

doi:10.13582/j.cnki.1672-7835.2024.05.006

汉语间接被动句的多模态组合范畴语法处理

姚从军,贺春雅

(湘潭大学 碧泉书院,湖南 湘潭 411105)

摘要:汉语被动句有长短之别,也有直接和间接之分,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。根据主语与“被”后某成分的不同关系,间接被动句分为“包括式间接被动句”和“排除式间接被动句”。传统分析方法借助移位、控制和述谓等操作分析汉语间接被动句,不仅否定了汉语间接被动句的原生成性,而且违反了语法规则,并出现了诸多难以解释的现象。多模态组合范畴语法 MMCCG 是单层结构语法,它从汉语间接被动句的表层结构出发,以毗连组合的方式生成其 MMCCG 树,从而证明了汉语间接被动句的原生成性,并避免了传统分析方法处理汉语间接被动句时出现的问题。

关键词:间接被动句;多模态组合范畴语法;广义斯科伦项

中图分类号:B81 **文献标志码:**A **文章编号:**1672-7835(2024)05-0042-08

一 问题的提出

“汉语被动句是汉语十大典型句型之一,出现的频率非常高,而且非常复杂,处理汉语被动句一直是语言学、逻辑学和自然语言处理 NLP 领域研究的重点和难点”^①。通常见到的汉语被动句是直接被动句,包括长直接被动句(如“小偷被警察抓住了”)和短直接被动句(如“小偷被抓住了”),其主要特征是谓语的逻辑宾语与主语同指(比如主语“小偷”的所指是谓语“抓住”的受事)。汉语还有大量的“间接被动句”,进一步可分为“包括式间接被动句”和“排除式间接被动句”。在间接被动句中,主语可能与谓语的宾语具有非同指的其他关系,也可能与“被”后任何句法位置和句法成分无关。下面的语句 1 和语句 2 就是汉语间接被动句。在 a 中,主语“明星”与谓语的宾语“肖像权”具有领属关系,这样的被动句就是“包括式间接被动句”。在 b 中,主语“日本队”与“中国女篮打出一波进攻小高潮”的任何句法位置和成分无关,这样的被动句就是“排除式间接被动句”。“排除式间接被动句”隐含了对主

语所指施加了某种不幸,因此又被称作“不如意间接被动句”。

语句 1 a. 明星被商家侵犯了肖像权;
b. 日本队被中国女篮打出一波进攻小高潮。

两类间接被动句各自又有长短之分,下面是与二者相对应的短间接被动句:

语句 2 a. 明星被侵犯了肖像权;
b. 日本队被打出一波进攻小高潮。

长、短间接被动句的区别由“被”后 NP 是否显现而决定的。转换生成语法认为,所有语句都有一个深层结构,日常使用的语句都是语言的表层结构,它是由语言的深层结构经移位、删除等操作转换而来的。比如,该语法认为,被动句是由主动句经移位而生成,短被动句是由长被动句经删除而生成。

在语言学研究中,转换生成语法具有划时代意义,其许多研究成果对人们学习和理解自然语言句法结构具有极其重要的意义。然而,转换生成语法采用的移位、删除等操作并不符合人们的

收稿日期:2024-05-16

基金项目:湖南省社会科学基金项目(23YBA081)

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

①姚从军,祖孟晨:《语言、逻辑与计算互动视角下汉语直接被动句的 MMCCG 处理》,《湖南科技大学学报(社会科学版)》2022 年第 1 期。

直觉。相比之下,组合范畴语法是融合了理性主义和经验主义的产物,它既能达到转换生成语法的生成能力,又不需要对句法结构进行分层,只对看得见、听得到的语言成分进行运算,否定移位、删除等操作,是单层结构语法,认为所有自然语言语句都是原生成的,且在句法和语义之间有一个透明接口,这与人类习得语言的过程更相似,也符合人们对自然语言的直觉。

组合范畴语法最初的发展更为关注词库的构建和句法理论的完善,这使得基于组合范畴语法的解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% 的数据钥匙,可能就在被忽略的语义之中。引入广义斯科伦项语义理论,很容易在组合范畴语法框架内刻画汉语间接被动句各成分之间的复杂依存关系,进而证明间接被动句的原生成性,同时解释在间接被动句中出现的传统分析方法无法解释的各种复杂现象,从语义角度弥补句法上分析汉语间接被动句的不足之处。鉴于此,本文基于融入了广义斯科伦项语义理论的多模态组合范畴语法 MMCCG 处理汉语间接被动句。

二 多模态组合范畴语法 MMCCG

在纯范畴语法 AB-演算中增加体现结合性和交换性的组合规则,产生了组合范畴语法,包括原生态组合范畴语法 CCG 和多模态组合范畴语法 MMCCG。然而,CCG 不加限制地使用体现结合性和交换性的组合规则,导致生成能力过强,产生了许多病句的 CCGbank 树。另一方面,CCG 把 λ -演算作为其语义理论,因该语义理论使用的全称量词和存在量词之间有着复杂的语义依存关系,故基于这样的语义理论分析自然语言时出现了量词辖域问题^①。此外, λ -演算语义理论还使 CCG 的优势得不到充分发挥,以至于在自然语言处理中缺乏应有的语义解析能力。由此可见,从限制 CCG 的句法规则和选择更合适的语义理论出发是明智之举。

MMCCG 是组合范畴语法发展的最新形态,

句法上它先用四个模态词 \star 、 \diamond 、 \times 、 \bullet 来定义斜线类型,再用类型化斜线限制组合规则的应用范围,从而降低 CCG 的结合能力和交换能力,排除了被 CCG 认可的病句,把组合范畴语法的生成能力控制在合理范围内。MMCCG 的语义理论—广义斯科伦项理论能贴近自然语言本身的句法构造去表征自然语言的量化语义,删除存在量词,简化了量词之间复杂的依存关系,从而解决了各种量词辖域问题,并能准确表述 λ -演算语义理论无法表述的诸多语言现象,提高组合范畴语法处理自然语言的能力。

下面给出 MMCCG 的句法理论,即 MMCCG 的组合规则^②:

规则 1 词法规则:类型提升规则

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

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

规则 2 句法规则一:函项应用规则

$$a. X /_{\star} Y : f \quad Y : a \Rightarrow X : f a \quad (>)$$

$$b. Y : a \quad X \setminus_{\star} Y : f \Rightarrow X : f a \quad (<)$$

规则 3 句法规则二:函项复合规则

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

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

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

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

规则 4 句法规则三:函项替换规则

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

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

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

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

斜线下标 i 是模态算子变元,可以取 $\{\star, \diamond, \times, \bullet\}$

^①姚从军,朱乐亚,邹崇理:《广义斯科伦项理论:一种新的量词理论》,《学术研究》2021 年第 5 期。

^②姚从军,朱乐亚:《语言、逻辑与计算互动视角下汉语“把”字句的多模态组合范畴语法处理》,《湖北大学学报(哲学社会科学版)》2022 年第 3 期。

x、•}中任一模态算子为值。f、g所对应的句法范畴分别被称为规则的主函子和从属函子。主斜线类型为/ \star 或\ \star 的句法范畴只能用作规则2的函子;主斜线类型为/ \circ 或\ \circ 的句法范畴只能用作规则3a-3b、规则4a-4b的主函子,以及规则2的函子;主斜线类型为/ x 或\ x 的句法范畴只能用作规则3c-3d、规则4c-4d的主函子,以及规则2的函子;主斜线类型为/ \cdot 或\ \cdot 的句法范畴可以用作任何规则的任何输入范畴,这样的斜线常常省略模态算子 \cdot ,简写为/ \cdot 或\ \cdot 。

有了这些模态算子,可以在CCG中生成的下面错误并列结构就不能借助MMCCG的函项复合规则<B进行生成。因为给“并且”指派了范畴(S\ \setminus S)/ \star S,“并且小宝说话”的范畴S\ \setminus S就不能通过规则<B与“睡觉”的范畴S\ \setminus NP进行组合运算,如图1所示。



图1 MMCCG阻止错误的并列结构生成

MMCCG的语义理论是广义斯科伦项量词理论。为解决基于CCG分析自然语言产生的各种量词辖域问题,Steedman发展了斯科伦项量词理论。他首先基于斯科伦项定义了广义斯科伦项^①,然后继续用 λ -项表征全称量词,而用广义斯科伦项表征所有非全称量词,形成了广义斯科伦项量词理论^②。有关该理论请读者参考注释中所提原文,这里不再赘述。将广义斯科伦项量词理论作为MMCCG的语义理论,有效地解决了各种量词辖域问题^③。

三 汉语包括式间接被动句的MMCCG处理

传统上对于包括式间接被动句的分析和处理主要有移位观点和控制观点。比如,对所有格型

包括式间接被动句“明星被商家侵犯了肖像权”来说,可以假定这个句子的谓语部分“侵犯了肖像权”存在一个空所有格短语,并且这个空所有格短语与“明星”同指,接下来需要解决的问题是如何建立这样的同指关系。一种观点认为,这种同指关系可通过对所有格短语直接进行空算子移位而建立。然而,如何利用空算子移位观点来分析这样的句子?“所有格提升”方法显然行不通,原因有三:一是因为这会违反Ross的左分支条件^④;二是对于语句“程序员被AI夺走了那份最珍惜的工作”来说,移位来自关系从句“那份…最珍惜的工作”内部,这就违反了复杂NP限制条件;三是移位假设无法解释为什么语句“桌子被我打断了一条腿”听起来非常自然,而语句“桌子被我看到了一条腿”听起来就让人无法接受。

Huang认为,应该把间接被动句分析为含有复杂谓语V'(比如语句1中的“侵犯了肖像权”)的语句,V'带有一个“外层宾语”,这个外层宾语控制空所有者;内层宾语从动词那里获得受事角色,外层宾语从V'那里获得受影响角色;受控的空所有者不是语迹,而是代语Pro;不是代语Pro移位,而是外层宾语进行了空算子移位,移至IP位置,然后与主句主语同指^⑤。依照这个假设,间接被动句“明星被商家侵犯了肖像权”的分析如图2所示。其中外层宾语是语迹“ t_i ”,内层宾语是“Pro肖像权”;“ t_i ”控制内层宾语中的空所有者“Pro”,“ t_i ”移位到 OP_i 且与“明星”同指。

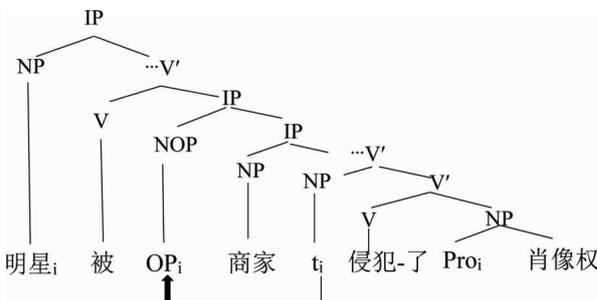


图2 “外层宾语”视角下的汉语包括式长间接被动句的生成

①Steedman M. *Taking Scope: The Natural Semantics of Quantifiers*. Cambridge, MA: The MIT Press, 2012, p.22.

②Steedman M. *Combinatory Categorical Grammar: An Introduction*. Philadelphia: The SOMESUCH Press, 2017, p.171.

③姚从军,朱乐亚,邹崇理:《广义斯科伦项理论:一种新的量词理论》,《学术研究》2021年第5期。

④Ross, J. R. *Constraints on Variables in Syntax*. Ph.D. MIT, Cambridge, Massachusetts, 1967, p. 207.

⑤Huang, C-T. James. “Complex predicates in control”. *Control and Grammar*. eds. Richard K. Larson, Sabine Iatridou, Utpal Lahiri, and James Higginbotham. Dordrecht :Kluwer Academic Publishers, 1992, pp.109-149.

与上述这些语法理论的解析不同,组合范畴语法是单层结构语法,不承认移位、删除等操作,其基本做法是从不同层次的语言表达式那里抽象出句法范畴,再通过范畴的运算来刻画自然语言表达式由小到大逐层逐级地毗连生成过程,从而表明所有自然语言语句都是原生成的。因此,从组合范畴语法的视角来看,所有汉语间接被动句都是原生成的,不是从其他语言结构通过转换而生成的,“被”字仅仅是间接被动句的指示词。下面先在融入了广义斯科伦项语义理论的 MMCCG 中处理汉语包括式间接被动句。

(一) 汉语包括式长间接被动句的 MMCCG 处理

- 语句 3 a. 程序员被 AI 夺走了一个工作机会;
b. 程序员我听说被 AI 夺走了一个

工作机会;
c. 程序员被 AI 夺走了一份珍惜的工作。

词条 1 被 := (S_被 \ NP) / ★ S: λpλz.pz;
一个 := NP[↑]_{3SG} / N_{3SG}: λpλqλy.q (skolem 'λx.px ∧ 有'xy)
一份 := NP[↑]_{3SG} / N_{3SG}: λpλqλx.q (skolem 'px)

在图 3 和图 4 中,基于融入了广义斯科伦项语义理论的 MMCCG,以表层组合的方式表征了谓语“夺走”、施事“AI”和受事“一个工作机会”之间的谓词-论元关系,刻画了“程序员”和“一个工作机会”之间的领有关系,其中组合规则和广义斯科伦项起着至关重要的作用。基于 λ-演算语义理论很难给出如此简洁且精准的语义表征。



图 3 汉语包括式有界的长间接被动句的 MMCCG 处理

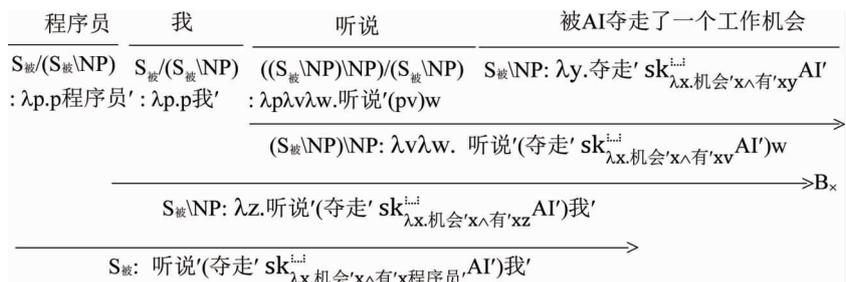


图 4 汉语包括式无界的长间接被动句的 MMCCG 处理一

同理,在语句 3c 的推演中(见图 5),以表层组合的方式不仅表征了主句谓语“夺走”、施事“AI”和受事“一份工作”之间的谓词-论元关系,

而且通过广义斯科伦项 sk^{i,i}_{λv.珍惜'v程序员'∧工作'v} 刻画了关系化从句中的谓语“珍惜”、施事“程序员”和受事“一份工作”之间的谓词-论元关系^①。

① sk^{i,i}_{λv.珍惜'v程序员'∧工作'v} 指称类型为 e 的个体,根据其限定子条件 λv.珍惜'v 程序员'∧工作'v,该个体就是变量 v 的取值,即那份程序员珍惜的工作。

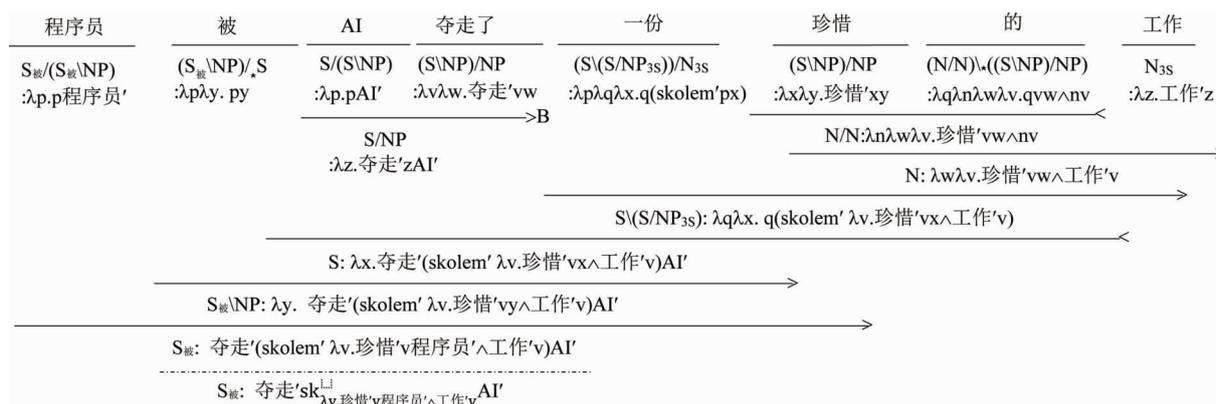


图5 汉语包括式无界的长间接被动句的MMCCG处理二

(二) 汉语包括式短间接被动句的MMCCG

处理

- 语句4 a. 程序员被夺走了一个工作机会;
- b. 程序员被夺走了一份珍惜的工作。

词条2 被: = (S_被 \ NP) / * S_被: λpλx.px
 夺走了: = VP_被: λx. 夺走'
 x sk^{x}_{λy.high'(prior'(夺走'xy))}
 一个: = NP[↑]_{3SG} / N_{3SG}: λpλqλa. q
 (skolem'λx.px ^ 有'xa)

一份: = NP[↑]_{3SG} / N_{3SG}: λpλqλx. q
 (skolem'px)

的: = (N/N) \ * ((S \ NP) / NP):
 λqλnλwλv.qvw ^ nv

在词条2中, sk^{x}_{λy.high'(prior'(夺走'xy))} 依赖于其 If (逻辑形式)一统领的 x, 指称“能夺走 x 的任何东西”, 表示潜在的施事主语, 这样的广义斯科伦项往往由中心语(如这里的“夺走”)的范畴引入。另外, VP_被 表示短被动句谓语。语句4a-b的推演分别如图6和图7所示。

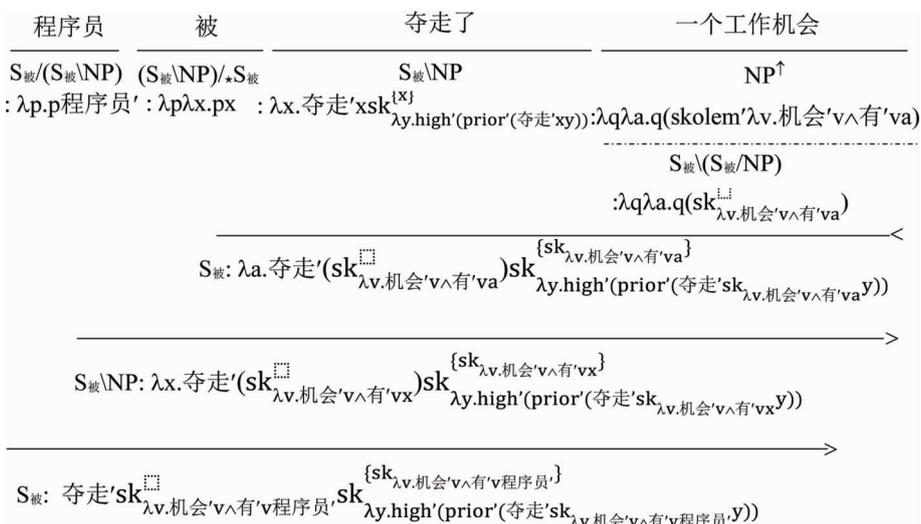


图6 汉语包括式有界的短间接被动句的MMCCG处理

在图6中,“夺走了”的宾语和潜在主语都用广义斯科伦项表述, 不仅表征了它们之间的谓词-论元关系, 而且表征了“程序员”对“一个工作机会”的领有关系; 在图7中,“夺走了”的宾语和

潜在主语也都用广义斯科伦项表述, 其中主句中的谓词-论元关系和关系从句中的谓词-论元关系都得到精准刻画。

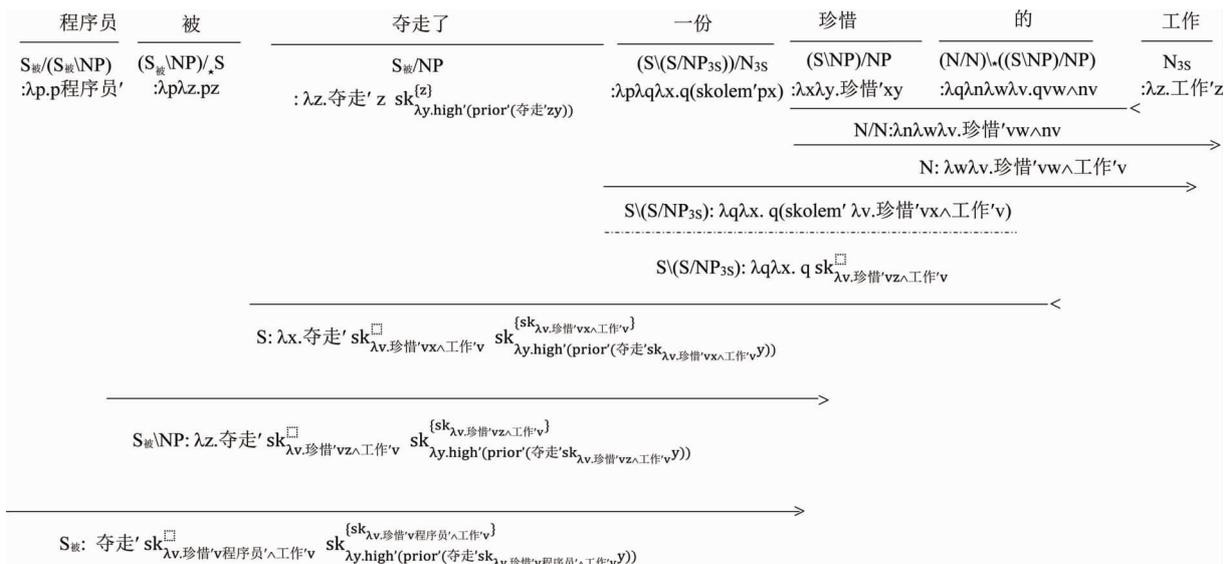


图 7 汉语包括式无界的短间接被动句的 MMCCG 处理

四 汉语排除式间接被动句的 MMCCG 处理

在直接被动句中,主语与谓语的受事宾语同指。在包括式间接被动句中,主语与谓语的受事宾语具有非同指的其他关系,比如领有关系。总之,在这两类被动句中,主语与“被”后某一句法位置或成分有关。但是,在排除式间接被动句中,主语与“被”后任何句法位置或成分无关。

在排除式间接被动句中,虽然被动句的主语与“被”后某一句法位置或成分无关,但被动句的主语与“被”后整个表达式在语义上有联系:“被”后整个表达式所表述的内容使被动句的主语所不如意,或者说前者对后者有负影响。传统语法理论对于排除式间接被动句的分析和处理有许多观点和方法,最为典型的是述谓观点及有关述谓的原则或公设。组合范畴语法是彻底的词汇主义语法,抛开了有关述谓的特设原则或公设,它将所

有信息都编码在汉语排除式间接被动句的词汇范畴上,再通过范畴的运算来刻画汉语排除式间接被动句由小到大逐层逐级地毗连生成过程,从而表明汉语排除式间接被动句是原生成的,不是由其他类型的语句转换而来的,“被”字仅仅是汉语排除式间接被动句的指示词。下面在 MMCCG 中处理汉语排除式间接被动句:

(一) 汉语排除式长间接被动句的 MMCCG 处理

- 语句 5 a. 原告被被告自证了^①;
- b. 原告我听说被被告自证了。

词条 3 被: = (S_被 \ NP) / ★ (S_被 \ NP):
 λpλy.py
 自证了: = (S_被 \ NP) \ NP: λyλx.自证'x ∧ 负影响'y(自证'x)

有了词条 3,我们就可以在 MMCCG 中处理语句 5a-b,如图 8 和图 9 所示。

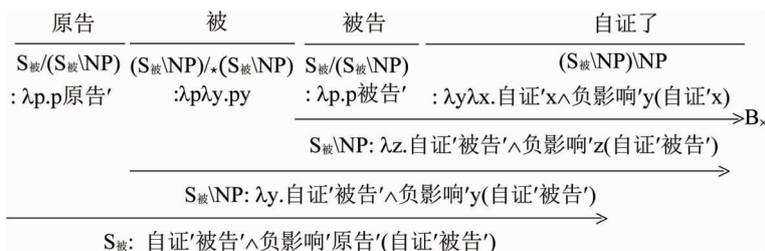


图 8 汉语排除式有界的长间接被动句的 MMCCG 处理

①被告自己提供证据证明自己的无辜,不是由原告触发的,但会对原告产生不利的影响。

最后得到的语义表达式“自证'被告'∧负影响'原告'(自证'被告)'"表明,被告自证了,并且被告自证给原告带来了消极后果。下图所得结果与此类似。

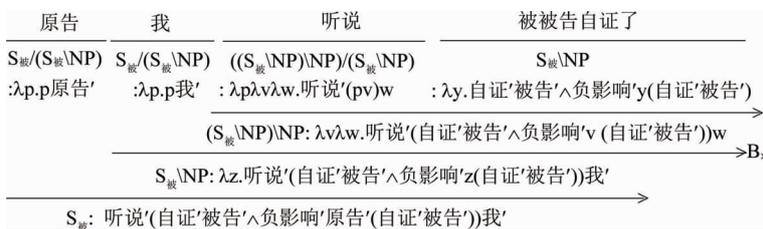


图9 汉语排除式无界的长间接被动句的MMCCG处理

MMCCG 可以处理汉语排除式长间接被动句的关键在于其彻底的词汇主义倾向,即把所有的句法和语义信息都编码在词汇范畴上。在这里,MMCCG 不仅把被动信息编码在“被”字的范畴上,还把这种被动句的“不如意”信息编码在“自证了”的范畴上,在最终的语义表达式中表征了“被”后整个表达式对主语所指的“不如意”谓述关系,这与人们对汉语排除式间接被动句的直觉相符合。另外,组合规则也发挥了重要作用。下面的分析与此相同。

(二) 汉语排除式短间接被动句的 MMCCG 处理

与汉语排除式长间接被动句相比,汉语排除

式短间接被动句的不同之处在于“被”后 NP 没有在语形或语音上显现出来,如下所示:

- 语句 6 a. 原告被自证了;
- b. 原告我听说被告自证了。

词条 4 被: = $(S_{被}\backslash NP)/\star VP_{被}: \lambda p\lambda z.pz$
 自证了: = $VP_{被}: \lambda y.$ 自证'sk^{y} _{$\lambda x.high(\text{prior}(\text{自证}'x)) \wedge \text{负影响}'y(\text{自证}'x)$}

在词条 4 中,sk^{y} _{$\lambda x.high(\text{prior}(\text{自证}'x)) \wedge \text{负影响}'y(\text{自证}'x)$} 表征实施自证行为的个体,该个体依赖于 y,但自证行为不是 y 所期望的行为,这样的广义斯科伦项也是由中心语动词(如“自证了”)的范畴引入。另外,VP_被表示短被动句谓语。基于 MMCCG,语句 6a -b 的推演分别如图 10 和图 11 所示。

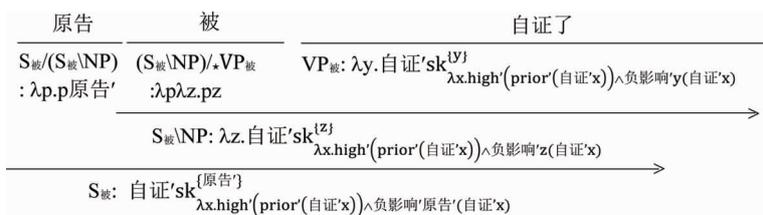


图 10 汉语排除式有界的短间接被动句的 MMCCG 处理



图 11 汉语排除式无界的短间接被动句的 MMCCG 处理

既然原告被自证,那么该原告一定是此自证所涉及的这场官司的原告。在原告发起的官司中,对

方(通常是被告)能够自证一定会对原告产生“不如意”影响,如原告官司打输了。推论结果中表征自证者的广义斯科伦项 $sk_{Ax.high}^{原告}(\text{prior}(\text{自证}x)) \wedge \text{负影响}(\text{原告}(\text{自证}x))$ 体现了两点:第一,自证者是受原告约束的广义斯科伦函项,即是该原告的对方;第二,在该原告发起的这场官司中,对方的自证行为对该原告产生了不利影响。

基于 MMCCG 可以处理汉语排除式短间接被动句,除了其彻底的词汇主义特性和组合规则的作用之外,广义斯科伦项语义理论也起着至关重要的作用。

结语

自然语言处理 NLP 是 AI 皇冠上的明珠,NLP 研究的突破极有可能导致 AI 的里程碑式发展,在

NLP 的追梦时代必然产生以自然语言语义为研究目标的逻辑语义学。组合范畴语法是目前为止最为简洁、广受计算语言学家青睐的能力语法,已成为对自然语言进行逻辑语义分析的理想工具。作为一阶逻辑系列量词理论和广义量词理论取长补短的产物的广义斯科伦项语义理论融入组合范畴语法后,可以使组合范畴语法的优势得到充分发挥。

基于“广义斯科伦项理论+ MMCCG”处理汉语间接被动句不仅从计算语言学视角证明了汉语间接被动句的“原生成性”,而且极大地提高了组合范畴语法处理汉语间接被动句的能力和准确性,在一定范围内改变了目前应用组合范畴语法时“重句法、轻语义”的状况,对语言学和语言逻辑的进一步发展必将起到较大的促进作用。

On Chinese Indirect Passive Sentences in Multi-modal Combinatory Categorical Grammar

YAO Congjun & HE Chunya

(Bi Quan Academy, Xiangtan University, Xiangtan 411105, China)

Abstract: Chinese passive sentences can be divided into long and short ones, as well as direct and indirect ones. How to deal with Chinese passive sentences has always been a difficulty for computational linguists. According to the different relationships between the subject and the components after “被(bei)”, while indirect passive sentences can be divided into “including indirect passive sentences” and “excluding indirect passive sentences”. Traditional analysis methods use syntactic operations, such as move, control and predication to analyze Chinese indirect passive sentences, which not only deny the original generation of Chinese indirect passive sentences, but also violate some grammar rules, and there are many phenomena hard to explain. Based on multi-modal combinatory categorial grammar, which is a single-layer structural grammar, the MMCCG deductive tree of various Chinese indirect passive sentences can be generated by surface composition, and the original generation of Chinese indirect passive sentences is proved from the perspective of computational linguistics, at the same time avoiding many problems faced by traditional methods in analyzing passive sentences.

Key words: indirect passive sentences; multi-modal combinatory categorial grammar; Generalized Skolem Term

(责任校对 龙四清)