的句子,比如“球弄丢了”“衣服洗干净了”和“饭做了”等等。依据不同的视角,“汉语被动句”这个概念也就有不同的内涵和外延。本文只处理带被动标记词“被”的被动句。

(1)a.王五被领导骂了。b.王五被骂了。

这两个句子都含有被动标记词“被”。在 a 中,“被”右有一个名词短语 NP,这样的被动句被称为长被动句。在 b 中,“被”右没有名词短语 NP,这样的被动句被称为短被动句。这两个被动句有一个共同特点:主谓语的受事指向主句主语,如上述主句主语“王五”所指的是“骂”的受事,这样的被动句被称作“直接被动句”。与直接被动句对应的是“间接被动句”,其中主句主语可能与主要动词的直接宾语具有非同指的其他关系(如“张三被李四抓住了一只手”“张三被抓住了一只手”),也可能与“被”后任何句法位置都无关(如“张三被李四自摸了”“张三被自摸了”),前者被称作“包括式间接被动句”,后者被称作“排除式间接被动句”,二者也都有长短被动句之分。本文只处理直接被动句,对于间接被动句,另撰文再论。

关于汉语被动句的来源问题,也是见仁见智。劲松认为,从句法角度来说,被动句从主动句转化

收稿日期:2021-07-22

基金项目:国家社会科学基金重大项目(17ZDA027)

作者简介:姚从军(1971—),男,湖北随州人,教授,博士生导师,主要从事现代逻辑和语言逻辑研究。

①洪心衡:《汉语语法问题研究》,新知识出版社1956年版,第98—108页。

②王力:《现代语法》,商务印书馆1985年版,第92—97页。

而来^①。笔者认为,汉语长被动句是原生成的,而不是由主动句转换而来的;汉语短被动句也是原生成的,而不是由长被动句转换而来的,它比长被动句有更长的历史起源。多模态组合范畴语法 MMCCG 是单层结构语法,它抛弃了移位和删除等句法操作,经由表层成分毗连组合得到各种被动句的句法-语义生成树,从计算语言学视角证明了被动句的原生成性。

二 多模态组合范畴语法 MMCCG

组合范畴语法 CCG 源于范畴语法 CG。CG 的基本做法是从不同层次的语言表达式中抽象出句法范畴,再通过范畴的运算来刻画语言表达式由小到大逐层逐级地毗连生成过程。CG 仅含两条规则,对自然语言的描述能力很有限。往 CG 中添加三类组合规则,产生了 CCG。CCG 虽然能够处理 CG 无法处理的一些自然语言语句,但也能对一些不合语法的自然语言语句构造推演树,应对其进行限制。多模态组合范畴语法 MMCCG 引入了定义斜线类型的模态词 \star 、 \diamond 、 \times 、 \cdot ,用类型化斜线限制规则的适用范围,使范畴语法走上了更加彻底的词汇主义道路,并且避免了 CCG 导致的过度生成性。

(一) MMCCG 的规则

(2) 词法规则(类型提升规则):

$$a. X \$: a \Rightarrow_{\text{LEX}} T /_i (T \setminus X) \$: \lambda f. fa; b. X \$: a \Rightarrow_{\text{LEX}} T \setminus_i (T /_i X) \$: \lambda f. fa$$

(3) 句法规则一(函项应用规则):

$$a. X /_{\star} Y : f \quad Y : a \Rightarrow X : fa (>); b. Y : a \quad X \setminus_{\star} Y : f \Rightarrow X : fa (<)$$

(4) 句法规则二(函项复合规则):

$$a. X /_{\diamond} Y : f \quad Y /_i Z : g \Rightarrow X /_i Z : \lambda z. f(gz) (>B)$$

$$b. Y \setminus_i Z : g \quad X \setminus_{\diamond} Y : f \Rightarrow X \setminus_i Z : \lambda z. f(gz) (<B)$$

$$c. X /_{\times} Y : f \quad Y \setminus_i Z : g \Rightarrow X \setminus_i Z : \lambda z. f(gz) (>B_{\times})$$

$$d. Y /_i Z : g \quad X \setminus_{\times} Y : f \Rightarrow X /_i Z : \lambda z. f(gz) (<B_{\times})$$

(5) 句法规则三(函项置换规则):

$$a. (X /_{\diamond} Y) /_i Z : f \quad Y /_i Z : g \Rightarrow X /_i Z : \lambda z. f z(gz) (>S)$$

$$b. Y \setminus_i Z : g \quad (X \setminus_{\diamond} Y) \setminus_i Z : f \Rightarrow X \setminus_i Z : \lambda z. f z(gz) (<S)$$

$$c. (X /_{\times} Y) \setminus_i Z : f \quad Y \setminus_i Z : g \Rightarrow X \setminus_i Z : \lambda z. f z(gz) (>S_{\times})$$

$$d. Y /_i Z : g \quad (X \setminus_{\times} Y) /_i Z : f \Rightarrow X /_i Z : \lambda z. f z(gz) (<S_{\times})$$

斜线下标 i 是模态算子变元,可以取 $\{\star, \diamond, \times, \cdot\}$ 中任何一算子为值。

(二) MMCCG 的语义理论:广义斯科伦项理论

一阶逻辑和 λ -演算结合产生了广义量词理论,CCG 使用广义量词理论表述自然语言的语义。在该语义理论中,因量词和全称量词之间存在复杂的语义依存关系而出现了辖域倒装、辖域非对称、伪歧义、伪解释和中间辖域等语义问题。比如,使用广义量词理论可得到“每个男孩喜欢一个女孩”的两个语义表达式: $\forall y (By \rightarrow \exists x (Gx \wedge Lxy))$; $\exists x (Gx \wedge \forall y (By \rightarrow Lxy))$ ^②。前者是存在量词取窄辖域的语义表达式,表示“每个男孩可能喜欢一个不同的女孩”;后者是存在量词取宽辖域的语义表达式,表示“每个男孩喜欢同一个女孩”,于是产生了辖域倒装问题。要解决这些问题,必须简化全称量词和存在量词之间的依存关系,Skolem 提出了斯科伦项理论:删除公式中的存在量词,用斯科伦函项 $sk(x, y, z \dots)$ 或斯科伦常项 $sk()$ 取代该存在量词约束的变元^③。斯科伦项的定义域由管辖该存在量词的全称量词决定:若存在量词 $\exists x$ 处于 $\forall y$ 的辖域内,则相应的斯科伦函项为 $sk(y)$;若 $\exists x$ 处于 $\forall y$ 和 $\forall z$ 的辖域内,则相应的斯科伦函项为 $sk(y, z) \dots$;若 $\exists x$ 不处于任何全称量词的辖域内,得到斯科伦常项 $sk()$ 。于是,前述两个语义表达式分别转换为 $\forall y (By \rightarrow (Gsk(y) \wedge Lsk(y) y))$ 和 $\forall y (By \rightarrow (Gsk() \wedge Lsk() y))$ 。

有了斯科伦项,不需要存在量词也可得到存在量词取宽、窄辖域两种解释,同时避免了广义量词理论中存在量词取宽辖域解释时出现的辖域倒装现象。但是,上述公式含有一个独立的谓词 G ,因此与语句表层语序并非完全一致。Steedman 把斯科伦项方法往前推进了一大步,引入了形如 $sk_{n,p,c}^E$ 的广义斯科伦项,其中 E 是环境,环境中的元

①劲松:《被动句的偏误和规范》,《汉语学习》2004年第1期。

②谓词 B 表示“是男孩”, G 表示“是女孩”, L 表示“喜欢”(Lxy 表示“ y 喜欢 x ”)。

③Robaldo L. *Skolem Theory and Generalized Quantifiers*. International Workshop on Logic. Berlin: Springer-Verlag, 2008, pp. 286-297.

素是管辖广义斯科伦项的全称量词约束的变元^①; p 是使用 λ -公式表述的名词性质, 比如“女生”的逻辑性质 $\lambda x.Gx$; n 是名词短语出现的编号, 目的是区分具有相同性质的不同个体, 比如在“三个男生走路, 三个男生坐车”中, 前后出现的“男生”要用带不同编号的广义斯科伦项区分开来(不用区分时, n 可省略); 基数条件 c 表示像“三个”“至多两个”和“大多数”这样的数量(基数为 1 时, c 可省略)^②。该语句开始出现的“三个男生”就可表述为 $sk_{1,x.Bx,3}^{\{1\}}$, 因为该广义斯科伦项没有出现在任何全称量词的辖域内, 如果环境 E 是空集, 这样的广义斯科伦项被称作广义斯科伦常项; 如果环境非空, 则得到广义斯科伦函项。广义斯科伦项 $sk_{n,p;c}^E$ 指称语义类型为 e 、性质为 p 的个体。众所周知, 就辖域交替而言, 包括复数在内的非全称量词与全称量词的表现非常不同, 在处理方式上应该区分开来。Steedman 继续用广义量词表示全称量词, 而用广义斯科伦项表示所有非全称量词, 形成了广义斯科伦项理论, 并把它作为 MMCCG 的语义理论, 这一做法不仅更准确地刻画了自然语言的量化现象, 解决量词辖域引起的各种语义问题, 还可以处理 CCG 无法处理的一些话题结构, 实现代词与其先行语之间的语义同指关系、所有者对中心语的领有关系, 等等。在 MMCCG 中, “每个男孩喜欢一个女孩”的两种解读就表示为: $\forall y (By \rightarrow Lsk_{x,Gx}^{\{y\}} y)$ 、 $\forall y (By \rightarrow Lsk_{x,Gx}^{\{1\}} y)$ 。 $sk_{x,Gx}^{\{1\}}$ 和 $sk_{x,Gx}^{\{y\}}$ 均表示具有性质 G (是女孩) 的个体, 前者指称一个确定的女孩, 后者的指称可以随着 y 的变化而变化, 两个语义表达式也没有使用独立的谓词 G 来表述广义斯科伦项所指对象的性质, 因为广义斯科伦项本身就是一个语义元素, 一开始就与名词的性质相联系, 这些性质以下标形式(比如 $sk_{x,Gx}^{\{y\}}$ 中的 Gx) 出现在广义斯科伦项中, 是广义斯科伦项的一部分。

三 MMCCG 对汉语长被动句的处理

(一) 汉语长被动句的生成及处理之辩

自从生成语法诞生, 在汉语长被动句的研究

中逐渐产生了两种对立的观点。Wang 认为, 汉语长被动句也是通过 NP 移位而产生, 一旦出现词汇“被”, 底层的宾语就会移到表层的主语位置, 这就是移位分析法^③。另一观点认为, “被”是带一个嵌入子句的动词, 该嵌入子句的宾语因与主句的主语相同而被强制删除, 成为 VP 补足语, 这就是补足语分析法。

NP 移位理论认为, “被”拥有抑制主语论元和吸收宾格的能力。一旦出现“被”字, 主句动词的主语位置就会出现空论旨角色, 主句动词的宾格也会消失。这样, 因为原宾语位置缺乏格位, 原宾语被迫移位到没有论旨角色的主语位置, 施事 NP 变成了介词短语 PP 的一个组成部分, 于是形成了被动句。根据此理论, 主动句“领导骂了王五”变为被动句“王五被领导骂了”的过程如图 1 所示(IP 是 Inflectional Phrase 的简称, 表示时态语句)。

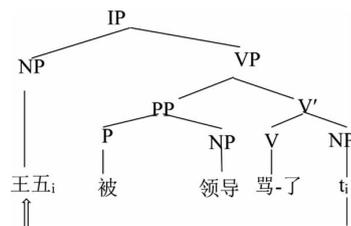


图 1 移位假说下汉语长被动句的形成图示

NP 移位理论面临一系列挑战。首先, NP 移位理论认为, 被动句底层宾语因受到“被”的影响而移到主语位置, 实际上它仍然扮演受事角色, 被动句的主语位置是非论旨位置。然而, 合法语句“小明故意被老师打了”带有主语取向性副词“故意”, 只有施事或历事 NP 方可接受“故意”修饰, 因此“故意”修饰的“小明”必须扮演施事或历事论旨角色。这样, “被”字句的主语只能是原生成的, 并接受主语应具有论旨角色, 这就与移位理论相矛盾。其次, NP 移位理论认为, “被 NP”是附加语 PP。然而, “被 NP”与 PP 区别很大。比如, PP 在句中的移动比较自由, 而“被 NP”就不能在句中自由移动。再者, 照应词的约束差别也证明了“被 NP”不是 PP。反身代词“自己”有主

^①E 中元素取决于对某个非确定斯科伦项进行确定化运算时, 管辖这个非确定斯科伦项的全称量词: 假设这个非确定斯科伦项处于 $\forall y$ 的辖域内, 那么 E 中元素为 y ; 如果这个非确定斯科伦项处于 $\forall y$ 和 $\forall z$ 的辖域内, 那么 E 中元素为 y, z ; 如果这个非确定斯科伦项不处于任何全称量词的辖域内, 那么 E 为空集。

^②Mark Steedman. *Combinatory Categorical Grammar: An Introduction*. Philadelphia: The SOMESUCH Press, 2017, pp. 196-198.

^③Wang Peter C T.A. *Transformational Approach to Chinese Ba and Bei*, Doctoral dissertation. University of Texas, Austin, 1970, p.56.

语取向性,即“自己”倾向于受主语约束,而在被动句“我被妈妈关在自己的房间里”中,“我”和“妈妈”都可做“自己”的先行语,故“妈妈”应是主语,而不是“被”的宾语。最后,“被 NP”是否是一个独立成分也有疑问。比如,对于并列结构“大宝被爸爸打了两巴掌,妈妈骂了两句”而言,“被”右边的 NP 与 VP 构成了一个独立成分。

补足语分析方法认为“被”是二元动词,事件补语的宾语因与历事主语相同而被强制删除。据此,“王五被领导骂了”的结构如图 2 所示。

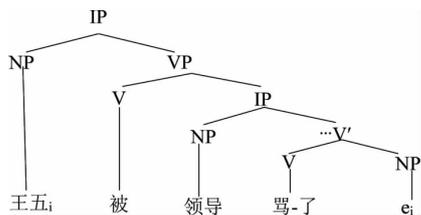


图 2 补足语分析法下汉语长被动句的结构

按照补足语分析方法,“被”有主语论元,可带主语取向性副词。“被 NP”不构成独立成分,不需要拥有 PP 的性质。在“我被妈妈关在自己的房间里”中,“我”和“妈妈”分别是主句主语和事件补语(从句)的主语,故均可约束反身代词。但是,补足语分析方法无法回答为什么要强制删除嵌入句的宾语。另外,在“王五说领导骂了他(自己)”中,嵌入句的宾语为显性代名词或照应词,为什么被动句不能如此?

MMCCG 不承认移位操作,直接否定了关于长被动句的移位生成观点。MMCCG 也不承认删除操作,认为“被”字右边的表达式主谓短语成分不涉及强制删除,与补足语分析方法迥异。MMCCG 既不把被动句之“被”字视为介词,也不视为动词,而是视为功能性成分,即被动句的标记词;“被”右边的主谓短语在整个被动句中作谓语,因此,汉语长被动句是带“被”字的主谓谓语句,带有两个主语:大主语(“被”左 NP)和小主语(“被”右 NP)。这种分析方法解决了上面提及的

两种传统分析方法带来的所有问题:

第一,补足语分析法需要回答强制删除嵌入句的宾语的原因,而 MMCCG 分析方法根本不承认删除,这就回避了补足语分析方法需要回答的问题。

第二,移位分析方法无法解释被动句主语位置的 NP 为什么能够右接主语取向性副词,而 MMCCG 分析方法坚持被动句的原生成地位,“被”左 NP 本身具有主语论旨角色,被主语取向性副词修饰乃天经地义。

第三,移位分析方法把长被动句之“被”当作介词,“被 NP”是介宾短语 PP,却不具有 PP 的一些性质。MMCCG 认为“被”是被动句指示词,“被 NP”不是独立成分,故不需要具有 PP 的性质。

第四,移位分析方法无法解释“我被妈妈关在自己的房间里”中的“自己”可回指“妈妈”,而 MMCCG 认为“妈妈”是小主语,可约束“自己”。

MMCCG 借助其自身的彻底词汇主义优势,不使用移位、删除等机制,以表层组合的方式进行句法-语义推演,在生成长被动句的句法结构树的同时,生成它的谓词-论元结构,这就从计算语言学的角度证明了长被动句是原生成的,和主动句并没有派生关系。

(二)简单的汉语有界和无界长被动句

(6)a. 小张被老师带走了。 b. 小张我听说被警察带走了。

c. 小张被校长派老师带走了。

(7)被 := (S_被 \ NP) / ★ (S / NP) : λpλy.py; 小张 := S_被 / (S_被 \ NP) : λp.p 张';

听说 := ((S_被 \ NP) \ NP) / (S_被 \ NP) : λpλvλw.听说'(pv)w;

派 := (((S \ NP) / NP) / VP) / NP : λxλpλzλy.派'x (pzx)y。

“被”词条上的斜线模态词“★”阻止“被”与右 NP 组合而成一个成分。下面给出(6)中三个语句的推演(见图 3、图 4 和图 5)。



图 3 MMCCG 对汉语有界长被动句的处理



图4 MMCCG对汉语无界长被动句的处理一



图5 MMCCG对汉语无界长被动句的处理二

此分析把“派”视作三元谓词,补语为“老师” 义被国际社会所谴责。
和“老师带走了小张”。 (9)所: = $(S\backslash NP)/(S\backslash NP): \lambda p\lambda w.pw$ 。

(三)含“所”字的汉语长被动句

(8)a.那份电报已被敌军所截获。 b.霸权主 所是。

于是,(8)中两个语句的推演如图6和图7



图6 MMCCG对含“所”字的汉语长被动句处理一



图7 MMCCG对含“所”字的汉语长被动句处理二

(四) 含代名词的汉语长被动句

(10) a. 张三被李四把他骗了。 b. 小宝被妈妈怀疑他拿了钱。

(11) 把 := (S_把 \ NP) / (S / NP) : λpλy.py;

骗 := ((S \ NP) / NP_{agr}) / NP_{pro,agr} : λyλxλa. 骗^{y}sk_a^{x};

拿 := ((S \ NP_{agr}) / NP_{pro,agr}) / NP : λx λyλa. 拿^{y}x;

怀疑 := ((S \ NP) / NP) / (S \ NP) : λpλy λx. 怀疑^{y}(py)x。

有了上述词条,便可得到(10)中两个语句的 MMCCG 树图(见图 8 和图 9)。



图 8 MMCCG 对含代名词的汉语长被动句处理一



图 9 MMCCG 对含代名词的汉语长被动句处理二

“拿”和“骗”有代名词论元,在语义表达式中使用了广义斯科伦项,从而实现了先行语与代名词之间的照应关系,扩展了 CCG 的表达力。

四 MMCCG 对汉语短被动句的处理

(一) 汉语短被动句的生成及分类

人们通常认为汉语短被动句是删除长被动句的施事 NP 而形成的,比如:

(12) a. 张三被老师骂了。 b. 张三被骂了。

历史事实表明,这一解释不正确,短被动句的出现早于长被动句。公元前 300 年的《韩非子》就使用了短被动句,而人们在公元 200 年的《汉书》中才发现长被动句的运用,比短被动句的出

现晚 500 余年。另外,汉语短被动句与长被动句有不同的句法表现。比如,在下面的语句中,副词“在学校”可以在长被动句中出现,却不能在短被动句中出现,出现了非对称性现象。

(13) a. 张三被老师在学校骂了。 b. *张三被在学校骂了。

非对称性现象还会以其他形式出现:

(14) a. 小张被校长派老师带走了。 a'. *小张被派老师带走了。

b. 霸权主义被国际社会所谴责。 b'. *霸权主义被所谴责。

c. 张三被李四把他骗了。 c'. *张三被把他骗了。

由上可知,短被动句没有无界形式、“所”字形式和复指代名词形式。如果认为短被动句是由长被动句转化而来,则无法解释上述非对称性现象。Ting 借鉴在英文 be 被动句分析中使用的论元移位法来分析汉语短被动句,认为汉语短被动句的表层主语是底层宾语移来的,扮演宾格角色^①。英语 be 被动句不允许带主语取向性副词,而汉语短被动句可带主语取向性副词(“小张故意被骂了”),故用英语 be 被动句的论元移位法分析汉语短被动句不可取。冯胜利认为,汉语被动句同句法中空算子运算有关,汉语被动句与英语“tough 移位”结构相似,可以借鉴“tough 移位”分析方法处理被动句^②。黄正德等认为,短被动句由助动词“被”和 VP 补足语组成,句中大代语宾语 PRO 因经历了 NP 移位而受历事主语控制,如图 10 所示^③。

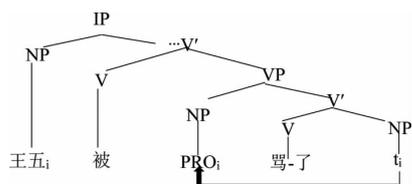


图 10 轻动词视角下的汉语短被动句生成图

MMCCG 认为,短被动句不是由长被动句转换而来,而是原生成的。短被动句的主语有“历事”角色,可带主语取向性副词。

(15) a. 王五被逮捕了。 b. 王五被捕了。

形如上述 a 的短被动句叫做短语型短被动句,可进行改写:

(16) a'. 王五被警察逮捕了。 a''. 王五被警察秘密地逮捕了。
a'''. 我们工厂被警察逮捕了两个工人。

形如上述 b 的短被动句叫做词汇型短被动句,“被”后面的动词是粘着语素,与“被”一起形成“被动动词”。两种短被动句具有不同特征,若对上述 b 也做出类似改写,则得到的语句不合法:

(17) b'. *两个间谍被警察捕了。
b''. *两个间谍被警察秘密地捕了。
b'''. *我们工厂被警察捕了两个工人。

现代汉语短被动句既有短语形式也有词汇形式,这与汉语的历史有关。在古汉语晚期,汉语具有高度单音节化特征,“被”作为被动标记词“见”的替代词开始在语言中出现。后来,汉语发展成高度双音节化语言,如果一个单音节动词变成了双音节动词,就产生了短语型短被动句;如果该动词保持单音节形式,它就要与“被”一起形成双音节“被动动词”,进而形成词汇型短被动句。

MMCCG 认为,短被动句的“被”也是被动标记词。在词汇形式短被动句中,“被”表示主句动词是“被动动词”。下面用 MMCCG 处理短被动句。

(二) 汉语短语型短被动句

(18) a. 小张被逮捕了。 b. 所有人被带走了。
c. 小张我听说被逮捕了。

(19) 被 := (S_被 \ NP) / ★ VP_被 : λpλz.pz;

逮捕了 := VP_被 : λx.逮捕'x sk_{λy.high'(prior'(逮捕'xy))}^{x}

带走了 := VP_被 : λx.带走'x sk_{λy.high'(prior'(带走'xy))}^{x}

上述词条把隐性主语表示为广义斯科伦项,由管辖动词引入,依赖于它自己 If(logic form)-统领的受事 x,指“很可能逮捕(带走)x 的任何东西”。

下面给出(18)中三个语句的 MMCCG 树图(见图 11、图 12 和图 13)。

MMCCG 还会阻止不合法的短语型短被动句,如图 14—图 16 所示。

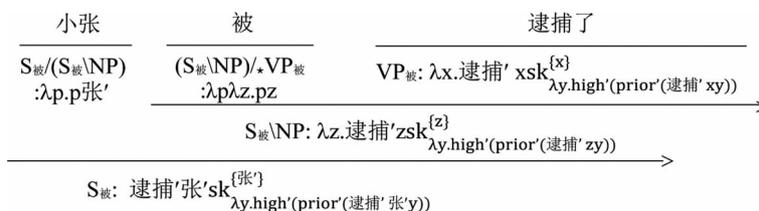


图 11 MMCCG 处理有界的汉语短语型短被动句一

①Jen Ting. *A Non-uniform Analysis of the Passive Construction in Mandarin Chinese*, Doctoral Dissertation, University of Rochester, 1995, p. 89.

②冯胜利:《汉语的韵律、词法与句法》,北京大学出版社 1997 年版,第 1—27 页。

③Huang James C T. “lexical structure and syntactic projection”. *Chinese Languages and Linguistics*, 1997, 10(3):45-89.

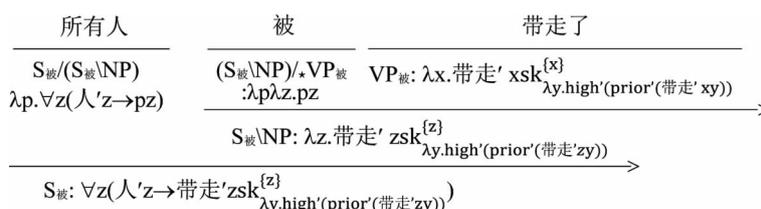


图 12 MMCCG 处理有界的汉语短语型短被动句二

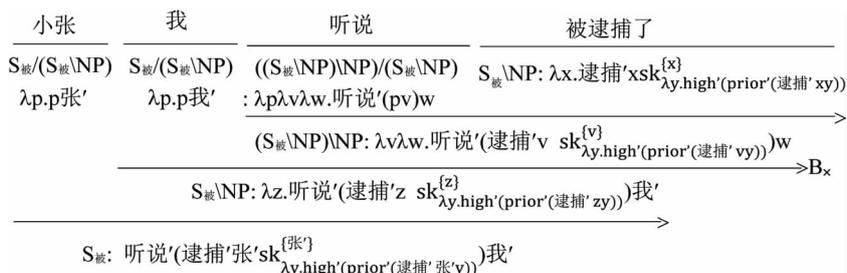


图 13 MMCCG 处理无界的汉语短语型短被动句

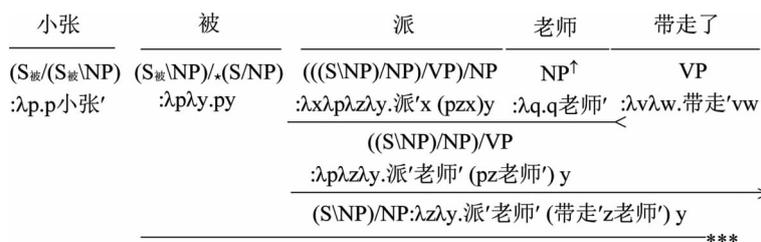


图 14 MMCCG 阻止无界的汉语短语型短被动句生成

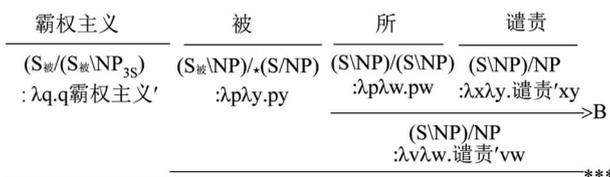


图 15 MMCCG 阻止含“所”字的汉语短语型短被动句生成



图 16 MMCCG 阻止含代名词的汉语短语型短被动句生成

MMCCG 能够阻止这三类短被动句的生成，再一次证明了 MMCCG 具有合适的生成能力。

(21) 被捕了: = $S_{\text{被}}\backslash\text{NP}:\lambda x.\text{被捕}'x$; 被俘了: = $S_{\text{被}}\backslash\text{NP}:\lambda x.\text{被俘}'x$

(三) 汉语词汇型短被动句

(20) a. 小张被捕了。 b. 小张我听说被捕了。

下面给出(20)中两个语句的 MMCCG 树图(见图 17a、图 17b)。

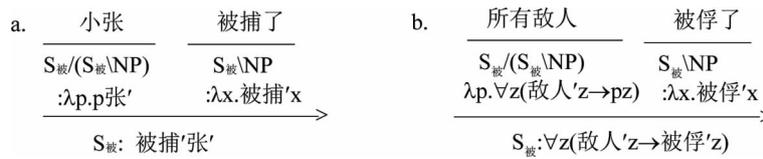


图 17 MMCCG 处理汉语词汇型短被动句

显然,被动动词“被捕”“被俘”是双音节语素,不能插入其他语素,故直接给出被动动词词条。

结语

汉语被动句是汉语十大典型句型之一,出现的频率很高,而且非常复杂,处理汉语被动句一直是语言学、逻辑学和自然语言信息处理领域的研究重点和难点。MMCCG 认为“被”不是介词或动词,而是功能性标句词,并采用了使相邻句法成分进行毗连组合生成句法-语义树的方法表明被动句的原生成性,认为句首 NP 可具有历事论旨角色,进而解释了传统分析方法无法解释的许多复杂现象。

CCG 最初的发展更关注词库的构建和句法

理论的完善,这使得 CCG 分析器在语言生成达到 90%以上之后,指标难以再提高。通往其余 10%的数据的钥匙,可能就在被忽略的语义之中。由于广义斯科伦项语义理论的引入,很容易在 MMCCG 框架内刻画被动句各成分之间的依存关系,如先行语对照应语的约束关系、中心语和主目语之间的谓词-论元关系,从语义角度弥补了句法上刻画被动句的不足之处,效果令人满意。作为 CCG 升级版的 MMCCG 是更加彻底的词汇主义语法,表现的现象比 CCG 深,能比 CCG 更快地评价语词之间具有的各种依存关系。基于 MMCCG 的分析器在分析速度和准确度上远胜 CCG。MMCCG 是目前自然语言信息处理的最理想工具,具有广阔的发展前景。

Processing of Chinese Direct Passive Sentences in Multi-modal Combinatory Categorical Grammar (MMCCG) Under the Interaction Perspective of Language, Logic and Computation

YAO Cong-jun & ZU Meng-chen

(Biquan Academy, Xiangtan University, Xiangtan 411105, China)

Abstract: Chinese passive sentences can be divided into direct and indirect sentences, as well as their length. How to deal with Chinese passive sentences has always been a difficult problem for computational linguists. The remarkable feature of Chinese direct passive sentences is that the patient of the predicate in subject sentence points to the subject in it. In the traditional analysis, syntactic operations, such as movement and deletion are often used to analyze Chinese passive sentences, which denies the original generation of passive sentences, and there are many phenomena that are difficult to explain. Multi-modal Combinatory Categorical Grammar (MMCCG) discards syntactic operations, such as shift and deletion, and generates syntactic semantic deduction trees of various passive sentences in the way of surface combination, so it can avoid all the problems in the traditional analysis of passive sentences, and prove the original generation of passive sentences from the perspective of computational linguistics.

Key words: Chinese passivesentences; Multi-Modal Combinatory Categorical Grammar; generalized Skolem Term

(责任校对 朱正余)