

# 北京大学

## 硕士学位论文

### 现代汉语 VP 的结构定界和结构关系判定

#### **Determining Boundaries and Constructional Relations of Verb Phrase in Contemporary Chinese**

姓 名: 詹卫东  
学 号: 19320036  
系 别: 中国语言文学系  
专 业: 现代汉语  
研究方向: 计算语言学  
导 师: 陆俭明教授 俞士汶教授

一九九六年五月十八日

## 现代汉语 VP 的结构定界和结构关系判定

### 摘 要

本文首先按语法功能标准将现代汉语短语进行了全面分类，对其中的动词短语（VP）做了详细的结构和分布描写。在此基础上，分析了计算机对 VP 进行结构定界和结构关系判定时产生歧义的情况，包括自圉型歧义和他圉型歧义，真歧义、伪歧义和准歧义等不同类型。

解决歧义问题的关键是找出 VP 的句法分布条件。

本文以跟“被”字结构相关的 VP 为例，具体分析了跟“被”字结构相关的 VP 的分布限制条件，较好地解决了“被”字句中 VP 的结构定界和结构关系判定的歧义问题。

最后，本文将归纳所得的语言知识加以形式化，交付计算机处理使用，使上述研究结果可以得到检验和进一步改进。

## **Determining Boundaries and Constructional Relations of Verb Phrase in Contemporary Chinese**

### **Abstract**

Firstly, this paper puts forward a classification of phrase in Contemporary Chinese according to the standard of distribution, and particularly, describes the constructions and the functions of Verb Phrase(VP).

Secondly, the author analyses the ambiguity of determining by computer boundaries and constructional relations of VP . They can be classified from two different perspectives. As viewed from the influence of ambiguity, ambiguous VP can be classified into two kinds: the closed-ambiguous type and the opened-ambiguous type; and viewed from differentiated types of relation between model ambiguity and concrete ambiguity, ambiguous VP can be classified into three kinds: the true-ambiguous type, the pseudo-ambiguous type and the quasi-ambiguous type.

Then, the author gives an example of determining boundaries and constructional relations of the construction “PP VP VP” in order to introduce how to solve ambiguous problems by the method of giving the restrained conditions of VP in certain contexts.

At last, the language knowledge about the restrained conditions of VP in certain contexts is formalized for a Chinese-English Machine Translation Prototype system.

目 录

第一章 引 言.....	1
第二章 关于现代汉语 VP .....	3
§ 2.1 VP 的内部构造.....	3
§ 2.2 VP 的外部功能.....	4
§ 2.3 分布条件.....	6
第三章 VP 的结构定界问题 .....	7
§ 3.1 可能发生 VP 结构体定界歧义的排列式.....	7
§ 3.2 自囿型结构体定界歧义.....	8
§ 3.3 他囿型结构体定界歧义.....	8
§ 3.4 真歧义.....	9
§ 3.5 伪歧义.....	10
§ 3.6 准歧义.....	11
第四章 VP 内部结构关系判定问题 .....	13
§ 4.1 他囿型结构关系歧义.....	13
§ 4.2 自囿型结构关系歧义.....	13
第五章 举例: PP VP VP 组合 .....	16
§ 5.1 PP VP 组合中 VP 要满足的条件.....	17
§ 5.2 VP VP 构成不同结构关系的限制条件 .....	24
第六章 语言知识的形式化表示.....	27
§ 6.1 复杂特征集与合一运算.....	27
§ 6.2 词的属性描述 .....	28
§ 6.3 短语的属性描述 .....	28
第七章 分析结果示例.....	30
结 语.....	36
附 注.....	37
参考文献.....	39
致 谢.....	40
附录一 现代汉语词和短语功能分类标记体系.....	41
附录二 涉及到 VP 的短语规则 (部分) .....	42
附录三 能跟“被”字结构组合的 VP .....	45

## 第一章 引言

中文信息处理经历字处理、词处理阶段，理论上和实践上都取得了令人满意的成果，目前已发展到用计算机进行短语处理的时期。但短语处理也碰到一些困难，突出的就是自然语言存在的大量歧义带来的短语定界困难，以及短语跟短语之间关系判定的困难。解决这些难题，首先需要我们人对汉语短语系统先有个清楚正确的认识，然后才能以一定的方式教给计算机掌握，使计算机在短语级也能达到较高的处理水平。

汉语语法学界对汉语短语的研究由来已久，但在短语的性质、短语的分类、短语的语法地位等一些基本问题上，至今仍然没有形成一致的看法。已有的研究成果无疑会对计算机处理有很多帮助，但面向计算机处理，在理论认识上和具体的语言现象描写上，都还是会跟传统的面向人的语法研究有不同。本文作为面向计算机进行汉语语法研究的一次尝试，集中讨论有关现代汉语动词短语（VP）的结构定界和结构关系判定的问题，归纳计算机对VP进行结构定界和结构关系判定时发生困难的不同类型，并结合具体实例的分析，相应地给出我们的处理策略。对那些不易说清的有关短语基本性质以及短语分类系统的问题，我们下面仅做背景性的简略说明。

以往对短语进行分类大都采用的是结构标准，但对一些结构“特别”、带有形式标记的短语，如“的”字结构、介词结构等，又另类处理，内部显得不是很整齐。而采用功能分类一般又只简单地把短语分成名词性短语、形容词性短语、动词性短语几类，分得过粗了一些<sup>[1]</sup>。我们对短语进行分类，跟词一样，也全面系统地采用分布特征（distribution）即语法功能作为分类标准<sup>[2]</sup>。

按结构标准分的定中、状中、述宾、述补、联合、连谓等结构类，很容易根据其分布特征并结合考虑中心成分的功能类（词类）来确定其功能类。比如定中结构归入名词短语、状中结构可分别归入副词短语、形容词短语和动词短语等等；介词结构本来就有功能类的意味，不妨仍沿用旧名为介词短语；数与数之间的组合比较特殊，跟一般的数量结构有明显功能差异，我们单列数词短语一类。“的”字结构、主谓结构的归属会发生一些问题。对“的”字结构，我们按其不同的句法功能性质并参考内部组成成分的功能类，分别归入名词性短语（如：买的、红的、老师的）、形容词性短语（如：高高的）和副词性短语（如：忽然的）三类；主谓结构的短语兼有体词性和谓词性，归入哪一类都不太合适，我们单独立类，只是没有与之相对应的词类。这样，我们得到如下现代汉语短语的功能分类体系<sup>[3]</sup>：

名词短语	时间词短语	处所词短语	数量短语	数词短语
动词短语	形容词短语	副词短语	介词短语	主谓短语

在我们的描写体系中，一般短语可以由单个词直接上升得到（如：动词“吃”可以看成是动词短语“吃”）；也可由词与词组合上升得到（如：动词“吃”跟名词“饭”组合得到动词短语“吃饭”<sup>[4]</sup>）。只是很少有介词能单独充当句子成分，大多数介词都要带宾语构成介词短语才能充当功能成分（“被”是例外）。介词短语一般必须由介词加上其他成分组合得到。还有主谓短语也不能由某类词直接上升得到，必须是组合形式。但是很明显，短语可以由词直接上升得到的做法也是贯彻功能分类思想原则的结果。词和短语作为语法单位，

其功能意义大于实体意义，词跟短语的形式界限在有些场合是可以模糊对待的<sup>[5]</sup>。从这一点上说，也跟以往的结构分类很不一样。模糊词和短语界限的做法，是结构分类体系在理论上所不能容许的。

上述关于短语基本性质及其分类体系的认识，即是本文讨论现代汉语 VP 的基础。需要进一步说明的是，本文的研究基本上在抽象的句法结构层面上进行，处理语料是受限的（restrained），而非真实的（real）<sup>[6]</sup>。对有关现代汉语 VP 的语言事实的把握，主要是采用研究者个人内省（introspective judgement）的方式。在分析归纳 VP 的各种内部组成、外部功能以及分布条件的基础上，我们尝试把语言知识加以形式化描述，交付实际运行的汉英机器翻译系统处理，通过实验结果反馈的信息来更新我们对 VP 的认识，改进以往的归纳，从而一步步提高计算机处理 VP 的能力<sup>[7]</sup>。

## 第二章 关于现代汉语 VP

考察 VP 可以从两方面来看。一是 VP 作为一个组合体，由于短语构造的递归性 (recursive)，它的组成成分中肯定含有 VP。从这个角度向里观察 VP，就是看 VP 的内部构造；二是 VP 作为一个组成成分，跟别的语类发生组合，构成的结构体有可能不属 VP，而属其它功能类。从这个角度向外观察 VP，就能更全面地概括 VP 占据的句法位置。

下面我们就从这两个方面，用推导式 (derivation)<sup>[8]</sup> 的形式来列举 VP 的句法分布情况。推导式中用到的符号标记参见附录一。

### § 2.1 VP 的内部构造

根据 VP 内部组合不同的形式特征，可将 VP 分为下面三组八个小类。

#### I. 带有显性形式标记的 VP

##### (1) 附加式

指动词或动词短语后附动态助词形成的 VP。动态助词是封闭的类，可以认为是这类 VP 的形式标记。例如：

vp → vp u      吃/vp了/u，吃完/vp了/u，坐/vp着/u，来/vp过/u

##### (2) 重叠式<sup>[9]</sup>

指动词以特定的方式重叠形成的 VP。例如：

vp → v v              看/v看/v  
vp → v m<一>v        看/v一/m看/v  
vp → v u<了>v        看/v了/u看/v

##### (3) 离合式<sup>[10]</sup>

指一些动词内部语素间关系松散，可以插入助词、数量词、趋向动词等成分造成的离合形式的 VP。例如：

vp → v u g            见/v了/u面/g，洗/v过/u澡/g  
vp → v u m g        睡/v了/u一/m觉/g  
vp → v v<起> g v<来>    洗/v起/v澡/g来/v

#### II. 定项组合的 VP

这一类 VP 内部组成成分项数是有限的确定的，都为两项组合，两项之间有的可以有特征词如“得”“不”等。包括下面三类结构方式。

##### (4) 述补式

包括有标记形式和无标记形式两种。有标记的含“得”“不”，属组合式述补 VP；无标记的一般是单个动词带补语的紧凑形式，属粘合式述补 VP<sup>[11]</sup>。例如：

vp → vp v            搬/vp出去/v，捆了/vp起来/v<sup>[12]</sup>，打/vp死/v  
vp → vp a            丢/vp光/a  
vp → vp pp          坐/vp在椅子上/pp，开/vp往北京/pp  
vp → vp u<得> ap    扫/vp得/u很干净/ap，洗/vp得/u干净/ap，吃/vp得/u光/ap

vp -> vp u<得> vp      逼/vp得/u到处借钱/vp  
 vp -> vp u<得> dj      打/vp得/u小王直求饶/dj  
 vp -> vp d<不> a      吃/vp不/d光/a

(5) 述宾式

由述语VP带NP、MP、VP、AP、DJ等等不同类型的宾语构成。例如：

vp -> vp np      做/vp作业/np  
 vp -> vp mp      看了/vp三天/mp  
 vp -> vp vp      企图/vp逃跑/vp  
 vp -> vp ap      爱/vp干净/ap  
 vp -> vp dj      同意/vp小王去北京/dj

(6) 状中式

中心语一定是VP，状语可由PP、DP、TP、SP、AP、MP等等充当<sup>[13]</sup>。例如：

vp -> pp vp      把书/pp拿走/vp  
 vp -> dp vp      刚/dp走/vp  
 vp -> tp vp      晚上/tp看/vp  
 vp -> sp vp      屋里/sp坐/vp  
 vp -> ap vp      认真/ap学习/vp  
 vp -> mp vp      一把/mp抓住/vp

III. 不定多项组合的VP

这一类VP内部组成成分项数可以是不确定的多项（两项或两项以上）。从技术处理的角度考虑，我们将其推导式写成两项组合，把多项并列分析为两项组合的套叠构造<sup>[14]</sup>，包括下列两类结构方式。

(7) 联合式

包括有标记的和无标记的两种情况。有标记的联合式VP含“和”、“并”等能连接谓词性成分的连词；无标记的是两个类似的VP并列连用。例如：

vp -> vp vp      [唱歌/vp跳舞/vp游泳/vp]/vp打球/vp  
 vp -> vp c vp      提出/vp并/c坚持/vp ，唱歌/vp或/c跳舞/vp

(8) 连谓式<sup>[15]</sup>

从结构形式上看，连谓式跟无标记的联合式VP同形。例如：

vp -> vp vp      [跑上去抓住小偷]/vp[狠狠打了两下]/vp

上面对VP的内部构造做了一个粗略的分类描写。每类的内部都可以根据分析需要再区分出更细的次小类来。譬如“扫得干净”跟“扫得很干净”这两个例子，虽然形式上都属述补式VP，但前者是表可能的，后者是表结果状态的，又可细分为两个小类。这里我们只大致说明有关VP最一般的结构形式，对每种结构形式的内部差别未做描写，而且也尚未列出全部的VP结构式。详见§2.2说明。

## § 2.2 VP 的外部功能

上面说明VP的内部构造时，已经把VP主要出现的句法位置基本上列举出来了。如：



VP 在述补、述宾构造中占据中心述语位置；在状中构造中占据中心语位置等等。从功能分布的一般意义上说，能占据那些句法位置都可称为 VP 的外部功能。这里要谈到的 VP 的外部功能，是想把 VP 能占据的其它一些句法位置做进一步说明。特指 VP 在这些句法位置上跟其它语类成分发生组合关系，构成的组合体不是 VP。下面我们仍用推导式来描述。

1. VP 占据主谓短语的谓语位置

- dj -> np vp 例: 老鼠/np 过街/vp
- dj -> vp vp 例: 游泳/vp 有益于健康/vp
- dj -> ap vp 例: 快乐/ap 在向我们招手/vp
- dj -> tp vp 例: 秋天/tp 是收获的季节/vp
- dj -> dj vp 例: 上进心强/dj 是年轻人的特点/vp

2. VP 占据主谓短语的主语位置

- dj -> vp ap 例: 学习/vp 很努力/ap

VP 充当主谓短语的主语时，谓语成分一般可以是 VP 或 AP，谓语是 VP 的可参见上面的例子。

3. VP 占据名词短语的中心语位置

- np -> np vp 例: 公路/np 建设/vp , 技术/np 改革/vp
- np -> ap vp 例: 经济结构的/ap 调整/vp

4. VP 占据名词短语的定语位置

- np -> vp np 例: 驾驶/vp 技术/np , 游泳/vp 姿势/np

5. VP 占据名词性“的”字结构的核心位置

- np -> vp u<的> 例: 带枪/vp 的/u

以上就是 VP 参与组合形成非 VP 短语类时 VP 的分布情况。我们可以看到，在上一节中，推导式左部的功能标记都是 vp，而这里推导式左部则是其他功能类标记，如：NP、DJ、AP 等。综合这两部分内容，就大致反映出现代汉语 VP 的分布面貌了。

补充说明一下，上面描述的 VP 分布，仍不足以概括现代汉语 VP 的全部句法表现。跟 VP 相关的推导式实际上也应包含类似下面这样的组合情况。如：

- tp -> vp f 例: 发展/vp 中/f , 游泳/vp 前/f
- dj -> np v<是> vp v u<的> 例: 他的病/np 是/v 想家/vp 想/v 的/u
- vp -> v<是> pp vp u<的> np 例: 是/v 在北京饭店/pp 举行/vp 的/u 招待会/np
- vp -> v<是> vp u<的> np 例: (他)是/v 坐/vp 的/u 飞机/np
- vp -> d<一> vp d<就> vp 例: (他)一/d 躺下去/vp 就/d 睡着了/vp
- vp -> p<被> np u<所> vp 例: 被/p 谎言/np 所/u 迷惑/vp

这些 VP 我们可以暂且称之为“特殊”分布形式，它们或者一般不在短语构造中担任中心成分；或者不易象上文所举 VP 那样做简单的结构描写。其他“特殊”的情况我们不多举例，这几例就足够说明 VP 在实际语料中分布的复杂性，以及概括归纳工作的困难了。尽管系统在实际处理时必须包含这些形式，但本文讨论的 VP 可以暂且不涉及这些“特殊”情况，只分析上文那些“一般”的 VP 类型。有关 VP 分布的更为详细的情况可参见附录二。

仅仅象上面这样列出 VP 的结构组合式和 VP 参与组合的结构式，只是描述 VP 分布状况的一个基础。要对实际语料中出现的 VP 的分布以及 VP 组合规律做归纳说明，还必须进

一步考察 VP 的分布环境，或者叫做分布条件。

### § 2.3 分布条件

所谓分布条件，就是指一个语类处在特定分布环境中需满足的条件。理论上可以包括句法条件、语义条件、语用条件等不同层次的条件。本文着重讨论 VP 在其分布环境中需满足哪些句法条件。举例来说：

(1) 我住在中国哈尔滨

正确理解的结构层次应该是：

[dj 我 [vp 住 [pp 在 [np 中国 哈尔滨]]]]<sup>[6]</sup>

“住”带 PP “在中国哈尔滨”构成述补式 VP 作整个 DJ 的谓语。

但计算机会产生错误的分析：

[dj 我 [vp [vp 住 [pp 在 中国]] 哈尔滨]]

VP “住在中国”带 NP 宾语“哈尔滨”构成述宾式 VP 作整个 DJ 的谓语。这种错误产生的原因，就是计算机不知道 VP 处在述宾结构的述语位置应满足什么条件。事实上，现代汉语中如果一个 VP 本身是带介词短语作补语的述补式 VP，它就不能再出现在述宾结构的述语位置上了。这是短语构造的结构条件限制之一。

(2) 很快恢复了生产

对这句话，结构层次应该是：[vp [ap 很 快] [vp 恢复了 生产]]

但对其中的 VP 语段“恢复了生产”的内部结构关系，计算机却会作两种理解，一种是正确地理解成述宾关系；一种就是可能错误地理解成连谓结构关系。这也是计算机对两个 VP 组合成更大的 VP 短语时，内部为何种结构关系，相应地要满足什么样的条件缺乏认识造成的。一般说来，两个 VP 组合，如果前一个 VP 是谓宾动词(如“恢复”)且没有带宾语，而后一个 VP 又是单个动词，那么后一个 VP 充当述宾结构 VP 的宾语的机会远远高于充当连谓结构 VP 的后项成分。

从上面所举的两个简单的例子可以看出，要让计算机正确识别短语中的 VP 结构体，以及判定内部结构关系，都要先搞清楚 VP 的分布条件，并让计算机能够掌握这些条件作为正确判别的依据。

一般语法书大都罗列了 VP 可能的用法，但给出条件的不多。跟以往面向人的语法研究相比，面向计算机的语法研究，对分布条件的要求更迫切、更苛刻。我们不能只满足于笼统地说 VP 能占据什么样的句法位置，还得说明 VP 要符合什么条件才能占据那些句法位置。换言之，就是不仅得说 VP 有哪些用法，更要紧的是还得说 VP 无哪些用法，尽管“言有易，言无难”。比如上一个小节我们列出了 VP 参与组合的基本结构形式，VP 可以充当主谓短语的主语，但显然不是任何 VP 都能有做主语的能力。这不仅跟 VP 的性质有关，同时也跟充当谓语成分的短语的性质有关。考察这里面的限制条件，也就是说明 VP 的分布条件，乃是我们下面要做的工作中最重要最关键的部分。

本章对 VP 做了鸟瞰式描述。下面几章我们将讨论 VP 结构体定界和结构关系判定碰到的困难以及我们的处理策略。几乎所有的分析环节都跟具体一类 VP 的分布条件密切相关。

### 第三章 VP 的结构定界问题

计算机分析自然语言最大的障碍就是歧义。排歧需要人先总结归纳出各种语言成分的用法条件。在进行这步工作之前，先对歧义问题的原因和面貌有个清楚的认识，无疑会使进一步归纳分布条件的工作更有针对性。本文准备首先在跟 VP 相关的结构层次歧义和结构关系歧义问题上做一番探讨。

计算机分析短语句子的语法结构时发生结构层次歧义，也就意味着结构体定界困难，主要是两个以上符号串的序列在组合时有多种可能性。我们以三个符号串的排列式来说明跟 VP 分析相关的定界歧义问题。

有一点需要说明，有些可能认为是结构层次歧义的情况(至少计算机的确是这样认为的)，其实是由结构体内部组成成分本身是同形异义词(包括同音词和兼类词等不同情形)造成的。例如：“会开了”可以有两种层次构造：

(1) [会 [开 了]]

(2) [[会 开] 了]

按(1)切分，“会”是名词，“会议”的意思，连带“开”的意思就是“开会、举行会议”；按(2)切分，“会”是助动词，“学会”的意思，连带“开”的意思就是“开汽车、驾驶某种交通工具”。两种不同的层次构造，完全由“会”这个同音词造成。

这种歧义情况不在我们的讨论范围之列。

#### § 3.1 可能发生 VP 结构体定界歧义的排列式

在考察机器可能碰到的 VP 结构体定界问题时，我们的研究对象首先是抽象的功能标记的序列，也就是所谓类型 (type) 排列式，而非实例 (token) 序列。如果从具体的实例出发，必然会碰到类似上面“会开了”那样的歧义情况，就要解决汉语中数量不少的词的兼类和同形词等等问题，这会使问题过于复杂而不便处理。因此我们从类型出发，考察短语类组合层面的歧义情况。

假设有 A、B、C 三个功能标记，它们连续排列为 ABC 形式。所谓可能发生结构定界歧义，是指存在两种组合可能：(1) B 跟 A 先组合后再跟 C 组合能形成更大的组合体；或者 (2) B 跟 C 先组合后再跟 A 组合也能形成更大的组合体。如果只存在上述两种情况中的一种，或者 A、B、C 三者根本不能组合成更大的结构体(即不能构成合法的语法形式)，则排列式 ABC 没有结构定界的问题。举例说明如下：

##### 1. VP U NP

这样三个功能类的排列形式，只有一种可能的组合方式即[[VP U] NP]，因为 U 是后定位功能成分，只能跟前面的成分组合，不可能跟其后面的成分发生组合关系。这种排列不会有定界问题。如：[[看 了] 三场电影]

##### 2. PP NP NP

PP 跟 NP 不可能发生结构关系，NP 跟 NP 发生组合关系后形成的结构体也不可能跟 PP 组合成更大的结构体。这样，这三个类的排列也不存在定界问题。



- a1 外部功能标记为 VP [喜欢 [看 京剧]], [提高 [修理 技术]]
- a2 外部功能标记为 DJ [游泳 [治好了 他的关节炎]]
- b [[VP VP] NP]
  - b1 外部功能标记为 VP [[贯彻 执行] 党的方针政策]
  - b2 外部功能标记为 NP [[骑马 射箭] 技术]

## 2. NP VP NP

有两种可能的组合方式，两种可能的功能标记：

- a [[NP VP] NP]
  - 外部功能标记为 NP [[方言 调查] 课], [[语法 分析] 方法]
- b [NP [VP NP]]
  - 外部功能标记为 DJ [小王 [看了 十本书]]

例 2 的两个功能标记分别都对应着唯一的一种组合方式，是典型的他围型组合歧义，而例 1 的三个标记中，VP 对应着两种不同的组合方式。也就是说，例 1 的歧义类型中有自围型的成分，组合方式不同但整个组合体的功能类一样，只是例 1 同时又包含了别的功能类，从大的方面说是他围型组合歧义，从局部看，是自围型歧义。

从上面两小节的分析可知，三个符号串的排列式造成的组合歧义，整体看来可以分为自围型和他围型两种歧义类型。但落实到具体某个排列式，很可能会有交叉。把一个排列式的歧义机械地归入哪种类型，并不是我们的目的，重要的是要弄清楚一个有组合层次歧义的排列式，组合歧义跟结构体整体功能的关系是怎样的。

### § 3.4 真歧义

上面我们从不同的层次组合是否导致整个结构体的外部功能标记不同这个角度，讨论了跟 VP 相关的层次组合歧义的两种不同类型。下面几个小节，我们从另一个角度，即排列式属类歧义还是实例歧义，来分析 VP 结构体定界问题的不同情况。

首先我们谈谈真结构层次歧义。

所谓真结构层次歧义，是指功能类组合时存在的歧义可以投射 (mapping) 到一个具体的自然语言形式上，即类的歧义可以在语言中很容易地找到同形异构异义的实例。例如：

#### 1. VP VP U<了>

上面这三个功能类的排列存在两种组合可能：

- a [VP [VP U<了>]] 如： [挤上巴士 [跑 了]]
- b [[VP VP] U<了>] 如： [[引起 争吵] 了]

a 跟 b 对应的两个实例分别都是单义的，只有一种层次结构。而下面的实例就是歧义短语了。层次上可以有两种不同的构造，意思上也可以有不同的理解。如：

拄着拐杖走了

可以理解为 a [拄着拐杖 [走 了]]

也可理解为 b [[拄着拐杖 走] 了]

做 a 式切分，意为“走了”，是“拄着拐杖走的”；做 b 式切分，可以表示以前不能“走”，现在能“拄着拐杖走”这样的意思，表示一种变化。

还有一种情形是在结构层次上可以做两种切分，但意义理解上没有什么差别。也就是通常所说的多切分的情况。如：

带着一家很快就回到上海了

可以按 a 切分 [带着一家 [很快就回到上海 了]]

也可按 b 切分 [[带着一家 很快就回到上海] 了]

两种层次都表示一样的意思。

从上面例子可以看出，真歧义的排列式，它投射产生的短语实例，在结构层次切分和意义理解上，存在三种情况。其中一种情况是真歧义排列式的充分必要条件。即同形异构异义的实例。上面这个排列式歧义是自圈型的。下面我们再举一个他圈型的真歧义排列式的例子。

## 2. VP D<不> AP

有两种组合可能： a [VP [D AP]] 如： [办事 [不 认真]]

b [VP D AP] 如： [洗 不 干净]

整个结构体按 a 式组合，功能类是 DJ，主谓结构，VP 作主语；按 b 式组合，功能类是 VP，述补结构。上面这两个实例都是单义的。我们可以再找两个例子，是同形异构异义的。如：

写不好 a [写 [不 好]]，不写也不好

b 这个字我 [写 不 好]

解释不清楚 a 论文对语言事实的描写很详细，但[解释 [不 清楚]]

b 小王[解释 不 清楚]事情的原因

这两个短语都是歧义实例。不同的组合层次，意义不同，整个结构体外部功能类也不相同。

## § 3.5 伪歧义

所谓伪歧义，是指功能类组合时有多种可能的方式，但映射到实例时，不同的组合方式没有意义理解上的不同；或者只有唯一的合理的切分，亦即实例事实上都是单一结构单一意义的。这也就是说，对类而言有多种组合可能，对实例而言则或者多切分或者单一切分，意思理解上都没有歧义。例如：

### 1. DP VP NP

从抽象的类的组合来看，这个排列式也有两种层次构造：

a. [DP [VP NP]]

b. [[DP VP] NP]

按 a 式组合，是副词性短语修饰一个动词性短语，整个结构体是状中式构造；按 b 式组合，则是副词性成分先修饰一个动词性短语，形成的状中式 VP 再带上一个名词性成分构成述宾结构 VP。下面的实例就可以有这样两种切分。

已经学过语言学 a [已经 [学过 语言学]]

b [[已经 学过] 语言学]

就这个短语而言，a、b 两种切分都可以接受。意义理解上几乎没有差别。是一般所说的多切分结构。但下面这个例子则只有一种合理的组合方式。

很看了几场电影      a    [很 [看了 几场电影]]

                         \* b    [[很 看了] 几场电影]

必须是“很”修饰整个述宾结构构成状中式 VP。不能是“很”先修饰“看了”，“很看了”不成结构。对这个例子而言，只有唯一的正确的层次组合，不是两可的。再比如：

## 2. PP VP NP

跟上面的排列式一样，也可以有 a [PP [VP NP]] 和 b [[PP VP] NP] 两种组合可能。但对应的实例通常却都是只能按 a 式组合的。如：

[在林子里 [追 野兔]]

[从南方 [跑到 北方]]

[把桌子 [撞断了 一条腿]]

也有似乎两可的例子，两种切分意思上没有什么差别。如：

按原订计划处理小王的问题

可以按 a 式组合： [按原订计划 [处理 小王的问题]]

也可按 b 式组合： [[按原订计划 处理] 小王的问题]

上面这两个排列式，都是伪歧义的情形。机器分析这类排列式时，应该很容易确定 VP 的前后界限。比如我们可以规定上两例都按 a 式组合。这样，就是对类的组合而言也不存在歧义了。这也正是我们称之为伪歧义的原因。一般说来，伪组合歧义都是自囿的。

## § 3.6 准歧义

所谓准歧义，是指功能类排列式在组合时有多种层次切分的可能性，对应到具体实例，一般只能选择其中唯一一种合理的组合方式，也只有一种意义理解。也就是说，有类的歧义，而具体实例则没有歧义。举例说明如下：

### 1. PP VP VP

有两种组合方式      a    [PP [VP VP]]

                         b    [[PP VP] VP]

a 式 PP 后面的两个 VP 先构成一个结构体，然后被 PP 修饰，整个组合成状中结构 VP。b 式 PP 先修饰相邻的 VP，构成的状中式 VP 再跟后面的 VP 组合，整个结构体的功能类可能是 VP，也可能是 DJ。是他囿型歧义结构。

由这个排列式映射的短语实例，要么按 a 式组合，要么按 b 式组合，一般只有一种切分方式，没有同形异构异义的现象。下面对比举例说明。

1A [把被子 [拿出去 晒一晒]]

1B [[把饭 吃完] 做作业]

2A [被黑社会 [敲诈 勒索]]

2B [[被老师 批评] 写了检查]

3A [在操场上 [跑步 打球]]

3B [[在操场上 跑步] 成了我的业余爱好]

4A [沿长江 [走了几个省 考察了一些历史遗迹]]

4B [[沿长江 旅游] 会有很多新发现]

上面四组八个短语，A 都是按 a 式切分的，B 都是按 b 式切分的。每句都是单义的。按 A 式切分的四句都是状中结构 VP；按 b 式切分的四句中，1B 和 2B 整个短语功能类为 VP，内部是连谓结构关系；3B 和 4B 整个短语功能类为 DJ，内部是主谓结构关系。

我们再来看一个自围型的准歧义例子。

## 2. VP NP NP

有两种切分可能： a [VP [NP NP]]

b [[VP NP] NP]

两种切分方式构成的组合体的功能类都是 VP。a 式是两个 NP 先构成定中结构再做 VP 的宾语；b 式是 VP 带 NP 宾语后构成的述宾结构又带一个 NP 宾语，述宾结构套叠构成述宾结构 VP。可以看下面的实例：

1A [出任 [公司 总经理]]

1B [[送 图书馆] 三箱书]

2A [买 [猪皮 大衣]]

2B [[运 济南] 十吨钢材]

这四个例子中，1A 和 2A 按 a 式切分；1B 和 2B 按 b 式切分。1A “公司总经理”作“出任”的宾语。1B 则“送图书馆”带“三箱书”作宾语。每句都只有一种切分方式，都是单义实例。

以上从排列式属类歧义还是实例歧义这个角度，把结构层次歧义分为真歧义、伪歧义、准歧义三类，有助于我们对歧义进行更深入的研究，为寻找有效的排歧条件打下了一个基础。三类中，准歧义其实是真歧义里的一种特殊情况。或者说，某个真歧义排列式，如果很难找到一个同形异构异义的实例，我们就可以认为它是一种比较特殊的情况，姑且把它叫作“准歧义”。从这个意义上讲，准歧义排列式也都是真歧义排列式。二者仅仅是排列式映射出来的同形异构实例数量多少的差别，没有实质上的不同。



## 第四章 VP 内部结构关系判定问题

上一章我们以三个功能类排列的组合层次歧义为研究对象，对跟 VP 相关的结构体定界问题做了系统的分类说明。这一章我们假定 VP 结构体界限确定，来看看结构体内部组成成分之间可能存在的句法结构关系歧义。

我们以两个语言成分组成结构体的情况为例说明。结构体内部结构关系不同可能连带结构体的外部功能类即整个符号串的功能标记也不同；也可能结构关系不同外部功能标记相同。这跟上一章结构层次歧义的情形类似，也可分为自囿型和他囿型两种歧义类型。

### § 4.1 他囿型结构关系歧义

他囿型结构关系歧义指结构关系不同组合体功能类也不同的歧义情况。下面就来看看跟 VP 相关的一些他囿型结构关系歧义。这些歧义组合都不含组合后形成不同的结构关系而整个结构体功能类相同的情形。换言之，就是这些他囿型结构关系歧义组合内部不包括自囿型歧义的情况。

#### (1) VP NP

这是现代汉语中比较常见的一种歧义组合。VP 跟 NP 组合，内部可以有述宾和偏正两种结构关系，分属 VP 和 NP 两个功能类。例：

- a 述宾关系——VP 炸/vp 碉堡/np
- b 偏正关系——NP 研究/vp 机构/np

上面两个例子各自都是没有歧义的。而象“出租汽车”、“学习文件”、“炒饭”这样的例子，就都是有两种理解的歧义短语了。

#### (2) NP VP

NP 跟 VP 组合可以有下面两种结构关系，分属两个功能类。

- a 主谓关系——DJ 大会/np 正式开始/vp
- b 定中关系——NP 城市/np 管理/vp

上面两个例子各自都只有一种意思。而象“企业赞助”这样的例子，内部就有主谓和定中两种可能的结构关系，是歧义短语的实例。

#### (3) AP VP

AP 跟 VP 组合也可以有两种结构关系，分属两个功能类。

- a 状中关系——VP 认真/ap 工作/vp
- b 主谓关系——DJ 干旱/ap 带来了饥荒/vp

这类结构关系歧义一般少有歧义实例。每个具体的 AP VP 组合一般都只有一种理解。

### § 4.2 自囿型结构关系歧义

自囿型结构关系歧义指结构关系不同但整个结构体功能类相同的歧义情况。严格说来，两个成分组合成 VP 的结构关系歧义，没有完全自囿型的。下面两个组合式例子，从整体上看，都属于他囿型歧义的范畴。但跟上一节纯粹的他囿型歧义不同，这里的例子是他囿

型结构关系歧义同时又包含了自囿型歧义的情况。

(1) VP VP

VP 跟 VP 组合可以得到两种功能类 DJ 和 VP, 属他囿型歧义。但当组合体功能标记为 VP 时, 内部组合关系还有多种情况, 从这点上说, 就是自囿型歧义了。我们下面撇开 VP 跟 VP 组合成 DJ 的情况不管, 来看看 VP 跟 VP 组合成 VP 时有哪些结构类型。

VP 跟 VP 组合成 VP, 结构关系可以有下面四种情况。

- a 述宾关系: 赞成/vp 打排球/vp
- b 述补关系: 搬/vp 出去/vp
- c 连谓关系: 关灯/vp 睡觉/vp
- d 联合关系: 写字/vp 画画/vp

上面四个例子都是单义的, 而象“想出来”, 分别可以理解为述宾关系或者是述补关系, 就是歧义实例了。比较:

- a. 他在里面呆久了, 想出来。 ( 述宾关系 )
- b. 这个问题我终于想出来了。 ( 述补关系 )

再比如“骑马打球”这样的例子, 可以理解为“骑着马打球”, “骑马”是“打球”的伴随方式; 也可理解为“骑马”和“打球”两种运动并举。前者是连谓结构关系, 后者是联合结构关系, 也是歧义实例。

需要说明一下, 象“打下去”这样的例子, 也可以有两种理解:

- (I) 我们终于把上山的敌人打下去了
- (II) 我们两家不能再这样打下去了

“打下去”内部语义关系不同, “下去”在(I)中表示位移趋向; 在(II)中表示动作行为的延续。但从结构关系上说, 两种情况下都属述补关系。因此这样的例子不是我们讨论的结构关系歧义的情况。

(2) VP AP

VP 跟 AP 组合也可以得到两种功能类 DJ 和 VP。跟上面(1)的情况一样, 当 VP 跟 AP 组合成 VP 时, 内部结构关系还有不同, 局部可以看作是自囿型组合关系歧义。下面举例说明。

- a 述宾关系: 喜欢/vp 安静/ap
- b 述补关系: 洗/vp 干净/ap

VP 跟 AP 组合成 VP, 内部可能有述宾和述补两种结构关系。上两例都是单义的例子。而象“说清楚”, 就是可以有两种理解的歧义实例了。既可以理解为“说”的内容是“清楚”(例如: 我问他讲得清楚不清楚, 他说清楚), 也可以理解成“说”的结果是“清楚”(例如: 这个问题你得说清楚)。前者是述宾关系; 后者是述补关系。

由上面的分析不难看出, 两个成分组合产生结构关系歧义, 从歧义跟整个结构体的功能类的关系角度, 有自囿和他囿两种情况; 从类歧义跟实例歧义的关系角度, 有类似结构层次歧义的真歧义和准歧义的情况。结构关系自囿型歧义的组合式, 一般都属真歧义的情况(如这一节的例子); 结构关系他囿型歧义的组合式, 既有真歧义的情况(如上节例 1, 2), 也有准歧义的情况(如上节例 3)。结构关系歧义没有伪歧义。不存在一个歧义组合式有类歧义, 而投射出来的实例却全部只有一种组合关系一种意义理解的情况。

上面两章我们对 VP 的结构定界歧义和结构关系判定歧义分别作了分类描写。主要意图在于为排歧打下一个基础。首先通过上面的分析归纳，我们对 VP 分析面临的歧义问题有了更清楚的认识。更重要的是，针对不同的歧义类型，相应的处理策略就有不同的考虑。他围型歧义相对容易在更大的语境中自动消歧，而自围型歧义的消解对外部语境的依赖更复杂，难度也就更大一些。比如：“学习文件”是他围型结构关系歧义的实例。它在语境(1)“一份学习文件”中是 NP；在语境(2)“他们正在学习文件”中是 VP。这很容易根据句法分布环境的限制要求来确定。“踢破球”是自围型结构层次歧义的实例。它在语境(3)“踢破球不如踢新球好玩”中是[踢 [破 球]]结构；在语境(4)“踢破球会被罚款”中是[[踢 破] 球]结构。人区分这两种不同的意思，更多的恐怕是根据常识或者说是意义关联上的因素。鉴于不同类型的歧义在排歧上有上述不同的情况，我们在考虑对策时，就要在自围型歧义的组合式上多下工夫了。真歧义、准歧义、伪歧义三类情况比较，其中伪歧义可以统一规定处理方式，适用于所有的实例就行了；准歧义因为少有歧义实例，这本身就说明这类组合式的排歧条件相对好找一些；真歧义最复杂，排歧一般都要跨出结构体自身范围，必须在更大语境中找制约歧义的条件。

综合起来看，目前比较适宜选择自围型准歧义的 VP 组合式，重点解决其结构定界和结构关系判定问题。解决的途径仍是我们在第二章中强调过的，尽可能归纳出 VP 的分布条件。下面我们就以一个 VP 组合式为例，具体说明我们的排歧办法。

## 第五章 举例：PP VP VP 组合

本章我们归纳总结有关 PP VP VP 组合的语言知识,解决这个组合式结构层次和结构关系判定的问题。为使讨论集中,我们只考察 PP 是“被”字结构的情况。以下如不特别说明,PP 都指“被”字结构。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
- (6) 被/p 塞进麻袋/vp 扔进了长江/vp

上面 6 个例句,(1)---(5)都是 PP VP VP 模式排列,(6)是 P VP VP 排列。出于使问题简化的考虑,我们回避实际语料中存在的“被”带谓词性宾语构成 PP 的情况。比如:“被砍头吓晕了”。VP“砍头”充当“被”的宾语。本文把这种情形作为不规范短语来处理(更为规范的说法应该是“被砍头的情景吓晕了”)。这样,我们就不用去判断“被”字后面紧跟的 VP 是“被”的宾语还是真正的上面我们例子中的 VP。在例(6)中,VP“塞进麻袋”就不是“被”的宾语,而是在结构上受“被”修饰的中心成分。我们规定,PP VP 结构中,不管“被”有没有体词性宾语,后面的 VP 总是被 PP 修饰构成状中结构的。因此,下文 P 跟 PP 同等对待。我们的目的在于为计算机处理提供规则依据,着重考虑的是结构类型上的普遍性和一般性,目前并不强调对各种情况的具体用法<sup>[18]</sup>都能面面俱到(这应该是长远目标),因此下面讨论时所举例句也不多,能够说明问题就可以了。

上两章已经谈到,PP VP VP 排列式在结构层次歧义类型中,属自囿型准歧义,而其中的 VP VP 组合,结构关系则有真歧义的实例。就我们这里举的六例中的前五个来说,(1)(2)(3)结构层次上是按 a 式切分:[PP [VP VP]]; (4)(5)是按 b 式切分:[[[PP VP] VP]。都是按 a 式切分,(1)中 VP 跟 VP 之间是连动关系;(2)中 VP 跟 VP 之间是联合关系;(3)中 VP 跟 VP 之间是述补关系。结构关系各有不同。都是按 b 式切分,(4)内部是连动关系;(5)内部是联合关系。结构关系也有差异。要使计算机能准确判定(1)(2)(3)是 a 式切分,(4)(5)是按 b 式切分,而且相同的切分方式内部还有不同的结构关系,并不是件容易的事情。需要我们从句法条件上去找制约因素。

要确定 PP VP VP 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,我们可以把问题转化为两项组合的情况来考虑。事实上,上文举的组合歧义例子,最终都得落实到两项组合的问题上来解决。在语言结构中,两项组合是最基础最广泛的方式。多项组合一般都可用层次分析分解为两两组合的结构。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。对本章这个组合歧义问题,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 VP VP 三项组合的歧义问题自然也就解决了。因此,实际上问题就可以转化为:

- (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?

这也就是我们在 § 2.3 中谈到的 VP 分布条件问题。下面我们就一步步来找出分布条件。首先我们看看 PP VP 组合的限制要求情况。

## § 5.1 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, 可以有两种角度。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on't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<sup>[19]</sup>, 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). 复数主语动词                | 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 <sup>[20]</sup> | 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          |
| (14). 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, 用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”属性的取值就简单的填为“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，都属同样的问题，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”字结构组合在目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，我们就可以为每个具体的动词贴上一个标签，来标记它跟“被”字结构组合的能力。如果计算机也掌握了这样的语言知识，对下面例句，就能确定其结构层次了。如：

A. (王教授)被美国一所大学邀请参加了研讨会

先分析出小的结构段：被美国一所大学/pp 邀请/vp 参加了研讨会/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”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] 参加了研讨会]

意义理解上也跟切分情况吻合，“王教授”是“被邀请”，“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 VP VP 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，事实上，正如我们在第二章中描述的，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。请看对比的例子：

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。例如：

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，直接用主动形式表达。如：“小王吃不完”。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，否定词跟“被”的关系比较密切，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 VP VP 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 VP VP 组合的结构层次也是非常有帮助的。

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，但还是有述补式、状中式构造的 VP 不能跟“被”字结构组合，还有必要把那些不能组合的形式排除出去。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(能跟“被”字结构组合的动词词表参见附录三)。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 143 个，占 12%；不能以光杆形式但能以复杂形式跟“被”字结构组合的，有 316 个，占 26%。譬如：“看”。不能单说“被看”，但能说“被他看到、被看了几眼”等。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 要么是用在特殊场合，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，我们这里不去详细区分。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 <sup>[21]</sup>
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。我们采用后者。

### 2. 重叠式 VP

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 VP

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

趋向述补 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。本文不做详细讨论。

#### 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”字结构后的 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”）

如果 VP 含其他类型的体词性宾语，宾语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：

被他吃完了三大碗饭 \*被小王吃完了饭（受事宾语）

被他写秃了三十支笔 \*被他写秃了毛笔（工具宾语）

也有的宾语可以不含数量成分，也能跟“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）

被吓了一身冷汗 被吓出冷汗

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。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、DJ 等宾语类型。一般说来，带 AP、DJ 宾语的都不能跟“被”字结构组合。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。

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、SP 等。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。

由副词上升形成的 DP<sup>[22]</sup>。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<sup>[23]</sup>，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：“被他昨天晚上吃光了”。这里似乎主要是语义因素在起作用，没有多少句法条件。

PP 充当状语的<sup>[24]</sup>。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。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（understanding）汉语 VP 的结构层次和结构关系，而非生成（generation），而且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，实际跟“被”字结构组合的 VP 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 的限制约束条件来判断。

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可以帮助计算机正确判定 PP VP VP 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 VP VP 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 式进行切分：

[被他 [拿铁链 拴住了]]

例(2)跟例(1)的情况类似，也应按 a 式切分：[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 式切分：

[[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 式切分：

[[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 VP VP 组合的结构层次问题了。但要判定其内部结构关系，仅有上面那些规则还显不够，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。

## § 5.2 VP VP 构成不同结构关系的限制条件

为叙述方便，我们给两个 VP 加上下标，分别记作 VP1 和 VP2。

VP1 跟 VP2 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<sup>[25]</sup>，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 VP VP 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 式切分：

[[被警察 罚款] 意味着经济损失]

不同于上文其他按 b 式切分的例子，这个短语的内部结构关系是主谓。因为这类组合情况少见（参见 § 3.6 例 3B、4B），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 VP VP 组合是 a 式层次切分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<sup>[26]</sup>，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，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或至多是附加形式；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，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并列联合关系的 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 各自的内部结构方式、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。本章开头所举例(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<sup>[27]</sup>。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 § 5.1 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”字结构的 PP 充任一种情况，相应地就可以跟 b 式层次切分联系起来考虑。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。如本章开头所举例(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 VP VP 组合的句法限制条件做了一个初步的概括总结。根据这些限制条件，我们基本上就可以正确标界其结构体（是 a 式切分还是 b 式切分），并判断内部结构关系（是述宾、述补、联合、连谓中的哪一种）了。譬如：

(1). 我被他拿绳子捆了起来

(2). 我被他派往雷区挖地雷

从句法上讲，(1)句“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 式切分：[被他 [拿绳子 捆了起来]。“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。(2)句“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 式切分：[[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<sup>[28]</sup>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.

可对实际的汉英翻译系统而言，仅分析到上面这样的程度，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。譬如：

(3). 我被他拦住不让走

从句法上讲，(3)句的结构层次和结构关系也比较容易判断。跟(2)一样，按 b 式切分，是连谓关系。但从语义上看，却跟(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(3)当作(2)来解释。因为的确还有同形的句子可以象(2)那样解释。

(4). 我被他蒙住眼睛不能看书

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。

有关语义层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 VP 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。

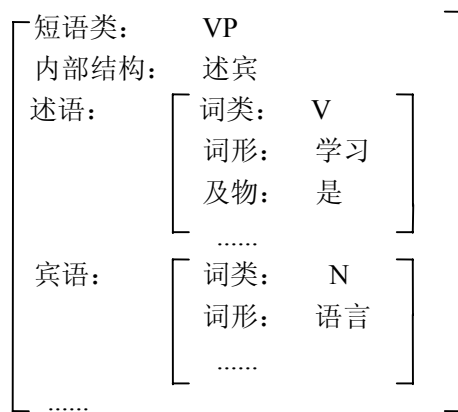
## 第六章 语言知识的形式化表示

用自然语言表述的句法规律，人能看懂，但由于自然语言表述的模糊性和不精确性，计算机掌握起来就比较困难了。我们必须把这些规律形式化（formalization），也就是把表义不严密的自然语言表述，换成用严格定义过的更加抽象的形式符号来表述。下面就把我们采用的形式化方法做一些基本的介绍。

### § 6.1 复杂特征集与合一运算

我们将语言知识形式化依据的语法理论是功能合一语法（Functional Unification Grammar）。这一计算语言学理论自八十年代由 Martin Key 提出以来，在理论上和技术上已经发展得比较成熟了。它最为显著的特点就是所谓复杂特征集（complex feature set）和合一运算（unification）。这二者结合运用，可以比较方便有效地对自然语言从词到句子加以形式描写。要对它们做详细全面的介绍，会涉及到很多细节问题<sup>[29]</sup>。我们这里仅就最基本的内容略作说明。

复杂特征集以特征结构（feature structure）的方式来描述语言成分的语法语义性质。特征结构是由[属性：值]对组成的集合。一个特征结构内部可以包括若干个[属性：值]对，这属于简单特征结构；也可以在一个特征结构内部再嵌套特征结构，就构成复杂特征结构。简单的说，[属性：值]对也可称为特征，可以看作是对所描述对象的语法语义各项性质的限制说明。在语言研究中经常讨论的语义特征、音位区别特征等等，也都是采用这种描述形式的。譬如语音系统中的一个音位，我们可以将它定义为一串区别特征的集合。对语言中的各种语言成分，我们都可以类似地用特征集合的形式对其加以定义。下面给出一个特征结构的例子：



“学习语言”的特征结构（部分内容）

由例子可以看出，特征结构是允许嵌套的。此外每个属性本身的性质还有差别，取值类型和取值范围也可不同。比如：“短语类”属性的取值范围是有限的十个，因为我们的系统只给出了现代汉语的十类短语；而“词形”这个属性取值范围显然就大得多。这两个属性的值都是一般的符号。“及物”属性的取值则只有“是”和“否”两个，非此即彼，属布尔

型值 (boolean)。我们在下文具体介绍词和短语的属性时再作详述。

合一运算就是把若干个特征结构合并成一个单独的特征结构,有点儿类似集合中的求并运算。如果两个特征结构中的各项特征相容,那么合一成功,得到的合一结果是原来几个特征集的并集;如果两个特征结构中的特征相抵触,那么合一失败,得到的结果就为空集(从这点上说跟单纯的集合求并运算又有些不同)。

合一运算可以方便地把原有的特征信息合并,构造新的特征结构;同时还能检查特征的相容性。从语言描写的角度看,也就是可以描述从词组合成短语语言成分的属性特征的继承和变化情况;以及判别哪些组合是允许的哪些组合是非法的。在下面 § 6.3 一节中,我们对短语结构规则加以限制约束,以及描述短语的属性特征,主要就是运用合一运算进行的。

## § 6.2 词的属性描述

所谓词的属性描述<sup>[30]</sup>,就是对词的用法加以详实的记录。一个记录就是一个复杂特征结构。这里我们以表(list)的形式来描述动词“想”的属性特征,用具体例子说明复杂特征集是怎样把有关词的语言知识加以形式化的。

“想”这个常用动词有多个义项。这里我们举其中三个来说明。“想 1”表示开动脑筋思索;“想 2”表示希望打算;“想 3”表示思念想念。不同的意思在用法上也有不同表现。词典中都作了详细记录:

- 想 {v1} [词类:v,谓词性主语:否,被:是,.....,名宾:是,动宾:是,形宾:否,.....,重叠:是,论元:2,语义类:自为] {主体:[语义类:人],客体:[语义类:抽象物]}
- {v2} [词类:v,谓词性主语:否,被:否,.....,名宾:否,动宾:是,形宾:是,.....,重叠:否,论元:1,语义类:自为] {主体:[语义类:人]}
- {v3} [词类:v,谓词性主语:否,被:否,.....,名宾:是,动宾:否,形宾:否,.....,重叠:否,论元:2,语义类:心态] {主体:[语义类:人],客体:[语义类:人]}

稍做比较不难发现,“想”不同义项对应的记录中有的属性特征取值一样,如“词类、谓词性主语”等;有的属性特征取值不同,如“被、名宾、重叠”等。不同的取值也就意味着用法上的差异。比如“想 1”可跟“被”字结构组合(“答案被他想出来了”),“想 2”跟“想 3”都不行;“想 1”跟“想 2”可以带 VP 宾语,“想 3”则不行;“想 1”不能带 AP 宾语,“想 2”则可以等等。

上面中括弧里有并列多项[属性:值]对,是简单特征结构;花括弧里“主体”属性的取值是用中括弧括起来的[属性:值]对,是复杂特征结构,用于描述一个动词的格关系(语义属性)。作为例子,这里只介绍了动词的部分属性。实际使用的词典中动词的属性描述比这还要丰富一些。此外还根据计算机处理的需要设置了一些标识符号。就不详细说了。

下面简要介绍怎样对短语的属性进行描述。

## § 6.3 短语的属性描述

词典中对每个具体语词的语法语义性质做了标记。规则库中的规则则对短语结构加以描述,包括短语的属性特征,以及搭配组合的语法语义限制。

对短语的属性,研究工作没有词做得多。我们目前只尝试使用了“内部结构、否定、被



动、语气、主语、中心语”等十多个。短语的很多性质是词向上组合时从词的属性继承得到的，也就是说词的部分属性，短语可以共享。下面以 PP 跟 VP 组合形成状中式 VP 为例，简略说明将有关短语的知识加以形式化的基本方法。PP 仅指“被”字结构。

```
vp -> pp vp $.内部结构=状中, $.被动=是, $.状语=%pp, $.中心语=%vp,
    %pp.格标=施事, %vp.被=是, %vp.否定=否,
    IF %vp.内部结构=述宾, %vp.动宾=是, %vp.宾语.短语类=vp FALSE,
    IF %vp.内部结构=组合式述补, %vp.补语.内部结构=单词 FALSE,
    IF %vp.内部结构=状中, %vp.状语.格标=施事 FALSE,
    IF %vp.内部结构=连谓 THEN %vp.后项.被=是 ENDIF, .....
```

规则开始部分是通常的短语结构推导式，用于描述短语的基本结构方式。后面的部分就是对短语属性的描述以及对组合条件的判断。“\$”表示整个短语，在这条规则中即指箭头左边的 VP。“%”用于指示箭头右边的标记的位置，在这条规则中“%VP”就指箭头右边第一个 VP。被逗号分隔开的等式都是合一等式。譬如“%VP.被=是”，就是拿动词的“被”属性（特征结构）跟一个布尔型的原子值“是”进行合一运算。如果计算机分析上文例句“答案被他想出来了”。在扫描到动词“想”时，计算机要判断这时该取“想”的哪一个义项，就可以根据这条合一运算的结果来决定。“想 1”，“被”属性的值为“是”，合一成功；“想 2”跟“想 3”“被”属性的值为“否”，合一失败。这样，就只有“想 1”能跟 PP 组成“被”字结构了。如果词典里再给出“想”的不同义项对应的英语词，分别是“think”、“want”、“miss”，计算机就可以挑选“想 1”相应的译词“think”而不是另外两个错译了。

IF 打头的合一等式用于表达一些复杂的组合形式的判断。我们在上一章中归纳出来的有关 PP VP 组合的结构制约条件，都可用这样的形式来严密表述。比如：当 VP 为连谓结构时，要求“%VP.后项.被=是”，即指连谓结构的后项 VP 必须能跟“被”字结构组合，整个连谓 VP 才能跟“被”组合成状中式 VP。“.”号相当于自然语言中表领属的“的”。其余等式也都对应着上一章我们用自然语言概括的那些有关的条件限制，就不一一说明了。

这一章我们简要介绍了将语言知识加以形式化的理论和方法的基本内容。词跟短语的属性描述都还有相当多的问题有待深入探讨。比如词的属性，设置多少属性项就是一个问题。属性描述要尽可能经济有效地反映词的实际用法，就要兼顾同类词的共性和具体词的个性。这需要在实践中不断探索<sup>[31]</sup>。再如短语的属性约束限制，比较突出的是概括过分（over-generalization）和概括不足（under-generalization）的问题。前者会导致计算机不能理解本是正确的句子；后者会造成计算机理解上大量的歧义现象。正如本文引言所说的，这需要根据计算机分析实际语料的结果，不断调整完善。本文最后一章，我们给出两个跟“被”字结构相关的 VP 的例子，来看看计算机目前能分析到一个什么样的程度。

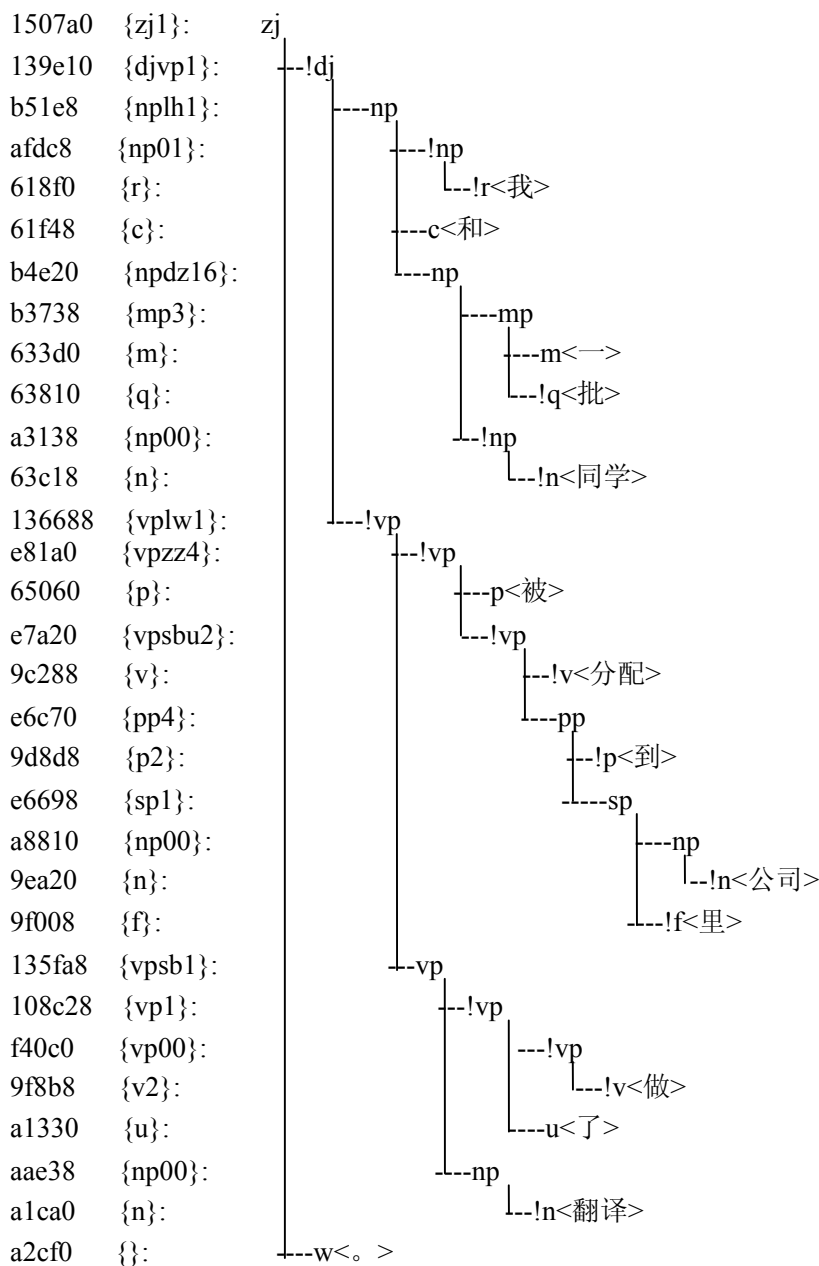
## 第七章 分析结果示例\*

示例一：

Source: 我和一批同学被分配到公司里做了翻译。

Target: I and a batch of classmates have been assigned to company to be translator.

== 源文根结点树 srcroottree ==



=== 译文根结点树 tgroottree ===

1507a0 {zj1}: SS  
 139e10 {djvp1}: ---!CS  
 b51e8 {nplh1}: ----NP  
 afdc8 {np01}: ---!NP  
 618f0 {r}: ---!R<I/我>  
 152cd8 {}: ----C<and/>  
 b4e20 {npdz16}: ----NP  
 b3738 {mp3}: ----NP  
 633d0 {m}: ----T<a/一>  
 63810 {q}: ---!N<batch/批>  
 153cd0 {}: ----P<of/>  
 a3138 {np00}: ---!NP  
 63c18 {n}: ---!N<classmate/同学>  
 136688 {vplw1}: ---!VP  
 e81a0 {vpzz4}: ---!VP  
 e7a20 {vpsbu2}: ---!VP  
 9c288 {v}: ---!V<assign/分配>  
 e6c70 {pp4}: ----PP  
 9d8d8 {p2}: ---!P<to/到>  
 e6698 {sp1}: ----NP  
 a8810 {np00}: ----NP  
 9ea20 {n}: ---!N<company/公司>  
 155968 {}: ----P<to/>  
 135fa8 {vpsb1}: ----VP  
 108c28 {vp1}: ---!VP  
 f40c0 {vp00}: ---!VP  
 9f8b8 {v2}: ---!V<be/做>  
 aae38 {np00}: ----NP  
 a1ca0 {n}: ---!N<translator/翻译>  
 151ca0 {}: ----W<“.” />

产生结点(b51e8:afdc8--61f48--b4e20) np: 我和一批同学

全局规则: {nplh1}

树形结构: np(!np(我) c<和> np(一批同学))

特征结构: [汉语短语分类:np,内部结构:联合]

产生结点(e81a0:65060--e7a20) vp: 被分配到公司里

全局规则: {vpzz4}

树形结构: vp(p<被> !vp(分配到公司里))

特征结构: [汉语短语分类:vp,内部结构:状中,被动:是,有宾语:否]{状语:p<被>,中心语:vp<分配>}

产生结点(135fa8:108c28--aae38) vp: 做了翻译

全局规则: {vpsb1}

树形结构: vp(!vp(做了) np(翻译))

特征结构: [汉语短语分类:vp,内部结构:述宾,有宾语:是]{述语:vp<做>,宾语:np<翻译>}

产生结点(136688:e81a0--135fa8) vp: 被分配到公司里做了翻译

全局规则: {vplw1}

树形结构: vp(!vp(被分配到公司里) vp(做了翻译))

特征结构: [汉语短语分类:vp,内部结构:连谓,被动:是,有宾语:否]{前谓:vp<分配>,后谓:vp<做>}

产生结点(139e10:b51e8--136688) dj: 我和一批同学被分配到公司里做了翻译

全局规则: {djvp1}

树形结构: dj(np(我和一批同学) !vp(被分配到公司里做了翻译))

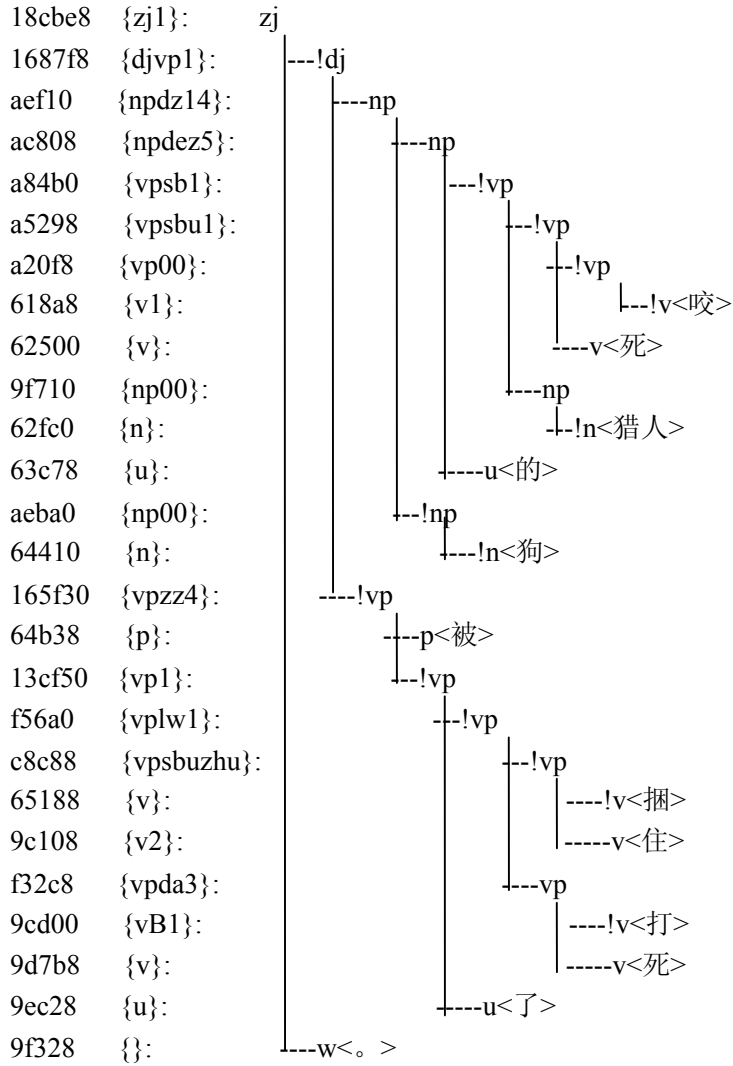
特征结构: [汉语短语分类:dj,内部结构:主谓,有宾语:否]{主语:np<我>,谓语:vp<分配>}

示例二：

source: 咬死猎人的狗被捆住打死了。

Target: Dog that bites hunter to death has been bound and has been beaten to death.

=== 源文根结点树 srcroottree ===



=== 译文根结点树 tgroottree ===

18cbe8 {zj1}: SS  
 1687f8 {djvp1}: ---!CS  
 aef10 {npdz14}: ----NP  
 aeba0 {np00}: ---!NP  
 64410 {n}: ---!N<dog/狗>  
 190820 {}: ----C<that/>  
 ac808 {npdez5}: ----VP  
 a84b0 {vpsb1}: ---!VP  
 a5298 {vpsbu1}: ---!VP  
 a20f8 {vp00}: ---!VP  
 618a8 {v1}: ---!V<bite/咬>  
 194a18 {}: ----PP  
 194388 {}: ----P<to/>  
 193820 {}: ----N<death/>  
 9f710 {np00}: ----NP  
 62fc0 {n}: ---!N<hunter/猎人>  
 165f30 {vpzz4}: ---!VP  
 13cf50 {vp1}: ---!VP  
 f56a0 {vplw1}: ---!VP  
 c8c88 {vpsbuzhu}: ---!VP  
 65188 {v}: ---!V<bind/捆>  
 192f30 {}: ----C<and/>  
 f32c8 {vpda3}: ----VP  
 18e5b8 {}: ---!V<beat/>  
 18f848 {}: ----PP  
 18f9b8 {}: ----P<to/>  
 18fba8 {}: ----N<death/>  
 192098 {}: ----W<“.” />

产生结点(aef10:ac808--aeba0) np: 咬死猎人的狗

全局规则: {npdz14}

树形结构: np(ap(咬死猎人的) !np(狗))

特征结构: [汉语短语分类:np,内部结构:组合定中]{定语:ap<咬>,中心语:np<狗>}

产生结点(13cf50:f5c70--9ec28) vp: 捆住打死了

全局规则: {vp1}

树形结构: vp(!vp(捆住打死) u<了>)

特征结构: [汉语短语分类:vp,内部结构:附加,有宾语:No]{附加语:u<了>,中心语:vp<捆>}

产生结点(165f30:64b38--13cf50) vp: 被捆住打死了

全局规则: {vpzz4}

树形结构: vp(p<被> !vp(捆住打死了))

特征结构: [汉语短语分类:vp,内部结构:状中,被动:是,有宾语:No]{状语:p<被>,中心语:vp<捆>}

产生结点(1687f8:aef10--165f30) dj: 咬死猎人的狗被捆住打死了

全局规则: {djvp1}

树形结构: dj(np(咬死猎人的狗) !vp(被捆住打死了))

特征结构: [汉语短语分类:dj,内部结构:主谓,有宾语:No]{主语:np<狗>,谓语:vp<捆>}

---

注\* 分析所用的汉英机器翻译系统由中科院计算所二室刘群设计。系统运行环境: Sun 工作站, Unix 操作系统, C++编程。

翻译系统在分析汉语源句后产生句法分析结果, 这里主要用一棵句法树来表示, 最左边的代码(如 1507a0 等)是计算机对规则所做索引的编码, 花括弧里是规则名, 同时我们还选取了一些子树的特征结构, 对分析这些片段时所用的规则以及分析得到的句法信息做了部分说明。生成英语译文只简单给出了一棵译文树。译文没有经过后处理, 时态和表达上还有一些毛病。总的说来, 计算机处理一个句子的中间环节是很多的。这里记录的只是其中比较重要的一小部分。大量的实际过程因为篇幅的限制都省略了。

## 结 语

本文的研究工作是在点面结合这个总体思想指导下完成的。

面上，我们描述了现代汉语 VP 的基本句法结构和分布面貌，并着重分析了 VP 结构体定界和结构关系判定的不同歧义类型；

点上，我们以 PP VP VP 组合为例，具体调查总结了现代汉语中 VP 跟“被”字结构组合的限制条件。根据这些条件限制，我们就可以判定：

- (1). 若后一个 VP 能跟“被”字结构组合，整个结构右向切分： a. [PP [VP VP]]；
- (2). 若后一个 VP 不能跟“被”字结构组合，整个结构左向切分： b. [[PP VP] VP]；
- (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；

因为是面向计算机处理自然语言的研究，本文还以一定的篇幅简单介绍了把语言知识形式化的基本内容。将上述自然语言表述的规则转化为精确定义的形式表达。所有的工作都通过计算机理解实际受限语料的结果来检验。

我们认为，遵循这样的研究思路，研究范围明确，针对性强，研究结果可检验，无论是将来深入研究还是拓宽研究，都可打下一个很好的基础。事实上，本文只集中讨论了 VP 的问题，但在分析实际语料时，要牵涉到其他所有的短语类。语言成分之间的系统相关性是十分显著的。在给 VP 定界时，跟 VP 同现的其他短语类 NP、AP、PP 等等，都会影响分析结果的正确性。要构造一个性能优良的语言分析规则系统，必须对每类短语的分布和句法限制条件都做调查总结。本文对 VP 的研究只是其中的一部分，而且基本上局限在句法层面，语义和语用的因素几乎没有考虑，所做的工作还多有缺欠，纵、横两个方面都有许多工作有待将来做进一步研究。



## 附 注:

- [1] 参见 吕叔湘 (1979)《汉语语法分析问题》，商务印书馆；  
李子云 (1986)《短语探讨》，载《安徽教育学院学报》社科版 1986.3。
- [2] 参见 朱德熙 (1982)《语法讲义》，商务印书馆；  
朱德熙 (1985)《语法答问》，商务印书馆。
- [3] 这里所说的“的”字结构统指“X的”。关于“X的”的分类以及“的”字结构的功能归类，参考了朱德熙 (1961)《说“的”》，载《中国语文》1961.2；  
短语的功能分类参考了：  
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。
- [4] 参见 王艾录 (1992)《词组和句子》，载《山西大学学报》哲社版 1992.1。
- [5] 参见 杨成凯 (1994)《关于汉语语法单位的反思---汉语语法特点散论之三》，载  
《汉语学习》1994.6；  
刘丹青 (1995)《汉语语法单位分级理论的再探讨》，载《汉语学习》1995.2；  
郭 锐 (1996)《汉语语法单位及其相互关系》，载《汉语学习》1996.1。
- [6] 关于受限汉语的思想可参见 俞士汶 (1995)《关于受限的规则汉语的设想》，载  
王均 主编《语文现代化论丛》山东教育社 1995 年版；  
鲁川 梁镇韩 (1987)《信息处理用规则汉语》，载《中文信息学报》1987.4。
- [7] 自 1994 年开始，中科院计算所二室与北大计算语言所合作开发汉英机器翻译系统。计算所二室负责系统设计和程序工作，北大计算语言所按二室给出的规范提供语言知识库。该系统从 1995 年 3 月开始进入调试阶段。笔者参加了全部调试工作。本文也正是在这项工作期间完成的。目前系统规模还不小，扩充和调试工作仍在进行。因此本文也可算作是一个阶段性的工作报告。
- [8] 推导式亦即产生式 (production)，其中“->”表示“定义为”或“由...组合而成”。另外还可称作重写规则 (rewriting rule)“->”表示“将...改写为...”。参见 Noam Chomsky 乔姆斯基 (1957) 著，邢公畹 等译 (1979)《句法结构》，中国社会科学出版社。
- [9] 重叠式的形式语法学界有不同看法，有的人不承认“v - v、v 了 v”等也是动词重叠式。
- [10] 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)。
- [11] 有关句法结构组合和粘合的观点，参见 朱德熙 (1982)。
- [12] 象“捆了起来”这样的 VP，一般认为是连谓结构。参见北大中文系现代汉语教研室编《现代汉语》教材，商务印书馆 1993。本文把它处理为述补结构。
- [13] 现代汉语中还有“讽刺说”、“上课睡觉”这样的例子。是 VP 作状语构成的状中式结构。我们把这种情况看作极少数特例，不作为一般的结构类型处理。
- [14] 有关句法结构套叠的观念，参见 陆俭明 (1990)《汉语句法成分特有的套叠现象》，载  
《中国语文》1990.2。
- [15] 连谓式 VP 包括一般的连动结构和递系结构。参见朱德熙 (1982)。本文只考虑两个 VP 连用的情况。象“称赞她很美丽”、“嫌这件衣服花哨”这样的短语，我们都处理为动词带主谓短语作宾语的构造。在我们的体系里，没有 VP 跟 AP 连用形成连谓结构 VP 的情况。
- [16] 为节省篇幅，本文不用下划线，而用括弧表的形式来标明层次。紧跟左括弧之后的标记标明该括弧内短语的功能类。跟正文用大写标记相区别，例句中都用小写标记。
- [17] 严格说来，DP VP VP 组合还有外部功能标记为 DJ 的情形，不完全是自围型歧义。如：  
[[常 跑步] 能长高] 整个结构体外部标记就是 DJ。
- [18] 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。
- [19] 《动词用法词典》词条目录中除去“标志着”、“意味着”两词，共得 1221 个动词。
- [20] 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。

[21] “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”。

[22] 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。

[23] 我们选用的是汉英机器翻译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。目前还在实验阶段，规模较小，共有 258 个形容词。

[24] 考察介词的范围可参见马真（1988）P17~18。

[25] 我们粗略考察了目前汉英机器翻译词典中的全部 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。

[26] 调查的动词范围同附注[25]，也是目前机器翻译用词典中的全部动词。

[27] 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。

[28] 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。这里给出的译文是人工润饰后的结果。

[29] 参见 冯志伟（1991）《Martin Key 的功能合一语法》，载《国外语言学》1991.2；

冯志伟（1990）《汉语句子描述中的复杂特征》，载《中文信息学报》1990.3；

沙新时 等（1993）《基于合一语法的通用句法分析器：设计与实施》，载

《中文信息学报》1993.2。

[30] 本文对词和短语所做的描述，都遵循中科院计算所二室刘群设计的汉英机器翻译系统的形式规范，在行文中略做了一些简化处理。特此向刘群先生表示感谢。

[31] 参见 俞士汶 等（1992）《现代汉语语法电子词典的概要与设计》

第三届中文信息处理国际研讨会会议论文 ICCIP'92 P186~191。

## 参考文献

(按作者姓氏拼音排序)

- [1] 曹敏 《计算机自动分析量词短语的方法及其规则》，载  
《中文信息学报》1990.1。
- [2] 董振东 《机器翻译中词典和文法的关系》，载《中文信息学报》1988.3。
- [3] 范继淹 《范继淹语言学论文集》，语文出版社，1986年版。
- [4] 冯志伟 《中文信息处理与汉语研究》，商务印书馆，1992年版。
- [5] 黄国营 《现代汉语的歧义短语》，载《语言研究》1985.1。
- [6] 李临定 《现代汉语动词》，中国社会科学出版社，1990年版。
- [7] 李临定 《现代汉语句型》，商务印书馆，1986年版。
- [8] 李子云 《汉语句法规则》，安徽教育出版社，1991年版。
- [9] 刘月荣等 《清华英汉机译系统中的歧义处理》，载  
陈力为 袁琦 主编《计算语言学进展与应用》，清华大学出版社，1995年版。
- [10] 陆俭明 《八十年代中国语法研究》，商务印书馆，1993年版。
- [11] 陆俭明 《陆俭明自选集》，河南教育出版社，1993年版。
- [12] 陆志韦 《汉语的构词法》，科学出版社，1957年版。
- [13] 陆致极 《计算语言学导论》，上海教育出版社，1990年版。
- [14] 吕叔湘 《汉语语法分析问题》，商务印书馆，1979年版。
- [15] 吕叔湘 《汉语语法论文集》，商务印书馆，1984年版。
- [16] 马希文 《以计算语言学为背景看语法问题》，载《国外语言学》1989.3。
- [17] 孟琮等 《动词用法词典》，上海辞书出版社，1987年版。
- [18] 钱峰 《计算语言学引论》，学林出版社，1990年版。
- [19] 王维贤 《动词小句的基本短语结构形式》，载《中国语文》1994.1。
- [20] 吴竞存 梁伯枢 《现代汉语句法结构与分析》，语文出版社，1992年版。
- [21] 郑怀德 孟庆海 《形容词用法词典》，湖南出版社，1991年版。
- [22] 朱德熙 《现代汉语语法研究》，商务印书馆，1985年版。
- [23] 朱德熙 《语法答问》，商务印书馆，1985年版。
- [24] 朱德熙 《语法讲义》，商务印书馆，1982年版。
- [25] 朱一之 王正刚选编《现代汉语语法研究的现状和回顾》，语文出版社，1991年版。

## 致 谢

首先我要衷心感谢我的两位导师陆俭明教授和俞士汶教授。他们渊博的学识、严谨的学风，给我树立了榜样，使我得以在三年研究生学习期间，学到了丰富的现代汉语和计算语言学专业知识，为今后的学习工作打下了坚实的基础。两位导师为我的毕业论文倾注了大量心血。俞老师给我提供了参加一个汉英机器翻译原型系统开发调试的机会，使论文的研究工作能够顺利完成；陆老师对论文的材料、观点和写作安排都给予了悉心指导。我深感获益匪浅。

北大中文系马真教授，北大计算语言所朱学锋副教授、段慧明老师、张芸芸老师、王惠老师、郭涛小姐，在我研究生学习期间，都给予我很多的帮助，提供了诸多的便利条件。在此一并表示诚挚的谢意。

北大计算语言所的陶晓鹏同学、周强博士，中科院计算所二室的刘群先生，在本文的研究和写作过程中，都曾给过我很多启发。也在此表示深深的感谢。

## 附录一 现代汉语词和短语功能分类标记体系

语素标记: g

词类标记\*: 名词 n 代词 r 处所词 s 时间词 t 方位词 f  
数词 m 量词 q 区别词 b 状态词 z 形容词 a  
动词 v 副词 d 介词 p 连词 c 助词 u  
语气词 y

短语标记: 名词短语 np 时间词短语 tp 处所词短语 sp  
数量短语 mp 数词短语 mcp 动词短语 vp  
副词短语 dp 介词短语 pp 形容词短语 ap  
主谓短语 dj

句子标记: 复句 fj 整句 zj

---

注\* 词类中还应包括叹词和拟声词两类。本文暂时没有处理包含叹词和拟声词的语料，所以在标记中没有列出。

## 附录二 涉及到 VP 的短语规则（部分）

- && {vp00} vp->!v :: \$.内部结构=单词
- && {tp7} tp->vp !f :: \$.内部结构=附加
- && {pp7} pp->!p vp :: \$.内部结构=介宾
- && {dp5} dp->!vp u<地> :: \$.内部结构=附加
- && {npdez5} np->!vp u<的> :: \$.内部结构=的字
- && {apsbu1} ap->!a v :: \$.内部结构=粘合述补
- && {apsbu4} ap->!a u<得> vp :: \$.内部结构=组合述补
- && {np3} np->!v g<性> :: \$.内部结构=单词
- && {npsz1} np->u<所> !vp :: \$.内部结构=所字
- && {npdz11} np->vp !n :: \$.内部结构=粘合定中
- && {npdz41} np->n !v :: \$.内部结构=粘合定中
- && {npdz42} np->ap !vp :: \$.内部结构=组合定中
- && {npdz43} np->r !vp :: \$.内部结构=粘合定中
- && {npdz44} np->sp !vp :: \$.内部结构=粘合定中
- && {vp1} vp -> !vp u<了> :: \$.内部结构=附加
- && {vp2} vp -> !vp u<过> :: \$.内部结构=附加
- && {vp3} vp -> !v u<着> :: \$.内部结构=附加
- && {vplh1} vp->!vp vp :: \$.内部结构=联合
- && {vplh2} vp->!vp w<“,” | “,” | “、” > vp :: \$.内部结构=联合
- && {vplh3} vp->!vp c vp :: \$.内部结构=联合
- && {vplh4} vp->!vp d<不|没> vp :: \$.内部结构=联合
- && {vplh5} vp->!vp d<还是> vp :: \$.内部结构=联合
- && {vplh6} vp->!vp w<“,” | “,” > d<还是> vp :: \$.内部结构=联合
- && {vplh7} vp->!vp d<没有> :: \$.内部结构=联合

- && {vplw1} vp->!vp vp :: \$.内部结构=连谓
- && {vpsbu1} vp->!vp v :: \$.内部结构=粘合述补
- && {vpsbu2} vp->!v pp :: \$.内部结构=粘合述补
- && {vpsbu3} vp->!v a :: \$.内部结构=粘合述补
- && {vpsbu4} vp->!v u<极了> :: \$.内部结构=组合述补
- && {vpsbu5} vp->!v d<不> v :: \$.内部结构=组合述补
- && {vpsbu6} vp->!v d<不> a :: \$.内部结构=组合述补
- && {vpsbu7} vp->!v u<得> d<很> :: \$.内部结构=组合述补
- && {vpsbu8} vp->!v u<得> r :: \$.内部结构=组合述补
- && {vpsbu9} vp->!v u<得> ap :: \$.内部结构=组合述补
- && {vpsbua} vp->!v u<得> vp :: \$.内部结构=组合述补
- && {vpsbub} vp->!v u<得> dj :: \$.内部结构=组合述补
- && {vpsb1} vp->!vp np :: \$.内部结构=述宾
- && {vpsb2} vp->!vp sp :: \$.内部结构=述宾
- && {vpsb3} vp->!vp tp :: \$.内部结构=述宾
- && {vpsb4} vp->!vp mp :: \$.内部结构=述宾
- && {vpsb5} vp->!vp mcp :: \$.内部结构=述宾
- && {vpsb6} vp->!vp q np :: \$.内部结构=述宾
- && {vpsb7} vp->!vp vp :: \$.内部结构=述宾
- && {vpsb8} vp->!vp ap :: \$.内部结构=述宾
- && {vpsb9} vp->!vp dj :: \$.内部结构=述宾
- && {vpsba} vp->!vp np np :: \$.内部结构=双宾
- && {vpsbb} vp->!vp np mp np :: \$.内部结构=双宾
- && {vpsbc} vp->!vp np dj :: \$.内部结构=双宾
- && {vpsbd} vp->!vp r :: \$.内部结构=述宾
- && {vpsbe} vp->!vp fj :: \$.内部结构=述宾
- && {vpzz1} vp->pp !vp :: \$.内部结构=状中
- && {vpzz2} vp->pp w<“,” | “,” > !vp :: \$.内部结构=状中
- && {vpzz3} vp->p !vp :: \$.内部结构=状中
- && {vpzz4} vp->ap !vp :: \$.内部结构=状中
- && {vpzz5} vp->dp !vp :: \$.内部结构=状中
- && {vpzz6} vp->r !vp :: \$.内部结构=状中
- && {vpzz7} vp->tp !vp :: \$.内部结构=状中
- && {vpzz9} vp->mp !vp :: \$.内部结构=状中
- && {vpzza} vp->sp !vp :: \$.内部结构=状中

- && {vpzzb} vp->c !vp :: \$.内部结构=状中
- && {djap4} dj->vp !ap :: \$.内部结构=主谓
- && {djvp1} dj->np !vp :: \$.内部结构=主谓
- && {djvp3} dj->tp !vp :: \$.内部结构=主谓
- && {djvp4} dj->sp !vp :: \$.内部结构=主谓
- && {djvp5} dj->r !vp :: \$.内部结构=主谓
- && {djvp7} dj->ap !vp :: \$.内部结构=主谓
- && {djvp8} dj->vp !vp :: \$.内部结构=主谓
- && {djvp9} dj->dj !vp :: \$.内部结构=主谓
- && {djdj7} dj->vp !dj :: \$.内部结构=主谓
- && {djshi1} dj-> v<是> np vp :: \$.内部结构=是字
- && {djshi2} dj-> v<是> np vp u<的> :: \$.内部结构=是字
- && {djshi3} dj-> d<不> v<是> np !vp u<的> :: \$.内部结构=是字
- && {djshi4} dj -> np !v<是> ap ::\$.内部结构=主谓
- && {djshi5} dj -> vp !v<是> ap ::\$.内部结构=主谓
- && {djyou1} dj->v<有> np !vp :: \$.内部结构=有字
- && {djyou2} dj->v<没有|没> np !vp :: \$.内部结构=有字
- && {djyou3} dj->v<有> q np !vp :: \$.内部结构=有字
- && {djcz1} dj->sp !vp(!vp np) :: \$.内部结构=存在
- && {djcz2} dj->np !vp(!vp np) :: \$.内部结构=存在
- && {djcz3} dj->sp !vp np :: \$.内部结构=存在
- && {djcz4} dj->sp !v u<着> q np :: \$.内部结构=存在
- && {djcz5} dj->np sp !vp np :: \$.内部结构=存在
- && {djcz6} dj->pp vp np :: \$.内部结构=存在
- && {djcz7} dj->pp !v u<着> q np :: \$.内部结构=存在



### 附录三 能跟“被”字结构组合的 VP

#### 一 能以光杆形式跟“被”字结构组合的动词

绑 包围 剥削 保护 保留 报复 逼 驳斥 踩 参考 操纵 拆 超过 抄  
吃 处罚 处理 传染 刺激 催 打 打倒 打破 代替 逮捕 耽误 挡 捣乱 得  
到 抵抗 颠倒 点 惦记 调查 叮 夺 发表 发现 罚 反对 粉碎 讽刺 奉承 否  
认 俘虏 攻击 鼓动 怪 忽视 怀疑 忌妒 检查 揭发 揭露 解决 解散 禁止 警  
告 拒绝 开除 控制 扣留 夸大 捆 扩大 利用 了解 领导 骂 埋怨 埋葬  
冒充 蒙蔽 虐待 排除 盘问 抛弃 碰见 批判 批评 批准 骗 聘请 破坏 欺  
负 欺骗 启发 牵连 强迫 抢 敲诈 撬 侵略 轻视 清理 驱逐 取消 杀 审  
查 审问 说服 偷 推荐 拖延 挖苦 威胁 侮辱 误会 误解 吸引 陷害 限制  
消除 消化 消灭 消弱 欣赏 压迫 压制 掩盖 邀请 咬 议论 引诱 隐瞒 遇  
到 遇见 原谅 允许 糟蹋 责备 侧 占领 掌握 镇压 指定 制止 抓 转移  
准许 捉 阻挡 阻止

#### 二 必须以复杂形式才能跟“被”字结构组合的动词

安 安排 按 熬 拔 掰 摆 搬 抱 背 比 闭 避 避免 编 表达 表示  
拨 布置 擦 裁 采 测量 测验 插 查 搀 缠 尝 唱 吵 炒 扯 称 冲  
抽 锄 除 穿 传 串 闯 存 搓 搭 打扮 打击 戴 带 逮 掸 弹 当 当  
做 捣 等 递 垫 钓 调<sub>1</sub> 调<sub>2</sub> 调动 调整 叠 盯 钉<sub>1</sub> 钉<sub>2</sub> 顶 丢 动员 冻  
斗 逗 读 堵 堆 夺取 躲 刹 发动 发挥 发扬 发展 翻译 反映 访问 放  
分 分配 分析 缝 扶 改 改造 改正 盖 赶 搞 搁 割 跟 跟随 雇 刮  
关 管 滚 裹 喝 合 哄 糊 花 画 换 集合 集中 记 记录 继承 夹 捡  
剪 交 浇 搅 揭 接 解 解放 解释 借 纠正 救 举 卷 掘 砍 扛 考  
烤 磕 克服 啃 恐吓 抠 扣 夸 扩充 拉 拦 捞 炼 晾 淋 搂 埋 买  
卖 迷 描 描写 摸 磨 抹<sub>1</sub> 抹<sub>2</sub> 拿 碾 撵 捻 捏 拧 扭 弄 挪 爬 拍  
排 派 派遣 榜 抛 泡 培养 陪 喷 捧 碰 批 披 劈 拼 泼 扑 铺 砌  
掐 牵 迁移 敲 切 取 去 劝 劝解 染 绕 惹 认 扔 揉 塞 扫 筛 晒  
删 闪 烧 拾 实现 试 试验 收集 收拾 刷 摔 拴 撕 送 算 锁 抬 贪  
污 烫 掏 逃 淘 淘汰 套 腾 踢 提 提拔 替 替换 剃 舔 挑<sub>1</sub> 挑<sub>2</sub> 贴  
听见 捅 涂 推 推翻 推广 吞 拖 托 脱 驮 挖 挽救 喂 捂 吸 吸收  
洗 吓 吓唬 掀 晓得 泄露 修 选 压 淹 腌 延长 掩护 掩饰 要 噎 应  
用 用 运 运用 砸 宰 凿 扎<sub>1</sub> 扎<sub>2</sub> 轧 炸 摘 沾 找 照料 召集 折磨  
整顿 整理 指导 指点 治 重复 煮 转<sub>1</sub> 转<sub>2</sub> 赚 撞 追 钻 做 作为 坐