

## 第二章 现代汉语短语句法语义属性范畴的确立

### § 2.1 短语句法知识表达的理论框架

#### 2.1.1 词组本位语法体系

本文把词组本位语法体系作为表达短语句法知识的基本框架。

众所周知，汉语语法研究发展至今，以某某“本位”冠名，或者虽不直接以某某本位冠名，但实质上也是大有“本位”意味的语法体系，已经有“词本位、句本位、词组本位、小句中枢、字本位”等等。对这些不同的本位体系作全面而具体的比较考察，不在本文的研究范围之内。但出于概括词组本位语法体系的精神实质的需要，这些语法体系成为事实上不容忽视的参照背景。将这些本位并肩而立，似乎很容易让人觉得：（1）一个语法体系中设置的大小不同的语法单位，重要程度不一样，应该区别对待；（2）要有效地揭示语言现象的规律，最重要的就是要找到地位最特殊的语法单位。

然而事实果真如此吗？

实际上，正如朱德熙先生在《语法答问》中已经精辟阐述过的那样，语法体系不同于语法理论，前者“在很大程度上是指的语法事实和语法规律的表述系统”，后者才是“揭示语法构造的本质和规律的”<sup>1</sup>。作为语法事实和语法规律的表述系统，不同的语法体系可以选择不同的语法单位作为表述的基础，这样会导致语法体系整体面貌的不同，以及简洁程度、解释效力上的一些差异。但是，却没有任何理由认为，语言系统中的大小单位，“本来”就有什么地位等级之分。归根结底，一个语法体系目标应该是很明确的，就是要来清晰地说明一个语言中的小单位是如何组成大单位的（反之亦然，即说明大单位是如何分解成小单位的），即必须回答这样一系列问题：篇章是如何组成的？句子是如何组成的？词组是如何组成的？……等等。在探索这些问题的答案的过程中，审慎的态度应该是，在任何时候都应该以最根本的理论目标为本，尽量去发掘有效的范畴来说明语言现象的规律，而不要把达到最终的真实目标的希望事先寄托在一个“特殊”的语法单位身上。

如果基于上述观点来审视“词组本位语法体系”，就可以把这一体系的要旨概括为两点：

（1）这一体系在句法语义层面上，相当严密而系统地阐述了**汉语词组的组织机制**；（2）这一体系认为在相当程度上汉语词组的组织机制基本就反映了**汉语句子的组织机制**。

所谓汉语词组的组织机制，实际上包含了两方面内容，一是汉语词组的构成模式，如主谓词组、述宾词组、偏正词组、……等等；二是词组的构成条件，即什么语言成分跟什么语言成分能组合成主谓结构，什么语言成分能跟什么语言成分能组成述宾结构，……等等。就这两方面内容来说，“词组本位语法体系”对汉语词组的语言事实所做的描写和刻画，代表了一个时代的水平。不过有些可惜的是，汉语学界过去对词组本位语法体系的评价，赞成也好，批评也好，似乎对这上面的（1）关注得都并不够，而是过多地强调了（2）。

我们认为，对于（1），应该继续发挥它的核心思想，使得基于这一体系的表述可以更纯粹，并能进一步揭示更多的汉语语法规律。而对于（2），客观地说，是词组本位语法体系的提出者在当时的研究背景下做出的过于乐观的估计。尽管这一主张有积极的理论意义，但理性地看，也许强调汉语句子的组织机制跟词组的组织机制在基本相同的同时也还存在不少有待探索的差异，更为客观一些<sup>2</sup>，对研究工作也会产生更为积极的推动。由于本文针对的对象是汉语的短语结构，重点是发掘汉语词组的组织机制，因此对句子的组织机制仅作为背景来考虑，而不是作为直接面对的前景。这里不进一步对（2）展开评述。下面仅就（1）涉及

的一些基本原则问题加以扼要概括。

词组本位语法体系在阐述汉语词组的组织机制方面，实际上有这样两个基本考虑：(1)以抽象的句法结构为研究对象并以结构类型为整个语法体系的基础概念；(2)始终强调以句法功能特征为建立句法范畴的最主要标准。

具体说来，《语法讲义》18章中有6章篇幅是在专门论述汉语词组构造的规律。而对汉语词组的描写，正是以“主谓结构”、“述宾结构”、“述补结构”、“偏正结构”、“联合结构”、“连谓结构”等不同类型的结构为纲展开的。至于“主语”、“谓语”、“述语”、“宾语”、“补语”、定语“、状语”、“名词”、“动词”、“形容词”等等一系列概念，都是在“主谓结构”等结构类型概念基础上导出的。作为基础概念，“主谓结构”、“述宾结构”等等则没有给出明确的定义。这些概念的基础是语言学家的语感，即感觉一些词组跟另一些词组相同与不同的“经验”<sup>3</sup>。从公理化方法的角度看，以这些结构类型作为基础概念，可以甚至不需要加以定义。它们的合理性取决于系统整体的描写能力和解释能力。如果在这些概念的基础上，运用正确的逻辑，能够推演出一整套比较好的可以揭示汉语词组的构造规律的概念和命题体系，就说明这些基础概念是合适的。反之，如果推演出的一整套体系解释语言事实不力，或者导出了跟语言事实不符的结论，就要怀疑这些基础概念是否合适或者可靠。

再看建立句法范畴。这方面最鲜明地表现在汉语的词类划分上。《语法讲义》中有7章篇幅是讨论汉语词类的。它正是以上述结构类型概念为基础，用功能特征为标准，建立起了汉语的词类体系。这样分出来的类，目的也非常明确，就是要能反过来根据一个词的词性判定它可能出现的句法结构位置。比如“通红”定为“状态词”，就可以判定它能出现在哪些结构位置，以及不能出现在哪些结构位置。根据分类体系对“状态词”的功能描述<sup>4</sup>，可以知道，“通红”不能出现在“很、不”等副词后面跟这些副词构成状中结构（如不说“\*很通红”、“\*不通红”）；也不能出现在述补结构的述语位置（如不说“\*通红极了”），但可以出现在述补结构中的补语位置（如可以说“脸晒得通红”），等等。

在我们看来，以上述两方面的思想原则为基础，可以在词组本位语法体系的框架下，做许多进一步拓展的工作。比如在词类信息的基础上对词的功能特征进行更细的描写。这里不妨来看一个例子。“是”跟“像”都是所谓的系动词，从功能上讲，有很明显的共性，都不能带“着、了、过”等动态助词。但同时这两个动词又有功能差异，“是”前面不能加“很”（“\*木兰很是女郎”），“像”前面可以加“很”（“木兰很像女郎”）；“是”后面不能带“极了”构成述补结构（“\*木兰是极了”），“像”则可以（“木兰像极了”）。这样，如果要准确地描写“是”、“像”在参与组合构成词组时的规则，就有必要把上述差别记录下来。从分类的角度讲，就是类可以不断的细分，小类下面还可以再分小类。句法分析要求对一个词的功能特征的信息掌握得越多越全面越好，但分类并不是达到这一目的的唯一手段，属性特征描述就是另一种可行的选择。比如就“是”和“像”这两个动词的情况来说，既可以选择在“系动词”的下面再细分不同的小类，也可以用“属性：值”这样的表达形式，直接描述这两个动词前面受“很”修饰、后面带补语的能力。《现代汉语语法信息词典》采用的就正是后一种做法。

另外还有一个可能的拓展方向，就是对更大的语言单位（比如词组），也像对词那样，对其功能特征进行细致的描写。就本文的研究目的来说，这个方向上的拓展将是我们重点探索的内容。下面首先从短语的功能分类开始。

### 2.1.2 短语的句法功能分类框架

跟词的句法功能分类一样，我们完全可以按照类似的思路对汉语的短语结构进行功能分类。同样地，建立短语句法功能类范畴的基础仍然是事先假定的结构类型概念。根据需要，

可以设定“主谓结构、述宾结构、述补结构、定中结构、状中结构、连谓结构、联合结构、附加结构、的字结构”等9个基本结构类型<sup>5</sup>。值得强调的是，结构类型的数量不是封闭的而是开放的。这体现在两个方向上，向外表现为基本结构类型可以再扩大。比如考虑“他所写的书”中的“他所写”，就可以增加一个“所写结构”这样的类型。考虑“高兴高兴”，可以增加“重叠结构”这样的类型。向内体现在每种结构类型还可以再细分类。比如“述补结构”就可以按照一些形式标准分成“组合式述补结构”和“粘合式述补结构”<sup>6</sup>。我们可以根据系统的实际需要来决定是否在这两个方向上进行一定的调整。

下表给出了汉语短语的基本结构类型和相应的实例。

序号	基础概念	一级导出概念		实 例
		主语	谓语	
1	主谓结构	主语	谓语	树叶黄了;小明喜欢看电视;感冒传染
2	述宾结构	述语 1	宾语	喝了三杯酒;学了三年;企图逃跑;送他香烟
3	述补结构	述语 2	补语	洗干净;做得非常好;好得很;吃得完;拿出来
4	定中结构	定语	中心语 1	一斤白菜;老师的眼泪;大红灯笼;削梨的刀
5	状中结构	状语	中心语 2	快跑;认真地学习;把饭吃完;明天见;屋里坐
6	连谓结构	前项	后项	开着窗户睡觉;打电话请医生;派助手办理;请他来
7	联合结构	前项	后项	小说和戏剧;又高兴又难过;批评教育
8	附加结构	中心语 3	附加语	红着;吃了;砍光了;努力奋斗过
9	的字结构	中心语 4	附加语	买菜的;老师表扬了的;冰凉的;慢性的

(表 2-1: 汉语词组基本结构类型一览表)

关于上表，有两点需要说明。

(一)在不严格的情况下，“主语、谓语、述语1、中心语1……”等等概念，都可以有两个不同的所指。一个所指是语言成分本身，这时“主语”是指处在主谓结构中前项位置上的那个成分（比如在“树叶黄了”中，“树叶”是“主语”）。还有一个所指是结构位置，这时“主语”可以理解为主谓结构中前项所占据的那个位置。比如我们问“这个动词能作主语吗？”，实际上这相当于问“这个动词能出现在主谓结构的前项所在的位置上吗？”<sup>7</sup>。实际上，是不是叫“主语”这个名称都无所谓，只是照顾到传统，为了称说方便，本文把这个结构位置以及处在该位置上的语言成分（下文以汉字和字母代码区分这两个不同的所指）称为“主语”。其他的概念，“述语 1、中心语 1、中心语 2、……”等等，都是如此。

(二)属于同一种结构类型的短语，从功能上讲，可以有不同情况，比如“小说和戏剧”、“又高兴又难过”都是联合结构，但从功能上讲，前者是名词性的，后者是形容词性的。再比如“买菜的”、“冰凉的”，从组合格式上讲，都是“的”字结构，但从功能上讲，前者是名词性的，后者是形容词性的。本文对短语的描述，将主要从功能着眼。

以上述结构类型概念为基础，下表初步确定了一个现代汉语短语功能分类体系<sup>8</sup>。

序号	标记	功能类名称	典型功能													
			a	b	c	d	e	f	g	h	i	j	k	l	m	n
1	dj	单句型短语		+		+										+
2	np	名词性短语	+			+			+	+				+		+
3	vp	动词性短语		+	+		+					+	+	+	+	+
4	ap	形容词性短语		+			+	+	+		+	+		+	+	+
5	dp	副词性短语									+					
6	pp	介词性短语									+		+			
7	sp	处所词性短语				+										+



细的功能标准来区分, 比如用【能否跟量词组成定中结构】这个标准可以帮助我们把 mp 跟 mcp 区分开, 而【能否用‘这时、那时’或者‘这里、那里’替换】这个标准, 则可以帮助我们把 tp 跟 sp 区分开。很显然, 用到的功能划分标准越多越细, 就可以把短语类型按照功能的不同分出越多的小类。从这个角度讲, 也说明对语言单位进行功能分类是相对的。具体的分类框架可以调整。评判分类框架的好坏则取决于分类的目的, 具体的说就是看应用这个分类框架去达成某个特定的目的, 效果如何。

(5) 短语句法功能分类的目的是为了刻画一个具体的短语的句法功能性质。举例来说, 如果给“看电影”这个短语贴上“vp”的功能标签, 就意味着能得到这样的推论, 即一般情况下, “看电影”可以出现在状中结构的中心语位置(“正在看电影”), 可以出现在连谓结构的前项或后项位置(“去看电影”), 不能出现在述补结构的补语位置, 等等。

(6) 正如不应奢求词的功能分类标记可以描述一个词的全部句法功能特征一样, 短语的句法功能标记, 也只是大致反映了一个短语主要的句法功能特征。拿词来说, 虽然名词的典型功能是在由数量词充任定语的定中结构中充任中心语, 但这并不表示所有的名词都要无一例外的必须具备这项功能, 比如“笔者”, 就不受数量词的修饰, 但“笔者”仍然是名词。理由很简单, 就是“笔者”不是别的什么词。它能作主语, 不能作谓语, 不能作补语, 不能作状语, 不能作连谓结构前后项成分, 等等。同样地, 短语的功能类标记也是如此, 虽然“看电影”不能出现在述补或者述宾结构中的述语位置上, 但这并不妨碍给“看电影”贴上一个 vp 的功能标签。在很大程度上可以认为是基于这样的理由, 即“看电影”不是“别的”什么短语。它不能作定语, 不能作补语, 不能受数量短语修饰, 等等。

(7) 既然功能类标记并不能反映一个短语的全部功能特征, 那么为了达到描述一个具体短语的全部功能性质的目的, 就应该跟对词语进行句法功能特征描述的方式一样, 在给出一个基本的大类体系基础上, 以“属性: 值”的方式来刻画一个具体的短语丰富的句法功能特性。事实上, 一个语言单位以句法功能为标准的任何分类体系都可以等价地转换成属性描述方式, 反过来任何属性描述的结果也都可以完全反映在一个分类体系中。从操作上讲, 以分类加属性描述的方法来表示语言成分的句法功能特征是合适的。

下面我们则将把这种做法引入到短语语义属性的表示上。

## § 2.2 短语语义知识表达的理论框架

### 2.2.1 广义配价模式

关于自然语言的语义分析理论, 大致有语义场理论 (Semantic Field) 和义素分析理论 (Componential Analysis)<sup>13</sup>、配价理论 (Valence Grammar)<sup>14</sup>、格语法 (Case Grammar)<sup>15</sup>、论旨理论 (Theta-Theory)<sup>16</sup>、概念依存理论 (Concept Dependence)<sup>17</sup>、语义网络 (Semantic Network)<sup>18</sup>、蒙太格语法 (Montague Grammar)<sup>19</sup>等。而从大规模描述自然语言语义知识的工程实践来看, 以英语和其他主要欧洲语言为描述对象的研究, 影响较大的有 WordNet、MindNet、FrameNet 等。中文信息处理方面则有“信息处理用现代汉语语义分类体系”<sup>20</sup>、“现代汉语述语动词机器词典”<sup>21</sup>、“董氏语义知识词典”<sup>22</sup>等比较有代表性的研究工作。

不同的语义理论在理论背景和侧重点上各有不同, 比如配价理论出自依存语法, 重视语义的句法对应性; 格语法在转换生成语法之后提出, 重视纯粹的动名语义关系的发掘。论旨理论作为生成语法理论的一个子系统, 强调把动词的句法搭配限制和语义搭配限制结合到一起统一处理。概念依存理论侧重表达概念之间的依存关系, 蒙太格语法则侧重通过演算得到句子的逻辑语义表达式, 等等。不同的语义工程在描述自然语言语义知识的深度和广度上以及语义知识的获取方式上也存在诸多差异。对于上述众多的语义理论以及语义工程实践, 本

文不准备展开评述，而只是在这些研究工作提供的广阔背景空间下，同时考虑汉语语义研究的实际情况，围绕下面这两个问题，来探讨如何组织汉语短语的语义知识。

(1) 语义知识是干什么用的？

(2) 在目的大致明确的情况下，如何组织语义知识比较适宜？

我们认为，自然语言的语义知识在信息处理中的作用可以概括为两个主要的方面<sup>23</sup>。一是帮助句法分析得到正确的结果；二是为刻画句子中各个语言成分所对应的概念之间的意义关系提供支持，这类似于回答新闻中的五个 W 和一个 How 的问题。这里可以举一个简单的例子来说明。比如对汉语中的两个短语，a “修理自行车的后胎” 和 b “修理自行车的师傅”。这两个短语都对应对应着线性序列模式 M：“V N 的 N”，这个序列在句法结构上至少有两种组合的可能性，即 A. [V [[N 的] N]] 和 B. [[ [V N] 的] N]。要判断一个具体的符合 M 的短语（比如上面例子 a 跟 b），应该按 A 切分还是应该按 B 切分，光靠组成这个短语的词的句法属性信息是不够的，而如果有关于“修理”、“自行车”、“后胎”、“师傅”等词语的语义知识，就可能便于作出正确的判断了。比如对 a、b 两例，可以包括这样一些语义知识，“后胎”是“修理”的【对象】，同时又是“自行车”的【部分】<sup>24</sup>，而不能是“修理”这个动作的【发出者】，“师傅”才是动作的【发出者】，等等。有了这些语义知识，就可以判断出 a 应该按 A 式切分，b 应该按 B 切分。而且在得到句法分析的正确结果的同时，也能得到 a 和 b 的语义解释，即两个短语中各成分之间的意义关系。a 中“修理”跟“自行车的后胎”之间是“动作——对象”的关系，“自行车”跟“后胎”之间是“整体——部分”的关系；而 b 中“修理”跟“师傅”之间是“动作——动作发出者”的关系，等等。

针对上述两个目的，上面提到的第二个问题就又可以分解为这样两方面来考虑。一是语义知识跟句法知识的配合；二是语义知识的深度和广度。

就第一个方面来说，我们认为，句法知识主要是在语言成分的线性序列基础上着力刻画一个语言成分参与组合时所能够占据的结构位置的性质。比如“‘修理’是动词”这个句法知识，就是在刻画它能跟名词（比如“自行车”）组合形成 V+N（如“修理自行车”）这样的述宾结构。而语义知识则关注跟线性位置无关的，语言成分所对应的概念之间的意义联系。比如在“老王修理的那辆自行车”、“这辆自行车破旧不堪修理起来太费劲了”中，虽然“修理”跟“自行车”之间没有直接的结构关系，但它们之间始终有“动作——对象”的意义联系。如果从搭配限制的严格程度来看，就是语义知识所描述的搭配限制比句法知识描述的搭配限制要严格一些。比如“修理”是动词，意味着它能跟名词形成 V+N 述宾结构，不跟形容词形成述宾结构，等等。而语义知识则用来说明“修理”能跟“自行车”这样的名词形成述宾结构，不能跟“气泡”这样的名词形成述宾结构，等等。显然，语义知识可以看作是在句法知识基础上，再以意义为标准对语言成分所做的细分类（当然分类的结果也得有形式上的验证）。而以汉语语法研究的实际情况来看，李临定（1990）<sup>25</sup>对汉语动词的研究，《动词用法词典》<sup>26</sup>对汉语常用动词的搭配描写，沈阳（1994）关于动词句位的研究，以及袁毓林（1998）关于汉语动词的配价研究，都是在句法知识基础上深挖动词的语义知识很有价值的研究工作。

就第二个方面来说，语义知识的描写越深越广，对句法分析的支持当然也就越大，但就操作上的可行性来讲，语义知识的获取又是非常困难的工作。语义知识实际上往往难以刻画到很深很广的程度<sup>27</sup>。已有的语义知识工程实践在这方面都进行了可贵的探索。这里我们不妨作一些简单的理论分析。配价理论主要是以动词为中心，记录一个动词在一个最小结构框架内能跟几个名词发生搭配关系，从而区别出 1 价、2 价、3 价动词等等。比如汉语中“咳嗽”只能跟一个名词发生搭配关系（“张三咳嗽”），它就是 1 价动词。“看”可以同时跟两个名词发生关系（如“张三看报纸”），它就是 2 价动词。“送”可以同时跟三个名词发生关系（如“张三送李四一本书”），它就是 3 价动词。格语法说是以名词为中心，实质上也是以动

词为中心,记录一个动词可以跟多少种意义类型的名词发生关系。比如“写”可以跟表示“工具”的名词发生搭配关系(“写毛笔”),也可以跟表示“处所”的名词发生关系(“写黑板”),还可以跟表示“对象/结果”的名词发生关系(“写那个情节”、“写信”),等等。论旨理论把上述不同语义类型的名词性成分称为该动词的不同论旨角色(或论元)。我们称之为“角色类型”。这些语义知识对汉语的句法分析都很有用,因此对一个动词的配价性质和角色类型,都有必要在词典中加以详细记录。跟我们对确立范畴的一贯态度一致,这些语义范畴的数量也是开放的,可以根据实际需要增补或者删并。还有,描述词语之间的语义搭配关系,也可以在精确度和概括度的不同水平上进行。如果强调精确度,就最好是词对词的搭配。如果要强调概括度,就描写词对类或者类对类的搭配。比如动词“突出”,能搭配的名词限于“包围圈”、“重围”等少数几个。要描写“突出”的配价性质,就应该强调精确度,直接记录能跟“突出”搭配的名词本身。再比如动词“吃”,能搭配的名词很广泛,要描写“吃”的配价性质,就可以概括描写为跟[可食物]类名词搭配,而不是像“突出”那样列举能够搭配的具体名词。在实际操作中,有的时候可以进行概括描写,有的时候可以精确描写。在我们看来,语义知识描写的深度和广度是跟应用目标直接相关的。针对不同的目标,在实施策略上完全允许有不同的选择。值得一提的是,要进行概括描写,前提是要有一个关于实词的语义分类体系。在2.2.2中我们将具体介绍一个比较简明的汉语实词语义分类体系。

除此之外,对动词语义性质的描述还应该在更深的层面上展开。那就是关于动词的搭配成分的变化情况。一般而言,动词的配价成分在动作行为起始到终止的过程中会伴随着变化。这些变化情况应该记录在动词的语义性质描述框架中。例如:性状变化、数量变化、领属关系变化、位置变化等<sup>28</sup>。与此同时,正如有相当多学者已经注意到的那样,描述语言成分间的搭配关系,并不限于动词跟名词之间的搭配。形容词跟名词之间,名词跟名词之间,动词跟形容词之间,以及副词跟动词之间,都存在语义搭配的问题,广义地讲,任何实义语言成分的搭配关系,都可以在语义层面上进行描写,而且不光是在词这一级单位上,完全可以并且有必要扩展到短语这一级语言单位上。也正是基于上述认识,我们把本文提出的语义知识表达框架称为“广义配价模式”(Augmented Valence Mode)。

在广义配价模式下,汉语实词语义知识是分为以下三个层级的内容加以描述的。

#### (1) 词语本身所属的语义类<sup>29</sup>。

在词典中,对每一个具体的实词,需要描述其语义类属性。比如“太阳”属于[天体]类、“红”属于[颜色]类、“拿”属于[搬移]类等。跟对词进行句法功能分类的目的类似,对词语进行语义分类,目的之一也是希望分类结果能用于控制句法分析。换言之,我们希望词语语义分类的结果达到这样的功效,即不同语义类别的词语在句法上应该有显著的差异表现,而同属一个语义类的词语在句法上应该有显著的相似表现。当然,也跟词的句法功能分类结果一样,要达到这样的目标光靠分类是很困难的,因此就要在更多的维度上对一个词的语义性质进行区分。于是下面引入第二层级的语义知识。

#### (2) 词语配价成分的个数,角色类型,及其选择限制。

从理论上讲,词语的配价成分这个概念对动词、形容词、名词都适用,但从重要性上讲,动词更为突出,动词的配价成分的情况也最为复杂。这里我们仅以动词为例来说明如何组织这一层级的语义知识。有关论述对名词和形容词同样适用,只不过就汉语的具体情况来说,相当多的名词和形容词在这一层级上的语义知识比较贫乏。

对动词来讲,配价数(即配价成分的个数)的取值范围<sup>30</sup>为:0、1、2、3。角色类型分成两部分<sup>31</sup>:一部分是核心语义角色,包括主体、客体、与事、工具、处所等;一部分是外围语义角色,包括空间和时间等<sup>32</sup>。选择限制则要求指明一个动词对其配价成分有哪些条件限制。实际上限制可以在句法层面描述,也可以在语义层面描述;可以是精确限制,也可以是概括限制;可以从正面来要求配价成分应符合某种条件,也可以从反面来要求配价成分应

避免哪些情况。表达形式可以根据具体动词的不同情况灵活选择<sup>33</sup>。

需要说明的是,本文认为应该区分动词对应的动作过程所在的“空间”和动作相关的物体所在的“处所”两个概念<sup>34</sup>。在我们的表述框架中,前者是外围的角色类型“空间”,在句法形式上是以“在+宾语”介词结构做状语来标示的(如“在教室里唱歌”);后者是核心的角色类型“处所”。在句法形式上除了可能以介词结构做状语来标示外(如“从屋里搬了出来”),还可以处所宾语来标示(如“(字)写在黑板上”<sup>35</sup>、“(书)放桌上”),特别是以宾语来标示的情况,一定是标示“处所”,而非“空间”;一般来说,动作总是在一定的“空间”中发生,因此动词“空间”属性的缺省值为“+”。实际表达中“空间”跟“处所”也可能重合(即有歧义)。比如“在院子里堆白菜”(比较“白菜堆在院子里”),“堆”的动作过程所在“空间”是“院子里”,所涉及的物体“白菜”所在的“处所”也是“院子里”。此外,一个动词能否带“处所”宾语,是由其词汇内容(lexical content)决定的。像通常说的“吃食堂”,“食堂”是“吃”的处所宾语,只是一种简单的处理方式。跟“睡床上”这种真正的“处所”比起来,“吃食堂”中的“食堂”的处所意味并不那么强就可以明显地暴露出来。与其说“食堂”是一种处所,还不如说它是一种用餐的“方式”更符合实际。因为,大学生几乎都“吃食堂”,但并不见得“饭都在食堂里”。事实上,有不少学生是“吃食堂的”,同时“他(她)们是把饭买回去,在宿舍里吃的”。而“睡床上”,决定了动作“睡”的主体所处的位置一定在“床上”。这种差别还体现在形式上,就是“吃”并不能带像“床上”、“大厅里”等含方位词的处所短语(\*吃大厅里)。进一步深究“吃”和“睡”的词义差别,不难发现,“吃”的词义内容中,它的“主体”或者“客体”所在的空间位置并不是突显的(salient),而“睡”的词义内容中,它的“主体”状态最需要强调,同时主体所在空间位置也是突显的。像“睡”这样的主体位置突显动词还有如“来”、“去”、“回”、“坐”等,此外还有像“搬”、“写”等客体位置突显的动词,这些动词都关注“主体”或“客体”的空间位置(即“处所”)。它们是能带真正的处所宾语的那一类动词。而像“吃”这样的动词,本文认为它不带处所宾语。对这种差别,可以通过词语配价成分的变化来描述。这就是下面“广义配价模式”第三层级的语义知识。

### (3) 词语配价成分的变化情况。

正如上文已经提到的那样,有些情况下,我们需要描述词语的搭配成分的变化情况。在这个层级上描述词语的语义知识,主要针对的是动词。目的是在比上两个层级更深的层面上,进一步对一些动词之间更细的语义差别进行区分。

下面表 2-4 中“洗”和“晾”就包含了配价成分变化的语义信息。

词	词性	广义配价模式					
		语义类	配价数	论元角色选择限制		配价成分变化	
				主体	客体	主体变化	客体变化
大衣	n	服饰	0				
父亲	n	人	1	[语义类:人]			
热情	a	态度	2	[语义类:人]	[语义类:人   事]		
洗	v	促变	2	[语义类:人]	[语义类:具体物 “天体...”] <sup>36</sup>	[性状]	[性状] <sup>37</sup>
晾	v	促变	2	[语义类:人]	[语义类:具体物 “天体...”]	[性状]	[性状   位置] <sup>38</sup>

(表 2-4: 以广义配价模式描述汉语实词语义信息举例)

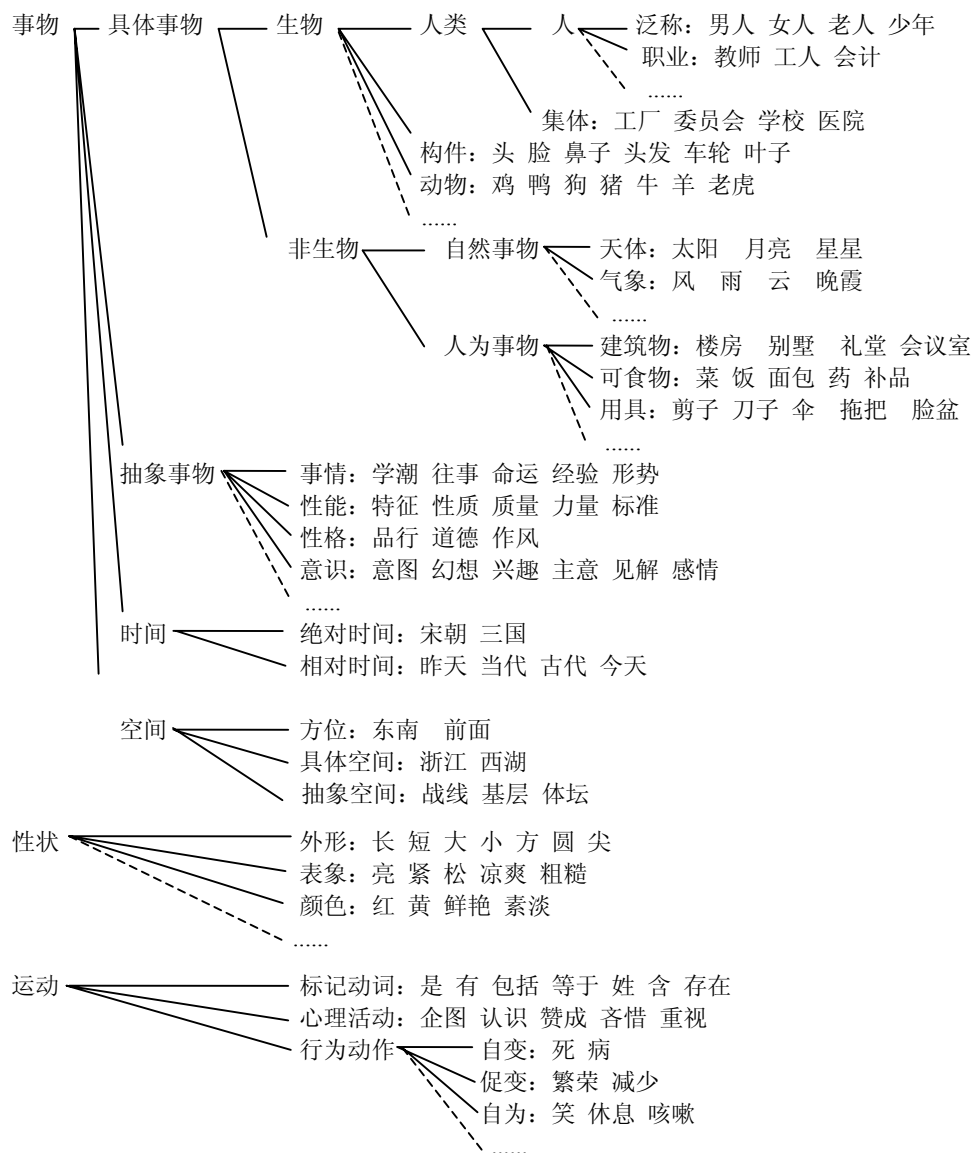
上述广义配价模式的第(1)、(2)两个层面,以往的语义研究已进行了一定的探讨。而第(3)个层面,即标注动词配价成分的变化情况,则是以往的语义研究较少涉及的。虽然

像在概念依存语义理论中也提到了用诸如[ATRANS]这样的原始概念来描述英语动词“give、take”等表达领属关系转变的情况<sup>39</sup>，但针对汉语的语义研究，还未见到描述动词配价成分变化情况的，尤其是汉语动词配价成分的变化在句法上“恰恰”对应着汉语别具特色的述补结构的使用。补充这个层面的语义信息，可能为汉语述补结构特别是补语语义指向的分析提供了一条比较有效的线索。

广义配价模式最基本的思想是分层次描写实词的语义性质<sup>40</sup>。这跟以词类信息加属性特征描述来刻画词语的语法性质的思想是非常一致的。上面构造的这种三级语义属性描述框架就很鲜明地体现了这一点。比如例中“洗”跟“晾”，虽属同一个语义类，但它们的差别可以在第三层级的配价成分变化性质描述中得到反映。请比较“洗干净了”——“晾干了”，（“洗”跟“晾”都可以接表示性状变化的补语）；“\*洗在盆子里”——“晾在院子里”（“洗”不能带表示处所位置的宾语，“晾”则可以带表示客体处所位置的宾语）。

### 2.2.2 汉语实词的语义分类体系及语义关系范畴

首先给出汉语实词语义分类体系的基本框架。“：”后面的是例词；虚线和省略号表示大类中还有其他没列出的小类（见下页图 2-1）。



（图 2-1：现代汉语实词语义分类体系）

在分类体系之外,一个语义知识表达模型中还应该表示两个语言成分之间语义关系的范畴。目前我们主要考虑各类实词跟名词之间的语义关系,初步设定了以下范畴。

词类	名词	形容词	动词	
名词	主体	主体 客体	核心语义角色	外围语义角色
			主体 客体 与事 工具 处所	空间 时间

(表 2-5: 汉语实词语义关系范畴)

对本文提出的语义分类体系以及语义关系范畴,还需要作如下补充说明。

首先,对词语的语义性质进行准确的描述,是否一定用分类的方法,以及就算用了分类的方法,具体又该怎么分,都还需要在实践中继续探索。具体分类的标准也可以从别的角度加以考虑,并不是唯一的<sup>41</sup>。

其次,关于实词间语义关系范畴,本文的研究工作在很大程度上只是提出了一个可供今后进一步调整的框架。动词跟名词之间语义关系,虽已有较多研究,但目前也仍然处在探索阶段。本文暂时确定的“主体、客体、与事、工具、处所”等核心语义角色以及“空间、时间”等外围语义角色,也都可以根据实际情况的需要增删并合。再拿名词跟名词之间的语义关系来说,也存在相当多的复杂情况,比如“木头房子”中,“木头”跟“房子”之间的关系是“材料——实体”的关系;“故事情节”中,“故事”跟“情节”是“整体——部分”的关系,等等。对汉语名词间这些复杂的语义关系情况,目前尚未见到系统的研究工作,本文也只是描述了亲属类名词(如“父亲、丈夫、兄长”等等)以及构件类名词(如“头、脖子、腿”等等)的配价性质。这些名词需要一个实体名词在现实语境中来帮助确定所指。我们把实体名词称作这类名词的“主体”配价成分。此外,有关动词跟形容词之间,动词跟动词之间的语义关系,也都还没有展开大规模系统地理论研究和实践工作。考虑到这些实际情况,我们对汉语实词间语义关系性质的描述主要是抱尝试态度,希望在实践中积累更多经验。值得一提的是,上面提出的汉语词语语义知识描写模式已经在本文第一章提到的汉英机器翻译系统中付诸实施,即描写了四万多汉语实词的语义类信息和配价信息。目前尚未大规模地描写动词的配价成分变化信息。可参见詹卫东(1998)<sup>42</sup>的介绍。

以上概要阐述了本文组织汉语句法知识和语义知识的理论框架。本小节对汉语语义知识表达模式的阐发是以词语这一级语言单位为对象展开的。在下面这一节中,我们将说明如何把前面谈到的句法语义知识的表达框架落实到对短语的句法语义属性的描述上。

### § 2.3 从词到短语: 属性变异描述

要描述短语的句法语义属性,理想的情况是,任意一个具体短语的属性特征(句法的和语义的)能够从其组成成分的属性特征按一定的规则(程序)推导出来。因为短语是开放集合,不可能像对词那样逐一进行描述,而只有通过由小的组成成分的属性信息推导得到更大单位的句法语义属性信息的办法来加以刻画。需要说明的是,从具体词的句法语义性质到具体短语的句法语义性质,这中间可以看作是一个“推导”的过程,即根据有关规则的描述来判定一个具体的词在向短语组合的过程中,句法语义功能发生了哪些变化。但对一条从词到短语的组合规则而言,则不是数学意义上严格的“推导”过程(因为我们并没有用来推导的,关于一条具体的组合规则的更基础的“元规则”),而是本小节标题所谓的“属性变异描述”,即我们在规则中“描述”了词向短语组合的过程中会发生哪些变异,这样,在分析具体的词语组合时,就可以利用已有的规则描述来进行动态的“推导”。

先来看句法属性的推导。实际上,布龙菲尔德(L. Bloomfield)曾经根据结构体中是否

存在与整体语法功能相同的组成成分，把结构组合体区分为向心（endocentric construction）和离心（exocentric construction）两类<sup>43</sup>。如果用推导的眼光来看布氏所做的这种区分，就可以这样来表述，有些短语在由词组合得到后，主要句法功能可以简单地由其中心成分的句法功能导出。比如“poor John”的功能可以由其中中心成分“John”的功能导出。这样的短语（结构体）就是向心结构；而有些短语在由词组合得到后，主要功能不能直接由其中任何一个组成成分的功能导出。比如“John ran”，其功能即不同于“John”，也不同于“ran”。这样的短语（结构体）就是离心结构。

就汉语的实际情况来看，这样的处理显得简单了些。相当多的情况是，一个短语的句法功能，部分与其组成成分的功能一致，部分又不一致。只是不同的短语，一致和不一致的比重会有不同。以“很喜欢”这个短语为例，整个短语的句法功能基本跟其中心成分“喜欢”一致，但也有不一致的地方，就是“很喜欢”不能跟“非常”构成状中结构，而“喜欢”可以。显然这个不一致是由于“很”造成的。如果只是说明“很喜欢”是向心结构，就不能全面反映它的功能特征。再比如像“公路建设”这样的短语构造，它的内部组成为“名词+动词”格式<sup>44</sup>。整体功能表现为不能做谓语，同时也不能前加修饰语构成状中式 vp。相反，它可以直接做定语（如“公路建设队伍”）；可以做宾语（如“进行公路建设”、“狠抓公路建设”）。显然，“公路建设”不像一般的“名词+动词”格式那样是构成主谓式 dj。它的功能特征更偏向 np。不过，值得一提的是，“公路建设”不能受数量短语的修饰。如不能说“\*一条公路建设”。很显然，“公路建设”的整体功能既不是从其中的名词“公路”，也不是从其中的动词“建设”那里简单继承来的。无论说它是向心结构，还是离心结构，都难以准确全面地反映它的实际功能特征情况。

从上面两个简单的例子不难看出，对一个短语的功能性质的描述，要想尽量符合实际情况，就应该比向心、离心这样简单的二分更进一步，参照描述词的功能性质的方式，尽可能全面地描述由词到短语的过程中，到底有哪些功能性质发生了变异。这包括**功能消失**与**功能增生**两个方面。

比如“有”、“看”等动词，单个前面不能加“很”。如不说“\*很有”、“\*很看”。但如果“有、看”等带上宾语构成“有才华”、“看了一些书”等述宾式 vp，功能就发生增值，前面可以受“很”修饰了，如汉语可以说“很有才华”、“很看了一些书”等<sup>45</sup>。因此描写“有才华”这个短语的功能特征，就要特别说明它的功能跟单个动词“有”相比，已经增生出能受程度副词“很”的修饰这一项特征。功能消失的例子也相当普遍，上面举的例子如“很喜欢”不能再受“非常”修饰，就可以视作“喜欢”可以受“非常”修饰这一项功能的丧失，还有如“公路建设”不能再带宾语，当然也应该视作“建设”这一动词参与形成 np 后主要功能的丧失。

上述有关短语句法属性推导的讨论，同样也适用于短语语义属性的推导。而且，正如词语的语义属性信息主要是用来描述词语间的搭配关系，短语的语义属性信息，主要目的之一同样也是用来描述一个短语的搭配能力。因此，从词推导短语的语义属性信息，着眼点仍然是，短语的组合功能跟词相比，发生了什么变化。只不过，我们把一些组合功能变化归结为短语的语义属性信息在起作用罢了。比如，动词“带”带上补语“来”之后组成 VP“带来”，其组合能力就发生了一些变化，“带”一般只能搭配表示具体物体的名词，如“带钢笔”、“带钱”等等。“带来”的搭配范围则扩展了，可以说“带来了一线希望”、“十月革命一声炮响，带来了马克思列宁主义”等等。显然，“带来”跟“带”相比，组合能力发生了变化，我们可以把这种变化记录在“带来”的语义属性信息描述中，即“带来”的客体配价成分的选择限制放宽。这类似于从词到短语，句法功能增值的情况。再比如“吃饱”，是动词“吃”带上补语“饱”形成的 VP，其组合能力跟“吃”相比也发生了变化，“吃”后面可以搭配不同语义角色类型的名词性成分，如“吃大碗”（工具）、“吃苹果”（客体）等，但“吃饱”能搭

配的语义角色类型就大大受到限制，只能说“吃饱饭”，“吃饱肚子”等，其他就很少有能跟“吃饱”搭配的名词性成分了，不能说“\*吃饱一个苹果”、“\*吃饱大碗”。我们把这种变化归结为，从动词“吃”向动词短语“吃饱”的组合过程中语义属性的变化使然。显然，“吃饱”的语义角色类型比“吃”受到更多的限制。这类类似于从词到短语，句法功能部分丧失的情况。

上文已经谈到，描述词语的语义搭配性质，可以在词对词的层面上精确描述，也可以在词对类的层面上概括描述。描述短语的句法语义性质也同样如此。既可以针对具体的短语组合（如“吃饱”）来描述，也可以针对短语的抽象组合模式（即类加类形成的短语类，如VP）来描述。无论是在具体短语的层面上进行描述，还是在抽象概括的短语类层面进行描述，目的都是一样的，就是要能准确反映一个具体短语的组合搭配性质。但相对来讲，关于短语的语义性质，在类的层面上进行推导，难度要大一些。比如王红旗（1995）和郭锐（1995）都曾讨论过汉语述补式短语的配价性质<sup>46</sup>。他们对述补式VP的语义性质的推导是希望在对具体述补式短语的配价分析基础上，能得到关于述补式VP的概括的语义推导公式。如果这样的公式被找到了，就可以像说明动词“吃”能跟哪些名词成分搭配那样，方便地说明“吃饱了”、“吃多了”等述补式VP能跟哪些名词性成分搭配了。但他们的研究工作表明，要在类的层面概括述补式VP的语义属性，尽管是可能的，但仍然是比较困难的，会有不少例外情况。而对于短语的句法属性推导，在类的层面概括相对就容易一些。比如上文举到的状中式VP的例子，如“很喜欢”，就可以很容易地在类的层面上概括为，**如果动词受程度副词修饰形成状中式VP，那么这类状中式VP就不能再受程度副词修饰了**。由此不难发现，选择在哪个层面上描述短语的句法语义属性，实际上反映了汉语短语研究的发展水平。短语的句法性质研究相对词而言还很不够，短语的语义性质的研究那就更少了，所以在类的层面上概括描述短语的语义性质就比描述其句法性质要困难得多。根据这样的实际情况，我们在概括短语的句法语义性质时，基本倾向就是，尽量在抽象类的层面上概括短语的句法性质，而主要在具体的词及短语的层面上考虑短语的语义性质的变异描述。理想的情况是，这样的具体描述积累到一定的程度后，有可能把针对具体短语组合的语义推导结果上升到对短语类的概括层面。

## § 2.4 小结

本章明确提出了对汉语的短语结构进行分析所需的句法和语义范畴。归纳如下：

### 一. 句法范畴

（一）观察短语的内部特征，确立的句法范畴：

- （1）内部结构类型，如“主谓结构、述宾结构、述补结构、连谓结构、……”等等；
- （2）组成成分名称，如“主语、谓语、述语、宾语、补语、附加语、……”等；
- （3）是否有疑问成分，如“好不好”、“多少人”可以表示疑问；
- （4）是否有否定成分，如“吃不了”、“没听见”表示否定；
- （5）是否有被动成分，如“被看见”、“让人拐跑了”表示被动；
- （6）是否有感叹祈使成分，如“多么美丽”表示感叹、“请进”表示祈使。

前两类范畴是用来描述一个短语的基本结构情况。比如“学习下棋”，内部结构是“述宾结构”，述语是“学习”，宾语是“下棋”。后四类范畴用于标明一个短语内部是否含有特定成分。这些范畴的作用将在下文具体论述短语结构规则时随文说明。

（二）审视各类短语的外部功能，确立的句法范畴：“=”后是属性的符号代码）

a: 作主语=zhuyu; b: 作谓语=weiyu; c: 作述语l=shuyul; d: 作宾语=binyu;

e: 作述语2=shuyu2; f: 作补语=buyu; g: 作定语=dingyu; h: 作中心语1=zhxyu1;  
i: 作状语=zhuangyu; j: 作中心语2=zhxyu2; k: 作连谓结构前项=lwqx; l: 作连谓结构后项=lwhx;  
m: 作联合结构前项=lhqx; n: 作联合结构后项=lhhx; o: 作中心语3=zhxyu3; p: 作中心语4=zhxyu4;

这里我们把“组成成分名称”跟“句法结构位置”两个概念区分开了(参见 2.1.2 中的有关讨论),前者,比如“主语”,即指“主谓结构”中第一项成分本身。后者,比如“zhuyu”,即指“主谓结构”中第一项成分所在的位置。其他“谓语、述语、中心语……”等等概念都是如此。本文中对“组成成分名称”,用汉字称说,而且不区分“述语 1”和“述语 2”、以及 4 个“中心语”;对“句法结构位置”,用拼音代码标记,区分“shuyu1”和“shuyu2”,“zhxyu1”、“zhxyu2”、“zhxyu3”、和“zhxyu4”等。

此外,还有两点需要说明,第一,尽管短语的功能类大致已经反映了一个短语出现在上述结构位置上的情况(参见 2.1.2 中的论述),但有时还需要在分类的基础上再辅之以属性特征描述,才能更准确地反映一个短语的实际句法功能。比如“正在洗澡”跟“打击盗版”虽然都是 vp,但前者不能作主语,后者可以(如“打击盗版对知识产权保护有利”)。对这样的差别,就可以通过这两个短语的“zhuyu”范畴的不同取值反映出来。第二,不同功能类的短语,在上述范畴上的表现并不是平等的。比如对于一个形容词短语,通常就不用考虑它能否出现在述宾结构的述语位置上,因为一旦它能出现在这个位置上,它也就不是 ap 而是 vp 了。要反映这方面的差别,从技术上讲,可以把上面“zhuyu”、“shuyu1”等范畴名目进一步细分成“apzhuyu”、“vpzhuyu”、“apshuyu1”、“vpshuyu1”、……等等,然后对这些更细的范畴分别设置所谓的默认值(default)来实现。比如“apshuyu1”的默认值设为“否”,而“vpshuyu1”的默认值就比较适合设为“是”,等等。但目前我们没有采用这种更精细的做法,对上述 16 个范畴,各类短语的默认值是简单的一视同仁,都取值为“是”。而如果一个短语某项句法功能范畴取值为“否”,就在规则中显性地说明。比如组合定中式 np(如“老师的女儿”)不能再直接作定语,这时就要显性地标明其“dingyu”属性值为“否”。这样处理有两个理由:(一)“言有易,言无难”。说一个短语有某项功能通常比说它没有某项功能显得稳妥和可靠。换句话说,我们认为,把一个短语的句法功能范畴默认值设为“否”,冒的风险比默认值设为“是”要大,所以我们把短语所有的句法功能范畴默认值都先设为“是”;(二)并不是所有的短语句法功能范畴对每类短语都平等地起作用。比如对 np 来讲,“buyu”和“lwqx”、“lwhx”这些范畴就没有意义。因为规则库中根本就没有由 np 参与构成述补或者连谓结构的情况,即汉语中 np 根本就不可能构成述补结构、连谓结构。因此,即使 np 的“buyu”、“lwqx”等属性值为“否”,也没有必要在 np 的规则中一一说明“\$. buyu=否、\$. lwqx=否”等等,即在具体的规则描述中,我们可以“合理地偷懒”。

## 二. 语义范畴

短语的语义范畴跟词的语义范畴完全一样,在广义配价的语义表达框架下,也是语义类,配价数,论元角色类型及语义选择限制、配价成分变化情况等内容。

但需要说明的是,一方面目前语义研究的基础还不够扎实,另一方面照顾到实际操作的方便,我们强调现阶段以句法为主,语义为辅的处理策略。对短语的语义属性描述,总的来讲是粗线条的。各种语义属性基本是简单地从中心词继承,碰到从词组合成短语后搭配性质发生显著变化的情况,则在规则中相应地加以描述说明。比如动词“搬”的“语义类”属性值是“搬移”,“配价数”的值是 2,它形成的述补式 vp“搬走”在这些语义性质上没有变化,“语义类”仍是“搬移”,“配价数”仍是 2,这些就简单地从中心动词“搬”那里继承。但“搬走”的搭配性质还是发生了一些变化,比如“搬走”不能再跟处所成分搭配了,请比较“搬屋里”与“\*搬走屋里”。对此,就要在述补式 vp 规则中加以描述说明,即“搬走”的“处所”配价成分取值为“否”。像这样的情况,在语言成分的组合中实际上是相当多的。下文论及主要短语的结构规则时,会结合具体问题的分析展开讨论。

关于上述句法语义范畴，还有两点补充说明：

(一) 本文强调上述句法语义范畴的开放性和层级性。所谓开放性，是指我们主张从方法论的意义上认识这些范畴，而不主张在本体论的层面上诠释这些范畴。也就是说，范畴数量和名目都是可以调整的。根据目标的不同以及反馈效果的好坏，可以对范畴进行增、删、合并、再细分等调整。比如词类范畴，就是相对的、开放的、可调整的。所谓层级性，就是指在抽象概括的不同程度上归纳语言知识。句法知识范畴跟语义知识范畴的差别，在我们看来，就在于抽象概括程度的不同，前者的抽象度高于后者。比如对词汇而言，名词这个句法分类结果，就比“事物”、“人”这些语义分类结果要抽象概括得多。从实际操作方面考虑，将范畴建立在不同层级上，更易于操作。比如“动词”这个范畴可以用来帮助确定一个词是否能带宾语构成述宾结构。而要具体看一个动词能带体词性宾语还是谓词性宾语，可以在属性特征的层次上描述，不一定要在词类的层级上描述<sup>47</sup>（词类的抽象概括度比属性的抽象度高）。至于要了解一个体宾动词具体能带哪些名词性成分作它的宾语，就可以在语义层级上用配价和论元选择限制等语义范畴来做出说明。这样实际上就是把一个词的搭配能力放在了不同的层面上进行描述。根据目标的要求，系统可以选择调用其中若干个或者全部层级的语言知识，这样一方面便于知识的模块化，一方面也有利于避免范畴间的相互干扰。

(二) 就一类短语所包含的成员数量及其复杂性来说，显然 dj、np、ap、vp 等应该看作是开放型的短语。而 mcp、mp、tp、sp、pp、dp 等则是相对封闭的（尽管封闭的程度有所不同）。对于封闭的短语类，可以作穷尽性的研究，而对于开放的短语类，则可以一面归纳具体的句法语义限制条件，一面探索描述其句法语义限制的一般方法。显然，后一个方面的研究工作对整个现代汉语短语结构系统的规则研究更具理论意义。同时，这方面的具体研究结果也对描述封闭类的短语结构规则有直接的支持作用。因此，下面第三章将重点对开放类的短语结构规则进行描述，其中也会根据说明问题的需要提及一些封闭类短语的规则（关于封闭类短语的结构规则可参见附录二）。

## 附 注

<sup>1</sup> 参见朱德熙（1985）《语法答问》，商务印书馆，P68。

<sup>2</sup> 参见詹卫东等（1997）《现代汉语短语本位语法体系在机器翻译中的应用及其问题》，载吴泉源、钱跃良 主编《智能计算机接口与应用进展》，电子工业出版社 1997 年版。（第三届中国计算机智能接口与智能应用学术会议论文集）

<sup>3</sup> 所谓“经验”，从形式上看，具体是指这样的一系列“能与不能”的判断，即一个结构的两个成分之间能不能插入“是不是”（如“天很黑”可以变成“天是不是很黑”）；能不能插入“不”（如“木头桌子”不能变成“\*木头不桌子”）；……等等；从意义上看，就是一个结构跟另一个结构是否具有“相同”的抽象意义，比如“拍桌子”跟“看电影”都有“动作——对象”这样的抽象意义关系，而“拍桌子”跟“看不完”之间则没有共同的抽象意义平行性，……等等。这些语感经验的判断结果是语言学家假设的短语结构关系范畴得以成立的语言事实基础。

<sup>4</sup> 参见陆俭明（1994）《关于词的兼类问题》，载《中国语文》1994 年第 1 期。

<sup>5</sup> 关于以句法结构及结构所确定的功能位置来确定语言成分的分类，还可参看陈小荷（1998）《从自动句法分析角度看汉语词类问题》，这篇论文在“98 现代汉语语法学国际学术会议”（北京大学）上宣读，并在北京大学计算语言学研究所的讨论班上作过报告。陈文列出了 8 种句法结构，13 种句法成分。

<sup>6</sup> 参见朱德熙（1982）《语法讲义》，商务印书馆，P125-139。

<sup>7</sup> 袁毓林（1998）《中文信息处理中的语言难题征答》，载《语言文字应用》1998 年第 4 期，P72。

<sup>8</sup> 表中“+”表示是该类型短语的典型功能；不填表示不是典型功能。坦率地说，所谓典型和非典型，是对一个短语在某个结构位置上出现频率的比较笼统的定性判断，目前主要是由个人语感来确定的。如果有大规模的标注了短语结构功能信息的汉语语料库，就可以量化地来区分各种不同功能类的短语的典型功能和非典型功能了。目前还做不到这一点。尽管如此，个人语感的判断结果对进行本文的研究工作也是可以接受的，并不造成原则上的问题。

<sup>9</sup> 这里为表格的简单起见，把连谓结构前项位置和后项位置并作一个区别特征加以看待了。实际上这两个位置是有差异的，在我们下文具体的短语结构规则描述中可以看到，pp 和 ap 短语都允许出现在连谓结构后项位置，但不能出现在前项位置。

- <sup>10</sup> 实际上,也不见得一定就需要这些类别来标识像“慢性、通红”这样的词语跟一般形容词的差异。理论上,要把“通红”这样的词跟其他形容词(比如“红”)区别开,可以通过给前者冠以“状态词”这样的词类名称,给后者冠以“形容词”这样的词类名称来实现。但值得注意的是,这仅仅是一种处理方式而已。完全可能采用别的处理方式来达到区别的目的,比如,可以仍把“通红”处理为形容词,同时附加一些特别的说明就可以了,比如“通红”不能单独作定语(\*通红房子——红房子),能作状态补语(脸晒得通红——\*脸晒得红),等等。不难看出,词的功能分类是相对的。
- <sup>11</sup> 关于这一点,最好的例子莫过于赵元任先生论及音位单位的多种可能性的经典论文了。参见赵元任(1957),The non-uniqueness of phonemic solutions of phonetic systems, In Martin Joos eds., Readings in Linguistics. 中译文见袁毓林(1992)主编《中国现代语言学的开拓和发展——赵元任语言学论文选》,清华大学出版社1992年版,P20-63。赵先生的文章虽然不是直接针对分类问题的,但实际上,音位单位的确立,完全可以看作是对语音基本单元的一种分类结果。一个音位,也可以理解为“一类音”。
- <sup>12</sup> 从这种观念来看词跟短语,二者之间的形式界限就比较模糊。在功能分类的框架下,这是允许的。而在结构分类的框架下则不可以。关于词跟短语这两级语法单位之间的关系,可参看杨成凯(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年第6期;刘丹青(1995)《汉语语法单位分级理论的再探讨》,载《汉语学习》1995年第2期;郭锐(1996)《汉语语法单位及其相互关系》,载《汉语学习》1996年第1期。
- <sup>13</sup> 关于语义场,词的义素分析等,可参见符淮青(1996)《词义的分析 and 描写》,语文出版社,第三、七、九章有关讨论。
- <sup>14</sup> 参见袁毓林(1998)《汉语动词的配价研究》,江西教育出版社1998年版。
- <sup>15</sup> 参见胡明扬(1980)《“格”辨》,载《语言学译丛》(第二辑),中国社会科学出版社1980年版;杨成凯(1986)《Fillmore的格语法理论》(上,中,下),载《国外语言学》1986年第1-3期。
- <sup>16</sup> 参见徐烈炯(1988)《生成语法理论》,上海外语教育出版社1988年版;汤廷池、张淑敏(1996)《论旨网格、原参语法与机器翻译》,载《中国语文》1996年第4期。
- <sup>17</sup> Schank, R. C., ed. 1975. Conceptual Information Processing. Amsterdam: North-Holland.
- <sup>18</sup> Quillian, M. R. 1968. Semantic Memory, In Semantic Information Processing, MIT Press. Simmons, R. F. 1973. Semantic Networks: Their Computation and Use for Understanding English Sentence, In Computer Models of Thought and Language.
- <sup>19</sup> 参见徐烈炯(1995)《语义学》,语文出版社1995年第2版,第5.6小节。邹崇理(1995)《逻辑、语言和蒙太格语法》,社会科学文献出版社1995年版。
- <sup>20</sup> 参见陈力为、袁琦主编(1995)《中文信息处理应用平台工程》,电子工业出版社。《信息处理用汉语语义词典》是“905”工程项目,国家“八五”重点科研项目“中文信息处理应用平台工程”(90年5月开始实施)的子课题。主要参加单位包括清华大学、北京语言文化大学(前北京语言学院)。
- <sup>21</sup> 林杏光主编(1994)《现代汉语述语动词机器词典》,北京语言学院出版社1994年版。
- <sup>22</sup> 参见董振东(1998)《语义关系的表达和知识系统的建造》,载《语言文字应用》1998年第3期。目前董振东先生已经将他开发的汉语语义知识库放在了因特网上供研究人员和有兴趣者免费下载,可以访问网站<http://www.how-net.com>查询。
- <sup>23</sup> 值得一提的是,论及语义知识在信息处理中的作用,也可以把推理包括进来。比如由“鸟会飞”,“麻雀是一种鸟”,可以推出“麻雀会飞”。但其中“麻雀是一种鸟”这一知识,既可理解语义知识也可理解为百科知识。这样一来,语义知识跟所谓百科知识的界限就非常模糊了。本文的倾向是尽可能关注语言内的语义知识,而避免涉及百科知识。尽管语义知识跟百科知识在不少情况下难以分清(比如上例),但的确也有不涉及百科知识而是比较纯粹的语义知识的例子,比如纯粹从逻辑上讲,“修理”的对象也可以是人,但汉语并不这么表达(\*“修理小王的师傅”)。还有如“吃饭”跟“抽烟”。南方话可以说“吃烟”,可见就逻辑而言,“吃”也是可以跟“烟”搭配的,只不过普通话中不允许。像这样的搭配知识,才是比较严格的语言内的语义知识。而像“麻雀是一种鸟”这样的知识,我们认为,首先是百科知识,之后才是语义知识。正是出于这些考虑,本文不特别强调语义知识的推理作用,而是把推理所依据的知识主要归结为百科知识。
- <sup>24</sup> 关于“后胎”是“自行车”的一个“部分”这样的语义知识,在我们的词典中是这样表示的:首先“后胎”的“语义类”属性值为“构件”(即它是另一个物体的一部分),其次它的“配价数”属性值为1(即它可以跟一个名词发生语义关联),它的配价成分的语义角色类型是“主体”,“主体”的语义选择限制则直接用“\*车”来表示,这里“\*”是通配符(不是一般语言学中表示错例的符号),“{主体: [语义类: \*车]}”的含义是:“后胎”的“主体”是包含“车”这个汉字的那些名词)。
- <sup>25</sup> 参见李临定(1990)《现代汉语动词》,中国社会科学出版社1990年版。
- <sup>26</sup> 孟琮等(1987)《动词用法词典》,上海辞书出版社1987年版。
- <sup>27</sup> 参见詹卫东(1997a)《词的语义分类在汉英机器翻译中所起的作用以及难以处理的问题》,载陈力为、袁琦主编《语言工程》,清华大学出版社1997年版(全国第四届计算语言学联合学术会议论文集,JSCL'97)。
- <sup>28</sup> 变化类型的数目可以根据分析目标的不同进行设置,并不是固定的。在本文的分析中,具有“性状变化”

- 或“位移变化”特征的动词的句法性质，讨论得相对多一些；具有“数量变化”和“领属变化”特征的动词则讨论得比较少。
- <sup>29</sup> 本文对汉语实词的语义分类，主要参考了张普（1995）《信息处理用现代汉语语义分析的理论与方法》，陈群秀、张普（1995）《信息处理用现代汉语语义分类体系：属性分类》（载《中文信息处理应用平台》，电子工业出版社 1995 年版）的研究工作，并根据我们对语义分类的认识以及本文研究工作的目标，进行了简化。
- <sup>30</sup> 本文把配价数的取值范围定在 0—3 之间，像“例如”这样的动词，不联系体词性成分，配价数为 0，像“送”这样的动词，可以联系三个体词性成分，配价数为 3。其他动词的配价数介于这二者之间。但需要说明的是，这样处理仅仅是一种简化的方式。关于配价成分应该如何记数，牵涉到对汉语动词整个配价系统的认识，同时也涉及到具体实施时的工作量问题。理想的情况也许是按照动词配价成分的核心语义角色类型的数目来定配价数，比如像“写”，可以搭配“主体”、“客体”、“工具”、“处所”等语义角色类型，配价数就为 4。但这是以对现代汉语整个动词配价系统的认识已经比较清晰为基础的。目前对动词核心语义角色类型的性质及在分析中的实际效用了解得还不是非常清楚，因此我们暂时走保守的路线，先简单地把配价数定在 0—3 的范围内，操作上则以句法结构形式为主要依据。这类似于沈阳（1994）《现代汉语空语类研究》中的做法。
- <sup>31</sup> 关于动词配价成分角色类型的选取主要参考了李临定（1990）《现代汉语动词》，和鲁川（1995）《现代汉语的语义网络》（载《中文信息处理应用平台工程》，电子工业出版社 1995 年版），同时也做了一些调整。“主体”、“客体”等等语义角色类型概念，本文不作严格定义。“主体”大致包括一般所说的“施事”、“当事”等语义格，“客体”大致包括一般所说的“受事”、“对象”、“目的”、“结果”……等等语义格。值得一提的是，我们不认为所有动词后表示空间方位的宾语成分都能算作动词的处所格。比如“喜欢南京”中的“南京”不能认为是处所宾语，也不能算作“喜欢”的处所类型的配价成分。理由是这里的“南京”在句法表现上跟“喜欢巴金的小说”里的“巴金的小说”是平行的，都是受事宾语，而不是跟处所宾语的性质接近。
- <sup>32</sup> 所谓“核心”和“外围”的语义角色，只是表述上的区分，我们并不在术语（或属性范畴）的意义上使用这两个概念（即在规则中并不用它们）。值得一提的是，格语法对不同类型格的区分是所谓的“必选格”和“可选格”。在我们看来，就对汉语句法分析的制约作用而言，区分“必选”跟“可选”的语义角色类型，并没有特别大的意义。
- <sup>33</sup> 参见詹卫东等（1998）《基于词组本位语法的语义模型》，载新加坡《中文与东方语言信息处理学会学报》Vol. 8, No. 1, 1998。
- <sup>34</sup> 也有学者用“内在处所格”和“外在处所格”这样的术语来称说，见袁毓林（1998）《汉语动词的配价研究》，江西教育出版社 1998 年版，P102。
- <sup>35</sup> “写在黑板上”的结构，本文分析为：[[写 在] 黑板上]，即“写”跟“在”先结合成述补式 vp，再带处所宾语。之所以采用这种分析方式，可参见下文 3.4.2 述补式 vp 相关规则的说明，也可参阅《语法讲义》，P182。值得一提的是，把“写在黑板上”分析成 pp “在黑板上”作补语，也不妨碍本文这里关于“处所”与“空间”对立的分析。
- <sup>36</sup> “洗”的“客体”配价成分，我们就用了从反面限制的办法，<sup>-</sup>号表示逻辑“非”。“具体物 天体…”表示“洗”的客体应该是“具体物”，但要排除“天体”等具体物。
- <sup>37</sup> 方括号里的变化项还可以进一步扩展。有的动词比如“举”，它的客体配价成分要发生“位置”变化，并且这个变化是有方向的，只能“向上举”，如果词典要标记“举”的这个性质，就可以在括号内的“位置”项内部进一步标明为“位置：向上”。
- <sup>38</sup> “|”号表示逻辑“或”的关系。
- <sup>39</sup> 参见俞士汶（1988）《自然语言的语义分析技术》，载《中国计算机用户》1998 年第 5 期。翁富良、王野翊（1988）《计算语言学导论》，中国社会科学出版社 1998 年版，P96-99。
- <sup>40</sup> 关于分层级描述词语的语义信息的思想，也可参见袁毓林（1998）《汉语动词的配价研究》。袁在对汉语动词的配价进行系统研究的过程中，也主张建立一个层级体系来刻画动词的配价性质。
- <sup>41</sup> 参见孙宏林（1994）《信息处理用汉语语义词典的描述方法》，载《现代语言学·第三届全国语言学会议论文集》，语文出版社；陈群秀（1996）《信息处理用现代汉语语义分类体系的设计思想》，载《计算机时代的汉语和汉字研究》，清华大学出版社。
- <sup>42</sup> 詹卫东等（1998）《基于词组本位语法的语义模型》，载新加坡《中文与东方信息处理学会学报》Vol. 8, No. 1, 1998。
- <sup>43</sup> Bloomfield, L. 1933. Language, Holt, Rinehart & Winston, New York. 中译本是袁家骅等译《语言论》，商务印书馆 1980 年版，P239。
- <sup>44</sup> 有关名词加动词格式的短语组合的功能的详细讨论，可参见马真、陆俭明（1996）《“名词”+“动词”词语串浅析》，载《中国语文》1996 年第 3 期。
- <sup>45</sup> 参见史金生（1988）《汉语向心结构的功能偏移》，载《绥化师专学报》社科版，1988 年第 4 期。
- <sup>46</sup> 王红旗（1995）《动结式述补结构配价研究》，郭锐（1995）《述结式的配价结构与成分的整合》，载沈阳、郑定欧主编《现代汉语配价语法研究》北京大学出版社。
- <sup>47</sup> 比如“企图”跟“吃”。它们的分布不同，“企图”后面不能出现体词性宾语，“吃”后面不能出现谓词性

宾语。如果要特别强调这种差别，可以把“企图”跟“吃”归到两个类中去。但也可以不在词类的层级上区别，而放到属性特征的层面上进行描述。《现代汉语语法信息词典》就是用“体谓准”这个属性来描述一个动词到底带哪种类型的宾语的。