

PP<被> + VP1 + VP2 格式歧义的自动消解*

詹卫东

提要 本文从计算机处理自然语言的角度着眼,指出了 PP<被>+VP1+VP2 格式对计算机造成的歧义问题,并通过归纳 PP VP 组合和 VP VP 组合的句法限制条件,找到了正确分析这个歧义格式的组合层次和内部结构关系的规则。

§ 0

PP<被>指“被”加宾语成分形成的介词短语,本文也称之为“被”字结构(用符号称说时,下文一般用 PP 代指 PP<被>);VP1、VP2 指动词性短语¹。请看例句:

- (1) 被老王/pp 拦住/vp 抓个正着/vp
- (2) 被警察/pp 扣了车/vp 罚了款/vp
- (3) 被工人们/pp 搬了/vp 出去/vp
- (4) 被老师/pp 批评/vp 写了检查/vp
- (5) 被爸爸/pp 打/vp 被妈妈骂/vp

五个例句在表层形式上按同一个格式排列,在短语结构组合方式和内部结构关系上各不相同。

前三例结构层次上是按 a 式: [PP [VP1 VP2]] 组合;

后两例结构层次上是按 b 式: [[PP VP1] VP2] 组合。

都是按 a 式组合,例(1)VP1 跟 VP2 之间是连动关系;例(2)VP1 跟 VP2 之间是联合关系;例(3)VP1 跟 VP2 之间是述补关系。结构关系各有不同。

都是按 b 式组合,例(4)内部是连动关系;例(5)内部是联合关系。结构关系也有差异。

人面对这些例句,理解判断上不会发生问题,都是单义的。如果让计算机来分析这些例句,要能准确判定例(1)(2)(3)是按 a 式组合,例(4)(5)是按 b 式组合,而且相同的组合方式内部还有不同的结构关系,并不是件容易的事情。这些例句都对对应着同一个格式,也就是本文题目中的 PP+VP1+VP2。对计算机而言,这是一个歧义格式²。什么时候该把这个格式分析为 a 式组合,什么时候该把它分析为 b 式组合,以及内部不同的结构关系受到哪些条件的制约,都需要人从句法条件上找制约因素,然后告诉给计算机作为判断的依据。

要确定 PP VP1 VP2 这个三项组合到底是什么样的结构层次以及内部是什么结构关系,

注* 本文是作者硕士论文研究工作的主要部分。在此向两位导师陆俭明教授和俞士汶教授表示深深的谢意。中科院计算所二室刘群先生、北大计算语言学研究所的老师和同学都曾给予作者很多帮助,借此机会也一并致谢。

可以把问题转化为两项组合的情况来考虑。语言结构中两项组合是最基础最广泛的方式。词和短语，两两之间的组合规律可以认为是覆盖了语言组合规律的大部分内容。对本文这个组合歧义格式，如果我们把 PP VP 两项组合和 VP VP 两项组合的限制约束条件都概括归纳出来，PP VP1 VP2 三项组合的歧义问题也就相应得到解决了。因此，需要解决的问题实际上就转化成：

(1) PP 跟 VP 组合，VP 要满足什么条件？

(2) VP 跟 VP 组合，不同结构关系对前后两个 VP 分别有哪些限制？

下面我们就分别来讨论这两个问题，从句法分布上找条件限制。

§ 1

首先我们来找 PP VP 组合中 VP 要满足的条件。

考察 VP 跟“被”字结构组合有什么条件限制，可以有两个角度。一是看什么样的 VP 不能跟“被”字结构组合。一是看什么样的 VP 能跟“被”字结构组合。从前一个角度看问题，倾向是从严；从后一个角度看，倾向是从宽。因为我们的研究是面向计算机处理汉语，所以条件限制倾向于严格一些。条件严格同时也就意味着条件明确。下面分别从这两个角度来考察 VP 跟“被”字结构组合的限制要求。

(一) 不能跟“被”字结构组合的 VP。

1. 有些动词本身就不能跟“被”字结构组合。

现代汉语中有相当数量的动词，无论是简单的光杆动词形式(bared form)，还是复杂的动词短语形式(combined form)，都不能跟“被”字结构组合构成合法组合体。我们考察了《动词用法词典》中的 1221 个动词³，其中有 762 个，占 62% 的动词无论采用什么形式都不能跟“被”字结构组合。下面简单分类列举一些不能跟“被”字结构组合的动词。

- | | |
|-------------------------|----------------------|
| (1). 连系动词 | 是 等于 姓 |
| (2). 形式动词 | 进行 给予 加以 |
| (3). 趋向动词 | 上 下 来 去 |
| (4). 助动词 | 能 肯 会 愿意 |
| (5). 对称性动词 ⁴ | 相会 交谈 交换 交涉 |
| (6). “~得”动词 | 认得 晓得 懂得 显得 获得 记得 |
| (7). 不及物动词 ⁵ | 奔 走 蹦 跑 闭 爆炸 咳嗽 |
| (8). “有”动词 | 有 没有 |
| (9). 心理动词 | 爱好 爱惜 保证 奢望 |
| (10). 态度动词 | 满意 赞成 |
| (11). 遭受类动词 | 遭到 受到 受 |
| (12). 一些动宾格动词 | 洗澡 鞠躬 丢脸 担心 告别 出力 提议 |
| (13). 一些谓宾动词 | 打算 同意 考虑 参加 |

(4). 其它 办理 帮忙 帮助 爆发 呈现

可以看到,上面这些类动词并不是按什么统一的标准分出来的。我们只是想说明不能跟“被”字结构组合的动词涉及面很宽,并不打算对这些动词作进一步的归纳概括。也就是说,我们只单纯的描写一个动词是否有跟“被”字结构组合的能力,而不对这种能力做解释。一个笼统处理的做法就是,为现代汉语动词建立一个“被”属性来标记一个动词是否能跟“被”字结构组合。象上面列举的这些动词,“被”属性的取值就简单地填为“否”。

对计算机处理来说,有两种情况还值得特别说明一下。一种是象“当”这样的,可以读阴平和去声两个声调,是两个不同的词,用了相同的字形。读阴平的“当”有(1)担任;(2)掌管两个义项,都不能跟“被”字结构组合;读去声的“当”有(1)当做 (2)认为 两个义项,都能以一定的形式跟“被”字结构组合。如:我被他当小偷了。 另一种情况是象“打”这样的动词,有多个义项,而不同义项下的“打”跟“被”字结构组合的能力不一样。比如:“打人”的“打”可以跟“被”字结构组合,如:我被他打了一下。 但“打电话、打篮球”的“打”却绝不能受“被”字结构修饰。还有如“跑”,一般认为是不及物动词,不能跟“被”字结构组合,但作为“逃跑”义项的“跑”却能以述补式 VP 形式跟“被”字结构组合。如:被他跑掉了。对计算机而言,上述两种情况其实没有差别,都要求我们在鉴别动词能否跟“被”字结构组合时,要按义项进行。特别是现代汉语中的单音节常用动词,一词多义的情况很普遍,不能笼统地说能不能跟“被”字结构组合,而应该在具体的义项下考察其跟“被”字结构组合的能力。

在鉴定有些动词能否跟“被”字结构组合时,会碰到模糊情况。譬如“喜欢”、“爱”这样的动词,就处在游移状态,虽然可以跟“被”字结构组合,但尚未普遍使用。事实上,语言总是在发展变化的,能说与不能说有相对性。我们不过多地纠缠这类问题。对大多数动词而言,能否跟“被”字结构组合在当前的共时普通话系统中还是容易确定的。

综合上面的原则和方法,我们就可以在词典中为每个具体的动词贴上一个属性标签,来标记它跟“被”字结构组合的能力。计算机掌握了这样的语言知识,碰到有些 PP VP1 VP2 格式的句子,就能确定其结构层次了。例如:

A. (王教授)被美国一所大学邀请参加了第八届地质学年会

先分析出小的结构段:被美国一所大学/pp 邀请/vp 参加了第八届地质学年会/vp

前一个 VP 是“邀请”,能跟“被”字结构组合;后一个 VP 的主要动词是“参加”,已知它不能跟“被”字结构组合,即在结构上后项 VP 不能受“被”字结构“统治”(govern),因此这个短语肯定就不能按 a 式切分,只能是按 b 式切分:

[被美国一所大学 邀请] 参加了第八届地质学年会]

意义理解上也跟切分情况吻合,“王教授”是“被邀请”,而“参加”则不是被动的,而是主动的。

从这个简单的例子就可看出,如果我们能准确判定一个 VP 能否跟“被”字结构组合,那对判定 PP VP1 VP2 组合的结构层次是很有裨益的。

上面只是就动词光杆形式跟“被”字结构组合的情况做了调查,VP 还有各种复杂的形式,

都需要一一考察其跟“被”字结构组合的限制。下面我们就来讨论不能跟“被”字结构组合的复杂 VP。

2. 有些 VP 句法结构不能跟“被”字结构组合。

“被”字结构不仅对动词有限制，对作为句法结构的 VP 也有一些结构形式上的限制。下面带否定标记的 VP 都不能跟“被”字结构组合。

(I) 状中式 VP。状语由否定副词“不、没、没有”等充当。试比较：

被打了 被狠打了一顿 *被不打 *被没打
被小王吃了 被小王全吃了 *被小王不吃 *被小王没吃

现代汉语“被”字结构的否定形式，一般都是在结构之外否定，也就是在结构上是由否定成分统治“被”字结构。比如：“不被打、没被打”这样的说法就是允许的。

(II) 述补式 VP。如果是否定的表可能述补式 VP，就不能跟“被”字结构组合。如：

被小王吃完了 *被小王吃不完
被小王挖深了 *被小王挖不深

现代汉语要表达上面这样不可能的意思，一般都不采用“被动”说法，而直接用主动形式表达。如：“小王吃不完这些西瓜”。或者用能愿动词结合“被动”形式表达。如：“这些西瓜不可能被小王吃完”。

(III) 含上面(I)、(II)情况的其他 VP。

如果述宾式 VP 的述语是由(II)中的 VP 充任的，就不能跟“被”字结构组合。如：

被打中了眼睛 *被打不中眼睛

如果连谓式、联合式 VP 的前后项中任意一项是由(I)或(II)中的 VP 充任的，就不能跟“被”字结构组合。如：

被敌人烧光抢光 *被敌人没烧光没抢光 *被敌人烧不光抢不光
被他吃了 *被他吃不吃 *被他吃没吃
被小王开门撞了一下 *被小王开门没撞一下 *被小王没开门撞一下

有的 VP 内部含否定标记，仍可跟“被”字结构组合。比如表结果状态的述补式 VP。请看对比例句：

A	B
被他洗得很干净	被他洗得越发不干净了
被他伤害得很深	被他伤害得不轻
被他刷得非常白	被他刷得一点都不白了

B 组例子中 VP 都含否定标记，但跟“被”字结构组合却都是合法的。事实上，在这里，否定标记出现的位置，或者更确切地说是否定标记在 VP 中的嵌套层次的深浅，才是影响 VP 能否跟“被”字结构组合的真正因素。上面几例中的否定词“不”都在 VP 的补语内部，从结构层次上看，嵌套得比较深。线性形式上表现为离“被”比较远。其中“不白、不干净”尤为明显。而上面(I)(II)(III)中的否定词要么是跟“被”在同一个结构层面上，如(I)(II)中的例子；要么就是处在跟“被”同结构层次的 VP 的中心语位置，如(III)中的例子，否定词都在 VP 中的

述语或并列中心语成分里。就结构层次来说，VP 包含的否定词跟“被”的句法关系比较密切，VP 就不大能跟“被”字结构组合了。

综合上面的分析，可以看出，只要 VP 中特定的结构位置有否定成分（主要就是状语位置和组合式述补的中缀位置），它就不能跟“被”字结构组合。这是一个比较严格的结构形式限制。这条规则同样能帮助计算机判定 PP VP1 VP2 组合的结构层次。如：

B. (李广)被敌人的铠甲挡着了没射中敌人的要害

首先分析出小结构段：被敌人的铠甲/pp 挡着了/vp 没射中敌人的要害/vp

前一个 VP “挡着了”能跟“被”字结构组合；后一个 VP 的主要动词是“射”也能跟“被”字结构组合，这跟上文例句 A 不同，但是不是由此就判定该按 a 式切分了呢？显然不能。尽管主要动词“射”能跟“被”字结构组合，但后项 VP 却是一个带否定标记“没”的状中式 VP，属不能跟“被”字结构组合的短语形式。因此，例 B 跟上文例 A 一样，仍然是不能按 a 式切分，只能按 b 式切分为：

[[被敌人的铠甲 挡着了] 没射中敌人的要害]

这个例子说明，如果我们清楚地知道某种形式的 VP 不能跟“被”字结构组合，对确定 PP VP1 VP2 组合的结构层次也是非常有帮助的。对计算机处理来说，我们可以仿照上文给动词加上“被”属性标记的做法，给动词短语也加上一个“否定”标签，来标记短语内部是否含有否定成分。譬如可以给“没射中敌人的要害”，还有如上文的“吃不完、打不中”等短语贴上“否定”=“是”的标签。带有这个标签的 VP，跟“被”字结构组合就受限制。而像“洗得越发不干净”这样的 VP，因为否定成分嵌在补语中，整个短语不能打上“否定”标记。用这样的办法，就能帮助计算机正确判断含否定成分的动词短语跟“被”字结构组合的层次了。

除了上面列举的不能跟“被”字结构组合的 VP 的情况，现代汉语中还有相当多 VP 能跟“被”字结构组合，但必须满足一定的条件。譬如泛泛地说，除了否定形式的构造，述补式 VP、状中式 VP 等等大都能跟“被”字结构组合，但还有述补式、状中式构造的 VP 不能跟“被”字结构组合，还有必要把能跟“被”字结构组合的 VP 的范围进一步缩小。下面我们就来考察这类情况。

(二) 必须满足一定条件才能跟“被”字结构组合的 VP。

除了（一）中概括的不能跟“被”字结构组合的 VP 情况，其他 VP 笼统地说就属于可以跟“被”字结构组合的了。但考察语言事实不难发现，能跟“被”字结构组合的 VP 还需满足各种制约条件。《动词用法词典》的 1221 个动词中，能以光杆形式跟“被”字结构组合的，只有十分之一左右，如“传染、处罚、欺负”等等；不能以光杆形式但能以复杂形式跟“被”字结构组合的多一些，占四分之一强。譬如：“看、炒、尝、淋”等等。我们不能单说“被看”，但能说“被他看到、被看了几眼”。还有相当多的动词，如“打、批评”等，即使本身能以光杆形式跟“被”字结构组合，组成的 VP 要么是用在特殊语境，要么也是粘着的，句法上不能自足（self-sufficient），往往被更大的结构包含。譬如“被打的工人、被老师批评的时候”等，“被”字结构加动词形成的 VP 都包含在更大的 NP 结构中。多数情况下，动词都是以复杂形式的 VP 跟“被”字结构组合形成相对独立的更大的 VP。在具体动词到底以什么形式跟“被”

字结构组合这个问题上，涉及到动词的个性因素和语义搭配因素较多，因为我们主要是着眼于计算机理解句子（understanding），而非生成（generation），因此对动词必须以何种形式跟“被”字结构组合，只重点做结构类型上的分析，而不去详细考虑区分个别复杂因素。下面我们仅就 VP 的不同结构类型来讨论：能以某种复杂形式跟“被”字结构组合的 VP，其内部组成成分有什么句法限制要求。

1. 附加式 VP

附加式 VP 由 VP 带助词构成。嵌套在内的 VP 可以是很复杂的。比如：

[VP [VP 抡起棒子 把敌人打翻] 了]

这里我们只管内部是简单动词形式的附加式 VP。单个动词附加“着”、“了”、“过”后能否跟“被”字结构组合，似乎跟动词本身性质以及意义有比较密切的关联。请看例子：

（杯子）被小王摔了	被小王摔过	*被小王摔着
?被小王看了	被小王看过	*被小王看着 ⁶
被敌人捆了	被敌人捆过	被敌人捆着

上面的这几个动词“摔、看、捆”，可以根据语义特征的不同分属不同的动词次类，来描写解释为什么有能力差别；也可以纯粹作为句法现象，象（一）中我们考察每个具体动词跟“被”字结构组合的能力一样，逐个加以分析标记。我们采用后一种方式。

2. 重叠式 VP⁷

重叠式 VP 中只有以 v u<了> v 形式重叠的能跟“被”字结构组合。其他重叠形式一般不能。例如：

被他闻了闻	*被他闻闻	*被他闻一闻
被小王推了推	*被小王推推	*被小王推一推

重叠式跟“被”字结构组合造成的 VP 一般是粘着的，不大能自足成句。

3. 离合式 VP⁸

离合式 VP 一般都有相应的紧凑动词形式。大部分离合式 VP 都是述宾构造，其相应的紧凑形式也就是以所谓动宾格方式构词的动词。这样的动词因为有宾语存在，所以大部分不大能跟“被”字结构组合。但也有极少数例外的情形。如：

被他将军了	被他将了一军
被人造谣了	被人造了谣

“将军”、“造谣”都是离合式动词，并且能以离合形式跟“被”字结构组合。有关离合式 VP 能否跟“被”字结构组合，可以象上文判断具体动词能否跟“被”字结构组合一样，逐个考察标记。在结构上并没有什么特别的要求。

4. 述补式 VP

述补式 VP 内部构造形式的情况比较多。上文已经把带否定标记的述补 VP 排除出去了。肯定的述补式 VP 中，大多数情况能跟“被”字结构组合，但表可能的述补式 VP 不能。例如：

被吃得精光	*被吃得光
被打得半死不活	*被打得死

被贴在墙上
被送到洗衣店

表可能是从意义上来说明述补式 VP 的特点。对计算机处理而言，我们更希望能找到形式上可供鉴别的依据。进一步比较考察可以发现，不管是表示可能，还是表示结果状态，一般只要是以简单的词形式充当补语的组合式述补 VP，都不大能跟“被”字结构组合。例如：

被染得很红 被染得红红的 *被染得红
被打扫得非常干净 被打扫得干干净净 *被打扫得干净

“染得红”、“洗得干净”都既可看作带可能补语的述补结构，也可看作带状态补语的述补结构，但因为补语都是由简单的词形式充任，所以不管在哪个意义上，它们都不能跟“被”字结构组合。

带趋向补语的述补式 VP 一般能跟“被”字结构组合。但如果趋向补语不是表示实际的趋向位移意义，而是其他引申义，跟“被”字结构组合的能力就会受影响。例如：“桌子被他搬下去了”只能表示实际位移义，不能表示动作持续。这说明“被”字结构对趋向述补 VP 的意义选择有限制。但同时我们也发现有更为复杂的情况。如：“被他这么拖下去”似乎就既可表位移也可表“拖”的动作持续，是一个歧义短语。这其中还涉及到动词的语义因素。这个问题将另作研究，本文暂不做详细讨论。

5. 述宾式 VP

一般说来，因为“被”字结构是把受事宾语提前放在主语位置的一种句法手段，这样它后面的 VP 大都不会再带宾语，也就是说“被”字结构后 VP 是述宾式构造的，应该比较少。但这里面还要区分一些不同的情况。主要跟宾语成分的性质有关。

先看看宾语是体词性成分的情况。

宾语如果是单纯表数量的成分，一般能跟“被”字结构组合。如：

被老师看了两眼
十个苹果被他吃了八个

宾语成分的语义如果跟“被”字前面的主语成分的语义有领属关系或整体与部分的关系，一般也能跟“被”字结构组合。如：

银行被偷走了很多钱 （“银行”领属“钱”）
他被强盗打着了眼睛 （“他”跟“眼睛”是整体与部分的关系）
桌子被撞断了一条腿 （“桌子”跟“腿”是整体与部分的关系）

有时这种语义联系可以是比较间接的，如：

我被妻子拽了一下衣角 （“我”跟“衣角”有间接的领属关系）

如果 VP 含其他类型的体词性宾语，还需作细致的分析。举例来说，如果是工具宾语，则一般得是受数量成分修饰的 NP，才能跟“被”字结构组合。请比较：

被他写秃了三十支笔 *被他写秃了毛笔 （工具宾语）

如果是结果宾语，不管含不含数量成分，都能跟“被”字结构组合。如：

被他盖了两间瓦房 被他盖了瓦房 （结果宾语）

从上面分析可知，含体词宾语的 VP，能否跟“被”字结构组合，制约因素不仅涉及到 VP 内部宾语成分的句法形式，同时又牵涉到语义上的一些限制，情况比较复杂。仅从句法形式上看，大致的倾向是，述宾式 VP 内部的述语和宾语一般不能同时都是简单形式，除非宾语跟主语有语义关联，如：儿子被父亲打手心。“打手心”是一个最简形式的述宾式 VP。这个短语结构能跟“被”字结构组合，是因为“儿子”跟“手心”有语义关联（领属关系）。其他的例子还有“抄家”、“挡路”等等。语义制约因素很难教给计算机掌握。目前只能笼统地规定跟“被”字结构组合的述宾式 VP 不能是最简形式的。对“打手心”这样的例子就当作特例处理。

下面再来看宾语是谓词性成分的述宾式 VP。

谓词性宾语成分包括 VP、AP、DJ⁹等短语类型。一般说来，带 VP、AP、DJ 宾语的都不能跟“被”字结构组合。只有带 VP 宾语的有极少数例外，比如：“他被认为是民族英雄”。似乎只有“认为”、“看作”、“当作”等极少数几个动词有这样的用法。而且作宾语的 VP 也局限于“是~”类型。

概括言之，述宾式 VP 跟“被”字结构组合的机会比较少。能跟“被”字结构组合的述宾式 VP 对其体词性宾语成分有句法形式以及语义上的要求，限制条件根据具体动词的不同而有比较显著的个性差异。

6. 状中式 VP

状中式 VP 的中心成分主要由上面谈到的五类 VP 充任。如果 VP 的中心成分不能跟“被”字结构组合。整个状中式 VP 肯定也就不能跟“被”字结构组合了。对中心成分满足跟“被”字结构组合条件的 VP，我们再来考察其状语成分要受哪些限制。

充当状中式 VP 状语成分的，包括 DP、AP、MP、TP、PP、SP¹⁰等短语类型。其中 SP 充当状语的肯定不能跟“被”字结构组合。剩下前五个各自情况都有不同，下面分别讨论。

DP 充当状语的。只有小部分能跟“被”字结构组合。包括下面两种情况：

名词、动词、形容词加“地”形成的 DP。如：

被他[天才地 发现了这一规律]

被一群姑娘 [羡慕地 围住问长问短]

被他[合理地 设计出来了]

能加“地”作状语的名词和动词不多。形容词稍多一些，但总的数量仍较少，可以在词典中逐一标注。

由副词直接上升形成的 DP¹¹。只有以下一些能跟“被”字结构组合：程度副词中只有一个“稍微”；范围副词“都、全”；频度副词“又、再、再三”；类同副词“也”；状态方式副词“悄悄、暗暗、紧紧、死死、亲自、一齐、一直、相互”等等。其他大部分副词，如表语气的、表时间的、表否定的等等都不能跟“被”字结构组合。例从略。

AP 充当状语的。我们粗略调查了 258 个形容词¹²，其中能作状语的有 102 个。这些形容词作状语，能否跟“被”字结构组合，涉及到的因素比较复杂。我们仍然采用逐个标明的办法。不能跟“被”字结构组合的如“安心 便宜 不少 不幸 长 迟 高 光荣 足够”等等；能组合的如“安全 广泛 精确 重 秘密 少 早”等等。

MP 充当状语的。一般都能跟“被”字结构组合。如：“被他一把抓住、被他一次看完了、被他一下子逮着了六只麻雀、被他三天就砍光了”等等。

TP 充当状语的。大都不能跟“被”字结构组合。特别是表示将来时间的，如：“明天、明年、下星期”等等，但也有 TP 充当状语，能跟“被”字结构组合的例子。如：“被他昨晚上吃光了”。这里似乎主要是语义因素在起作用，没有多少句法条件。我们仍采用逐个标记的办法。

PP 充当状语的¹³。不能跟“被”字结构组合的有“被 比 对 对于 给 跟 关于 和 让 替 同 为 自 自从”等等；能跟“被”字结构组合的有“按 按照 把 本着 除 从 趁 根据 据 连 往 向 以 用 在”等等。也采用逐个标记的方法。这里例从略。

总的来说，表示动作行为的方式状态的状语成分，跟动词的关系紧密，结构上一般也就易于跟动词处在同一个层次。而表示时间、语气、动作关涉对象等的状语成分，跟动词的关系则松散一些，结构上一般也就倾向于远离中心动词。状语的次序问题比较复杂，涉及到的因素很多，还有待进一步研究。由于本文着眼于计算机如何理解汉语 VP 的结构层次和结构关系，而非用计算机来生成汉语句子，因此可以假定计算机要处理的输入句子都是正确的“被”字句，实际上也就很少会有上面那些不合法的状语成分出现了。述宾式 VP 也一样，假定跟“被”字结构组合的 VP 已经肯定是合法的。从这个角度看来，在掌握语言知识的精确程度方面，理解比生成的要求稍低一些。

7. 联合式 VP

联合式 VP 可以看成是由上面各种类型的 VP 同类组合得到的。不管有没有连接词，我们都暂且把联合的前后两项限制为是同类 VP，以简化处理。这样，上面那些结构形式的 VP 跟“被”字结构组合的限制条件，直接移植到联合式 VP 上就可以了。

8. 连谓式 VP

两个 VP 在结构关系上不属上面我们提到的任何一种 VP 组合方式，就可认为是连谓式 VP。能跟“被”字结构组合的连谓式 VP，必须满足一个必要条件：即连谓两项中的后项 VP 能跟“被”字结构组合。而充当连谓后项的 VP 能不能跟“被”字结构组合，就可以参照上面已经给出的，“被”字结构对各类 VP 的约束条件来判断。

上面我们对八类 VP 跟“被”字结构组合的约束条件进行了粗略的分析。归纳出来的这些约束条件基本上就可以帮助计算机正确判定 PP VP1 VP2 组合的结构层次。例如：

- (1) (疯狗)被他/pp 拿铁链/vp 拴住了/vp
- (2) (我)被他/pp 跑过来/vp 打了一拳/vp
- (3) (我)被/p 分配到公司/vp 作翻译/vp
- (4) (我)被老师/pp 批评/vp 写了检查/vp

例(1)–(4)都可先分析出小语段，均为 PP VP1 VP2 模式的排列式。

例(1)中前一个 VP “拿铁链”是个简单形式的述宾构造 VP，不能跟“被”字结构组合；后一个 VP “拴住了”是由述补结构充当中心成分的附加式 VP，满足跟“被”字结构组合的条件。这样，很容易就判定应按 a 式进行组合：

[被他 [拿铁链 拴住了]]

例(2)跟例(1)的情况类似，也应按 a 式组合：[被他 [跑过来 打了一拳]]

例(3)中前一个 VP “分配到公司”是述补结构，满足跟“被”字结构组合的条件；后一个 VP “作翻译”的中心动词“作”即不能跟“被”字结构组合，整个 VP 自然不能跟“被”字结构组合。因此按 b 式组合：

[[被 分配到公司] 作翻译]

例(4)跟例(3)类似，“写了检查”不能跟“被”字结构组合，整个短语应按 b 式组合：

[[被老师 批评] 写了检查]

可见，根据上面总结的那些条件规则，基本可以回答 PP VP1 VP2 格式的结构组合层次问题了。但要判定其内部结构关系，仅有上面那些规则还是不够，还需要进一步考察 VP VP 组合形成不同结构关系的限制情况。

§ 2

这一节就再来看看 VP VP 构成不同结构关系的限制条件

VP1 跟 VP2¹⁴组合可能形成述宾、述补、联合、连谓、主谓五种结构关系。前四种结构关系造成的短语功能类是 VP；最后一种主谓关系造成的短语功能类是 DJ。

汉语中相当多的动词排斥谓词性主语¹⁵，如：“吃、逮捕、打、跑、收购、讨厌、推荐”等等，都不允许其主语是谓词性成分。这样的动词肯定是不能占据 VP2 位置的。VP1 从形式上不易给出明确的限制条件。包含“被”字结构的 VP1 有的可出现在主语位置上。例如：

被警察罚款意味着经济损失

单从表层线性形式上看，这个例子也是 PP VP1 VP2 模式的排列：

被警察/pp 罚款/vp 意味着经济损失/vp

从结构层次上来说，是按 b 式组合：

[[被警察 罚款] 意味着经济损失]

不同于上文其他按 b 式组合的例子，这个短语的内部结构关系是主谓。因为这类组合情况少见（象“意味着”这样的动词不多），而且从结构类型上说也是一种弱势组合，我们对这类情况就不详细专门讨论了。下面有关两项 VP 组合的讨论都不涉及到构成主谓结构关系的情况。

结构关系的判定实际上是跟结构层次的判定同时进行的。结合 a、b 两种组合层次歧义的判定，我们可以把 VP1 分成非状中式 VP 和状中式 VP 两类，来考察 VP1 VP2 构成不同结构关系的限制条件。

（一）VP1 是非状中式 VP

VP1 是非状中式 VP，我们就可以相应地跟 PP VP1 VP2 格式按 a 式组合的情况联系起来考虑。这种情况下，VP1 跟 VP2 可能形成的结构关系有述宾、述补、联合、连谓四种。下面分别考察不同结构关系对两项 VP 的限制。

1. VP1 VP2 构成述宾关系

VP1 能跟 VP2 构成述宾结构关系的句法条件比较严格。VP1 的主要动词必须是谓宾动词。我们调查了 915 个动词¹⁶，谓宾动词有 259 个，占 28%。在这 259 个动词中，只有 44 个动词是只能带谓词宾语的，占全部的 5%。如：“企图 打算”等等；其他 215 个动词既可带谓词性宾语，也可带体词性宾语。如：“同意 参加 支持 赞成”等等；VP1 本身一般得是简单形式或附加形式，不能是其他复杂形式。做宾语的 VP2 主要受跟 VP1 语义搭配的限制，形式上难以归纳出约束条件。上文已经说明过，除极少数情况外，VP1 跟 VP2 构成述宾结构关系的 VP，一般不能跟“被”字结构组合。

2. VP1 VP2 构成述补关系

两个动词性成分直接结合，不依靠“得”连接，要构成述补结构关系，限制条件也是比较严格的。VP1 一般得是能带补语的动词的简单形式，可在词典中逐一说明；VP2 必须是能充当补语成分的那些动词，比如：“死、丢、到、倒、伤、完”等等，以及趋向动词“上、下、进、出、出来，下去”等等，VP2 基本上是一个封闭集，也可以在词典中一一注明。

3. VP1 VP2 构成联合关系

两个动词性成分不依靠连词连接直接组合形成联合关系的 VP，一般说来也是有很多限制的。但这些限制不太容易句法化，比较笼统。这里我们从简化处理考虑，规定 VP1 跟 VP2 各自的结构方式、所属动词次类、基本语义类都应该相同。比如：被敌人 [烧光了 抢光了]。“烧光了”跟“抢光了”就满足上述条件，内部是联合关系。本文开头所举例(2)也属同样的情形。

4. VP1 VP2 构成连谓关系

关于两个动词性成分组合形成连谓结构关系，上一小节我们已经从消极排除的角度给出了笼统的限制条件。这里进一步来看看对 VP1 和 VP2 有什么形式上的具体限制。

VP1 VP2 构成连谓关系，VP1 为“来、去”时可以是单个动词，如：“来打球、去看电影”，其他动词一般不能是单个动词形式¹⁷。而“来、去”形成的 VP 不能跟“被”字结构组合，综合这两个条件，就得到 VP1 不能是简单动词形式的限制。此外，当 VP1 是表反复问的联合形式时，VP1 一般跟 VP2 形成主谓结构关系而非连谓关系。如：“打不打由你决定”。只有 VP1 是“来、去”时，才有可能形成连谓结构，如：“去不去买菜”，可以作为特例处理。这样，我们就限制 VP1 不能是这种有疑问标记的联合形式。连谓后项 VP2 主要是要满足跟 VP1 搭配的语义限制，从形式上不太容易找到制约条件。就我们针对的问题而言，VP2 必须要满足 § 1 中给出的，能跟“被”字结构组合的那些条件。

(二) VP1 是状中式 VP

VP1 是状中式 VP。我们只简单讨论状语成分由“被”字结构充任这一种情况，相应地就可以跟 b 式组合联系起来考虑。这种情况下，VP1 不能再带宾语、补语。VP1 跟 VP2 也就不能构成述宾、述补结构，只可能有联合、连谓两种结构关系了。

VP1 跟 VP2 要构成联合关系，按上文对联合结构的规定，要求 VP2 也必须是由“被”字结构充当状语的状中式 VP。如本文开头所举例(5)。

VP1 跟 VP2 要构成连谓关系，VP2 就不能也是由“被”字结构充当状语的状中式 VP（那样 VP1 就跟 VP2 构成联合结构了），同时也不可以是能跟“被”字结构组合的那些 VP，因为如果 VP2 能跟“被”字结构组合，整个组合就成了按 a 式切分的短语了。如：“被警察抓住关进了监狱”中的 VP2 “关进了监狱”能跟“被”字结构组合，它就不能按 b 式切分，跟前面的“被警察抓住”构成连谓结构，而只能是按 a 式进行层次切分，“关进了监狱”先跟“抓住”构成连谓结构，再跟 PP 组合：[被警察 [抓住 关进了监狱]]。

至此，我们把有关 PP VP1 VP2 组合的句法限制条件做了一个概括总结。现整理如下：

- (1). 若 VP2 能跟“被”字结构组合，整个结构右向切分：a. [PP [VP1 VP2]]；
- (2). 若 VP2 不能跟“被”字结构组合，整个结构左向切分：b. [[PP VP1] VP2]；
- (3). 按 a 式切分，两项 VP 之间可以有述补、联合、连谓三种结构关系，
述补和联合都有严格的形式限定条件，
不属述补和联合关系的就判定为连谓结构关系；
- (4). 按 b 式切分，整个结构内部可能有联合、连谓两种结构关系，
联合关系有严格的形式条件限制，
不属联合关系的就判定为连谓结构关系；

根据这些限制条件，我们基本上就不但可以正确标界其结构体（是 a 式组合还是 b 式组合），而且可以判断其内部结构关系（是述宾、述补、联合、连谓中的哪一种）了。如：

- (1). (我)被他/pp 拿绳子/vp 捆了起来/vp
- (2). (我)被他/pp 派往雷区/vp 挖地雷/vp

从句法上讲，例(1)“拿绳子”是简单形式的述宾构造 VP，不能跟“被”字结构组合，“捆了起来”可以组合。结构层次很容易定为按 a 式组合：[被他 [拿绳子 捆了起来]。由 § 2(一)的分析可知，“拿绳子”跟“捆了起来”不可能形成述宾、述补、联合等结构关系，只能是连谓关系。例(2)“派”本身可以跟“被”字结构组合，“派往雷区”是述补结构，也满足跟“被”字结构组合的条件；“挖地雷”是简单形式述宾结构，不能跟“被”字结构组合。层次也就很容易判定为是按 b 式组合：[[被他 派往雷区] 挖地雷]。同样，“被他派往雷区”跟“挖地雷”也只能形成连谓关系。

从语义上讲，这两句因为结构关系清晰，因而内部语义关系也较容易判断。(1)句可以理解为“我被他捆了起来”+“他拿绳子”。对译的英文¹⁸就是：

I have been bounded by him with rope.

(2)句可以理解为“我被他派往雷区”+“我挖地雷”。对译的英文是：

I am dispatched to mine field by him to dig landmine.

§ 3

本文通过分别归纳 PP VP 组合和 VP VP 组合的限制规律，基本上消解了计算机在分析

PP<被>+VP1+VP2 格式时碰到的歧义问题。但上文所举例句都是比较典型的“被”字句。对实际语料中的一些特殊情况，或者没什么特殊但超出上文我们归纳的范围之外的情况，这里还有必要做一些说明。

“被”是比较特殊的介词，它在一定条件下允许宾语省略。如：

(1). 被/p 塞进麻袋/vp 扔进了长江/vp

同时“被”也可以带谓词性宾语。如：

(2). 被/p 砍头/vp 吓晕了/vp

例(1)(2)在格式上同形，但结构方式上迥异。例(1)是“被”的宾语省略，可以看成是本文所讨论的格式；例(2)“砍头”是“被”的宾语，根本不在我们讨论的范围之内。实际上，我们认为，例(2)的更为规范的说法应该是“被砍头的情景吓晕了”。如果这样表达，就是很普通的“被”字句格式了。计算机可以正确处理。本文在分析被字句格式歧义时，注重考虑的是结构类型上的普遍性和一般性，对实际语料中“被”字句用法的复杂性则只略有兼顾，不做全面详察¹⁹。

此外，对实际的汉英机器翻译系统而言，仅分析到上两节那样的程度，结构层次跟结构关系都清楚了，但实际上还没有真正深入到语义层的理解。最终还可能处理不了。譬如：

(3). (我)被他/pp 拦住/vp 不让走/vp

从句法上讲，这里例(3)的结构层次和结构关系也比较容易判断。跟§2中例(2)一样，按b式切分，是连谓关系。但从语义上看，却跟例(2)不同，是“我被他拦住”+“他不让我走”。计算机对此类现象的判断理解就碰到困难了。它很可能把例(3)跟例(2)做同样地解释。因为的确还有同形的句子可以象(2)那样解释。如：

(4). (我)被他/pp 蒙住眼睛/vp 不能看书/vp

例(4)应按b式切分，意思是“我被他蒙住眼睛”+“我不能看书”。对这样的语义歧义，排歧方法恐怕还得是挖掘分布条件。只不过要求更高，条件就更难找了。上面(3)、(4)这两个例子，似乎连谓后项VP“让”跟“能”的差异是影响语义解释的主要因素。限于篇幅，只能另文详细讨论了。

有关语义上的排歧问题涉及的因素比较复杂。跟人理解外语有水平高低差别一样，计算机对自然语言的理解也是由少到多，由浅到深，逐步提高的。目前我们处理得相对较好的是在短语的结构定界和结构关系的判定上，语义分析还有待将来做深入研究。我们可以把已经掌握的知识先告诉计算机，在计算机运用这些知识处理实际语料的过程中，一方面可以检验我们总结的语言知识正确与否；一方面又可以发现新的难以解决的问题。这就可以引导我们一步步去探索语言结构的规律，形成一个良性的信息反馈循环系统。

附注：

¹ 本文对现代汉语的词和短语都贯彻功能分类的思想，有关词和短语功能分类的详细情况，以及本文所用的功能标记符号说明可参见：

-
- 俞士汶（1993）《关于计算语言学的若干研究》，载《语言文字应用》1993.2；
俞士汶（1992）《关于汉语短语结构体系及其描述方法的说明》内部资料；
詹卫东（1996）《现代汉语 VP 的结构定界和结构关系判定》，北京大学硕士学位论文。
- ² 我们也称这种歧义格式为准歧义格式。参见詹卫东（1996）。
- ³ 《动词用法词典》词目录中除去“标志着”、“意味着”两词，共得 1221 个动词。
- ⁴ 参见朱德熙（1982）《语法讲义》§ 13.3，商务印书馆，1982 年版。
- ⁵ 及物动词指那些能带受事、结果、对象等等真宾语的动词，只能带施事宾语、处所宾语或时量动量数量等宾语类型的动词，都属不及物动词。如：走了三个人，去广州，咳嗽了一下。其中“走”、“去”、“咳嗽”都不是带真宾语，属不及物动词。
- ⁶ “看”是个同形词（不同音）。读阴平时，是“看护”的意思，可以说“被小王看着”；读去声时，而是“观看”的意思时，就不能说“被小王看着”。
- ⁷ 重叠式的形式语法学界有不同的看法，有的人不承认“v—v、v 了 v”等也是动词重叠式。
- ⁸ 离合式在形式上是相对凝固的短语组合，概念上相当于一个词。参见朱德熙（1982）《语法讲义》，商务印书馆。
- ⁹ AP、DJ 分别指形容词性短语和主谓短语，是短语的功能分类标记。参见附注 1。
- ¹⁰ DP、MP、TP、PP、SP 分别指副词性短语、数量词短语、时间词短语、介词短语、处所词短语。可参见附注 1 的说明。
- ¹¹ 副词的内部分类主要参考了马真（1988）《简明实用汉语语法》，北京大学出版社。本文对副词的考察范围也基本局限在该书中所论述的副词。
- ¹² 我们选用的是中科院计算所二室汉英机器翻译模型系统词典中的形容词。词典中对形容词充当状语的能力已经做了描述。机器翻译系统的词典是根据北大计算语言所编制的《现代汉语语法电子词典》（电子版本）自动生成的。该系统目前还在实验阶段，规模较小，共有 258 个形容词。
- ¹³ 考察介词的范围可参见马真（1988）P17~18。
- ¹⁴ 这一节用 VP1 和 VP2 来表示两项 VP 组合的情况，以区别于 PP VP1 VP2 格式中的 VP1 和 VP2。
- ¹⁵ 我们粗略考察了中科院计算所二室汉英机器翻译模型系统词典中的全部 915 个动词，其中 94% 的动词排斥谓词性成分充当主语。不排斥的动词只占 6%，有下面这些：
安排 避免 变成 表示 不如 产生 称 成 成为 当 得 得到 等于 定 发生
分为 符合 过去 花 恢复 回答 叫 决定 开始 可能 可以 浪费 能 能够
批准 取得 涉及 省 是 属于 提高 通过 同意 突出 忘 吸引 显得
消灭 需要 研究 要 引起 影响 准备
- 这些动词允许谓词性成分充当主语。如：游泳安排在下午进行。
- ¹⁶ 调查的动词范围同附注 15 的说明。
- ¹⁷ 象“出国学习”这样的连谓构造，前项 VP“出国”可以看成单词，也可处理为述宾式短语。我们倾向于采用后一种处理办法。
- ¹⁸ 目前计算机能正确分析出汉语短语的结构层次和结构关系，但译文形式尚不能达到很好的效果。这里给出的译文是经过人工润饰的。
- ¹⁹ 有关“被”字结构较为全面的用法描写可参看李珊（1994）《现代汉语被字句研究》，北京大学出版社；李临定（1986）《现代汉语句型》，P205~225，商务印书馆。