

# 确立语义范畴的原则及语义范畴的相对性\*

## The Principles of Setting Up A Semantic Category System

詹卫东 北京大学

**摘要** 本文以语义知识在中文信息处理应用中所能发挥的作用，及其发挥作用的方式为出发点，来讨论一个语义知识系统确立语义范畴所应遵循的大原则。大致可以概括为三：一是目标驱动原则（实用主义原则）；二是形式验证原则；三是可形式化原则。在此基础上，本文还试图阐明，跟语言模型中其他层面的范畴一样，语义范畴与其看作是语言中固有的本体，不如看作是研究人员构造的产物，因而对语义范畴的相对性应有清醒地认识，否则会造成许多不必要的争论，结果妨碍人们去探寻真正有价值的问题的答案。

**关键词** 句法范畴 语义范畴 论元 配价 语言模型

**Abstract** First of all, this paper explains what people can do and in what way people do depending on so called semantic knowledge in Chinese Information Processing systems. And then, the author discusses the principles of setting up a semantic category system. Generally, there are three main principles need to be taken into account. The first is the principle of goal-driven or pragmatism principle. The second is the principle of proof-test through linguistic forms. The third is the principle of computability. The author also tries to clarify that one of important characteristics of semantic categories is relativity. And this is the same as so called syntactic categories. It's worthy to note that a researcher could be involved in some unnecessary disputes instead of paying more attention to the valuable questions if the relativity of semantic categories is not understood enough.

**Keyword** Syntactic Category, Semantic Category, Argument, Valence, Linguistic Model

近年来语义研究在整个语法研究中的重要性日益显著，与此同时，有关语义的一些基础问题在研究过程中又时常引起争论，比如动词配价的性质问题，动词论元角色的语义类型如何确定，动词配价与句式之间的关系问题，等等[参考文献 10, 11, 12, 16]。对此，我们可以结合计算机处理汉语的实际需求来看待有关语义范畴的争议，或许能够提供给我们一个更为有效的参照系。本文尝试从这种思路出发，就确立语义范畴的基本原则进行讨论，并试图阐明语义范畴跟语言中的其他范畴一样，都具有相对性，一些表面的争议问题其实不过是一些伪问题，从而将研究者的关注点引向真正的问题所在。

### 1 确立语义范畴的原则

跟语言学中的大多数问题（比如词类问题，主宾语问题等）一样，确立语义范畴时碰到的问题如果能从目的着眼来审视，可以看得更清晰，也就是说，首先应该弄清楚，我们要拿语义范畴来干什么用？

#### 1.1 语义知识的作用

1.1.1 从知识的表现形式上讲，语义知识包括两种类型：一种是所谓范畴型知识，即可以表述为“属性：值”（即复杂特征集）这种形式的知识；一种是所谓规则型知识，即可以表述为“条件->动作”（即产生式规则）这种形式的知识（文献[19]）。举例来讲，在一部语义词典中，可以记录有下面这样的信息：

---

\* 本文研究得到国家重点基础研究发展规划项目“面向中文信息处理的现代汉语动词论旨结构系统和汉语词语语义分类层级系统研究”（项目编号：G1998030507-1）资助。

苹果 := [语义类:可食物,物态:固体]; 汤 := [语义类:可食物,物态:液体]; 圆 :=[语义类:物形]  
吃 := [语义类:获取,论元:2]{施事:[语义类:动物|人类],受事:[语义类:可食物,物态:固体]}

需要说明的是,以“属性:值”形式记录的范畴型语义知识又包括两种情况,一种可以称之为简单范畴型语义知识,刻画的是一个对象的基本语义属性(上面以[ ]括出的内容),比如[语义类:可食物];一种可以称之为复杂范畴型语义知识,刻画的是两个以上对象之间的基本语义关系(上面以{ }括出的内容),比如用来描述动名语义约束关系的{施事:[语义类:动物|人类]}(这里“|”表示逻辑上的“或”关系)。不难看出,复杂范畴型语义知识的“属性:值”形式表示中,“值”是用若干简单范畴型语义知识来刻画的。

同时我们还有下面这样的规则型的语义知识<sup>1</sup>:

如果一个名词的语义性质跟一个动词的受事论元的语义要求吻合(条件),那么该名词可以跟这个名词组成“动作——受事”的组合形式(动作)。

显然,上述范畴型语义知识加上规则型语义知识,可以用来说明“吃苹果”(能说)和“吃汤”、“吃圆”(不能说)之间的差异。

1.1.2 就实际的自然语言处理应用系统来说,语义知识在不同的应用领域都可以发挥作用。概括地说,语义知识的作用是用来排除自动分析过程中碰到的各种歧义,这可以通过两种比较典型的情况——排除句法结构分析的歧义,以及在多义词义项判定中确定多义词的义项——来加以说明。

先来看语义知识在句法结构分析中所起的作用。

短语结构 P1“修汽车的后胎”跟 P2“修汽车的王师傅”在句法层面上,都可以抽象为“V+N1+的+N2”这样的序列(记作 M),但其结构分析结果不同,前者是 A.[修 [汽车 的后胎]];后者是 B.[[修 汽车] 的 王师傅]。要让计算机能够将 P1 的结构分析为 A,将 P2 的结构分析为 B,需要用到“施事”、“受事”、“整体——部分”这样的语义范畴。对于序列 M,有规则:如果 N1 跟 N2 的语义性质跟 V 的受事论元的语义要求吻合,并且 N1 跟 N2 之间有“整体——部分”的关系,那么 M 的结构应分析为 A;如果 N1 的语义性质跟 V 的“受事”论元的语义要求吻合,N2 的语义性质跟 V 的“施事”论元的语义要求吻合,那么 M 的结构应分析为 B。

再来看语义知识在多义词义项判定中所起的作用。

“想主意”(以“想<sub>1</sub>”代表)跟“想女儿”(以“想<sub>2</sub>”代表)中的两个“想”意思不同,前者是“思考,思索”的意思(形式上表现在可以变换成“想出主意”,不能变换成“想念主意”);后者是“思念,想念”的意思(形式上表现为可以变换成“想念女儿”,不能变换成“想出女儿”),对此,需要“想”的“受事”论元的语义约束信息来帮助计算机做出判断。“想<sub>1</sub>”的“受事”论元要求不能是[语义类:人]这样的名词,相反,“想<sub>2</sub>”的“受事”论元要求是[语义类:人]这样的名词。“主意”跟“女儿”两个名词的“语义类”属性取值分别为[语义类:方法]和[语义类:人],计算机可以通过匹配知道“想”跟“主意”搭配时是“想<sub>1</sub>”,跟“女儿”搭配时是“想<sub>2</sub>”。基于上述语义知识,计算机不仅能在简单述宾结构中确定“想”的义项,也能确定“张三想的主意”(复杂的定中结构)中“想”的义项。

1.1.3 无论从自然语言处理系统的实践,还是从语言学研究本身的理论追求来看,都可以发现,人们构造语义知识系统的根本目的,实际上跟句法知识系统是同样的,都是为了描述(在某种意义上也就是解释)某个“语言形式”是否能说/可被接受,以及能说/可被接受的若干“语言形式”之间是否存在一定的变换关系。这一点在语言本体研究中也许因理论的厚重而妨碍了研究者的认识,而如果以自然语言处理系统为应用背景来观察,就比较清晰。实际上,如果我们能够透过“句法”与“语义”表面的理论分野(理论形式上的差别)去看它们背后的共性,就不难了解到,我们完全可以用看待“句法范畴”的眼光来审视所谓的

“语义范畴”。

举例来说，只需要凭借“名量结构作定语，不作状语”、“副词作状语，不作定语”这样一些所谓的句法知识，我们就可以判定“一份学习”和“正在文件”不能说/不被接受，而能说的短语结构 P3 “一份学习文件”和 P4 “正在学习文件”中的两个形式“学习文件”之间存在着某种差异。值得指出的是，差异实际上都可以（也必然）通过一定的变换体现出来，比如 P3 中的“学习文件”可以变换成 P3' “学习过的文件”，P4 中的“学习文件”不可以变换成 P4' “学习过的文件”（即“正在学习过的文件”不能说）<sup>2</sup>，等等。

如果我们希望说明“吃苹果”能说，“吃紫色的狂怒的思想”不能说；并且想要描述“吃苹果的人”跟“吃苹果的方法”之间的差异，一般来讲，仅凭借“名词”、“动词”、“述宾结构”、“定中结构”这样一些句法范畴就不大够，这时就需要通过诸如“施事”、“受事”、“语义类”、“抽象事物”、“具体事物”、“人”、“食物”等等一些所谓的“语义范畴”来达到目的。

简而言之，人们是用“句法范畴”跟所谓的“语义范畴”在干差不多同样的事情，而且干事情的方式也是一样的（之所以要区分它们，是为了整体理论模型的清晰和更好的模块化，可以看作是技术处理的结果，而非“事情本来如此”，下文还有更多讨论）。

基于上述认识，当我们面对似乎非常抽象，不易把握的“语义范畴”时，现在就有两个可以让人相对感到“踏实”一些的办法了，一是以目的（要干什么事）或者说是实践（能做什么事）作为思考背景和评判标准；二是可以拿相对清楚一些的句法范畴作为攀比对象，如果对一个语义范畴的性质感到不好理解，不妨拿一个句法范畴来比较比较看。

本文讨论确立语义范畴的原则以及对语义范畴相对性的论述都建立在上述认识基础上。

## 1.2 确立语义范畴的三个原则

1.2.1 原则之一：语义范畴的设置应该是目标驱动的，遵循实用主义的原则，即在明确的目标下，所确立的语义范畴如果做到够用就可以了（参见文献[1, 2]）。

拿自然语言的理解和自然语言的生成这两个不同的目标来说，理解对语义范畴的精度要求相对就可以低一些，生成对语义范畴的精度要求则相对较高。比如：有学者指出应该区分“当事”（*essive*）和“感事”（*experienter*），理由是汉语的人可以感觉到“水很冷”跟“我很冷”之间的区别。这当然是不错的。不过，如果从句子理解的角度来看，要得到正确的句法分析结果，把这两句中的“水”和“我”都分析为“冷”的“施事”（或者更为概括的“主体”）也未尝不可，并不一定需要再设置“当事”和“感事”两个论元角色。而对生成来说，仅有“施事”这个范畴，办起事来可能就不利索了。因为人们希望计算机可以自动生成 A. “这么冷的水”，而不要生成 A' “这么冷的我”，这时候如果有“当事”和“感事”这两个语义范畴，并在这两个范畴的基础上构造一条规则：“当事”可以在“ap+的+np”的 np 位置出现，使得整个结构成为定中 np；“感事”不能在上述格式中的 np 位置出现，我们就比较容易区分出 A 是能说的，而 A'是不能说的。

同样是做自然语言理解，如果是涉及到篇章范围的理解，可能需要的语义范畴就更多一些，如果只在短语结构范围内理解，需要的语义范畴就相对少一些。比如：涉及到篇章推理时，需要知道动词之间的语义推导或蕴含关系。只有描述了“买”和“拥有”这两个动词的**致使**关系（*Cause Relation*），才能从“张三买了一本书”，推导出“张三拥有了一本书”这个结果；只有描述了“打鼾”和“睡觉”这两个动词的**整体-部分**关系（*Proper Inclusion*），才能知道“张三正在打鼾”蕴含了“张三正在睡觉”。如果只在短语结构范围内进行理解，目标是给出句法分析结果和简单的语义信息，一般就用不着这些语义范畴。

不难看出，应用目标实际上也提示了人们确立语义范畴时所选取的语言分析对象的“尺寸”。像“施事”、“受事”这样的语义范畴，是分析短语结构/句子级语言对象的结果，而像

“蕴含”这样的语义关系范畴，则是分析篇章级语言对象的结果。人们构造语义知识模型和大规模语义知识库的实践也清楚地显示出，以不同尺寸的语言单位作为观察对象，是跟不同的应用目标相适应的，由此形成的语义知识的组织方式因而也呈现出不同的面貌<sup>3</sup>。

这条原则提醒一个语义知识系统的构建者去思考某个语义范畴需不需要设置，即必要性（necessity）的问题，并且在思考这个问题的时候，务实地走“实用主义”路线。

1.2.2 原则之二：语义范畴的设置应该以形式上有可观察的依据为准则。

如前所述，人们需要语义范畴跟需要句法范畴的理由是一样的，相应地，确立语义范畴跟确立句法范畴的直接依据也是一样的，就是都要有可观察的形式证据。尽管人们谈到“意义”这个不可捉摸的东西时多少有些靠理性“品味”而不是靠“观察”，但实际上所有的所谓“意义”问题，都必然要“体现”为“形式”问题。

拿“施事”这个语义范畴来说，它在形式上的依据是：

- (1) 充当“施事”角色的 np 可以出现在 vp 前，并且跟 vp 形成句法上的所谓“主谓结构关系”；
  - (2) 充当“施事”角色的 np 一般不能出现在 vp 后跟 vp 形成句法上的所谓“述宾结构关系”；
  - (3) 充当“施事”角色的 np 可以出现在“被+np+vp”格式中的 np 位置；
  - (4) 充当“施事”角色的 np 一般不能出现在“把|用|在|对|…+np+vp”格式中的 np 位置；
- ……

这就好比我们确立句法上的“名词”范畴，也同样是在用这类的形式依据：

- (1) 可以出现在“mp+np”后（mp 表示“数量结构”），跟 mp 形成所谓的“定中结构关系”；
  - (2) 不能出现在“vp+np”格式中的 np 位置形成所谓的“述补结构关系”；
  - (3) 不能出现在“很|不+np”格式中的 np 位置形成所谓的“状中结构关系”；
- ……

再看“上下位语义关系”范畴的形式依据（下面的例子转引自网上有关 WordNet 的论文，也可参见文献[21]）。如果要说明“A pistol is more dangerous than a rifle”（手枪比步枪更危险）可以接受，而“A pistol is more dangerous than a gun”（手枪比枪更危险）不容易接受，就需要建立“上下位关系”这样的语义范畴，并说明“手枪”跟“步枪”没有上下位关系，是同位关系，二者的共同上位概念是“枪”，同时建立一条这样的规则：比较句中两个比较项不能满足上下位关系。

上下位关系有时还可以跨越句子，用来解释小句间关联的合法性，比如“I gave him a good novel, but the book bored him”（我给了他一本不错的小说，但那本书让他感到乏味）的小句衔接是可接受的，而“I gave him a good novel, but the catsup bored him”（我给了他一本不错的小说，但调味番茄酱让他感到乏味）的小句衔接性就不容易让人接受。语义知识模型可以用——“书”和“小说”之间有上下位语义关系，而“番茄酱”跟“小说”之间没有上下位语义关系——来对上述可观察的现象加以解释。

以上这些例子都可以说明，语义范畴跟句法范畴干着同样的事情，都是在对人们观察到的“语言形式”进行刻画。反过来说，它们赖以存在的理由也非常明确，就是能够刻画语言形式的某些特征（能说/不能说，形式之间有什么变换关系等等）。当然，语义范畴本身也必然负载着“意义”内容，只不过有“意义”不是它们的直接存在理由。这就好像“名词”这个所谓的句法范畴也有“意义”（名词的意义大概是“表示事物”吧），但语言学上需要“名词”这个范畴的直接依据并不是因为它“表示事物”，而是“名词”这个范畴能够用来刻画一系列的语言形式特征。同样地，人们当然可以用“事件中自主运动的有意识的主体”（文献[3]）这样的描述来定义“施事”，但这不是“施事”进入“语言学殿堂”时应有的扮相，“施事”这个范畴之所以能在语言学中立足，直接理由跟“名词”这个范畴能在语言学中立足的理由完全一样——因为“施事”这个范畴能刻画一系列的语言形式特征。

事实上,以往有关语义的研究中人们提出的语义范畴大都有对应的形式依据,只要是在比较语言形式差异的事实基础上提取的语义范畴就一定如此(文献[9, 10, 11, 12, 13, 14]),即满足我们提出的这条原则。不过,一些研究者倾向于用“意义描述”的方式去定义语义范畴,而不是用“形式特征描述”的方式去定义语义范畴,结果把语义问题搞得更为模糊。比如,根据徐烈炯、沈阳(1999)的介绍,Dowty(1991)把语义角色确立为“典型施事”和“典型受事”两大类,然后用一些特性,“意志性”、“感知性”、“使动性”、“变化性”、“渐成性”、……等来界定“施事”、“受事”等语义角色范畴。在我们看来,这样做只不过是将一个抽象的概念,用另外一些抽象程度差不多的概念作一些替换罢了。与其在所谓的“意义”里打转,为什么不面对我们直接可以面对的,而且应该直接面对的“语言形式”呢?

现在汉语语言学界基本都公认了“词类是以词语的功能(形式分布)标准作为分类原则”,其实这个原则完全适用于语义范畴,即“语义范畴也是以句法分布标准来确立的”。当然,就像人们可以给“名词”这个句法范畴附上一些“意义”上的说明一样,我们也不反对给“施事”这个所谓的“语义范畴”附上一些“意义”说明。

上述原则提醒一个语义知识系统的构建者思考一个语义范畴能不能设置,即能行性(practicability)的问题。只有给出一个语义范畴的形式特征说明,语言模型中才能够设置这个语义范畴。换言之,这条原则要求研究者应该把一个语义范畴落实到形式特征上<sup>4</sup>。

### 1.2.3 原则之三:语义范畴应该尽量做到可形式化。

所谓将语义范畴形式化,包含两个层面的含义,一层意思是指在知识表述上,基于语义范畴的知识应该以类似于上文提到的“属性:值”和“条件->动作”那样的形式化方式来表达;还有一层意思是指语义范畴要尽可能做到根据可观察的形式上的特征来加以严格的定义。前者是知识表示形式层面的形式化,可以看作是给语义知识披上一件形式化的“外衣”;后者是知识内容层面的形式化,可以看作是语言模型中的语义模块加入一个货真价实的形式化的“瓢”。

将语义知识打扮成形式化的样子并不特别困难,有像“属性:值”和“条件->动作”这样现成的形式化工具可以利用(当然具体落实上,还有一些技术处理上的技巧可以讲究),而要将语义知识的“瓢”(内核)给形式化,不是一件容易的事情。举例来说,存在着像下面这样的困难:

有些语言形式之间的差异可以“感觉”到,当然也可以在形式变换中体现出来,但不容易确立相应的“语义范畴”去刻画差异。比如 A. “跑步去”、B. “散步去”、C. “走路去”三个短语结构,我们可以感觉到 A 中“跑步”跟“去”之间存在两种关联: I. “跑步”是“去”的方式(形式上可以变换为:“以跑步的方式去到某地”), II. “跑步”是即将要进行的行为动作(形式上可以变化为:“去跑步”); B 中“散步”跟“去”之间只有一种关联:“散步”是即将要进行的行为动作(形式上可以变化为:“去散步”); C 中“走路”跟“去”之间也只有一种关联:“走路”是“去”的方式(形式上可以变换为:“以走路的方式去到某地”)。要刻画上述三个短语结构之间的差异,不大自然找到合适的“语义范畴”。从这三个具体的短语结构拓展开去,问题实际上可以进一步表述为:目前还没有合适的“语义范畴”来刻画汉语中两个 vp 之间的关系,尽管人们已经发明了“施事”、“受事”这样的“语义范畴”来刻画 vp 跟 np 之间的关系。

再比如一些比较复杂的歧义形式缺乏“语义范畴”去分化。“买多了”这个短语结构有歧义,既可以理解为“买这个动作行为的发生次数多”(形式上的特征是可以变换为“多买几次”、“你刚开始买菜不习惯没关系,买多了就会习惯的”),也可以理解为“买回的东西超出某个标准”(形式上的特征是可以变换为“买得太多了”),还可以理解为“买的东西数量上确实多”(形式上的特征是可以变换为“买多了有奖”)。对于上述歧义,目前还没有合适的“语义范畴”去加以分化,也即没有合适的“语义范畴”去刻画 vp 跟 ap 之间的关系。

通过上述例子可以知道,语言事实中只有部分可观察到的形式差异能够被现有的语言模型(包括其中的语义模块)说明,还有相当多的形式差异没有被刻画(指“以可形式化的严格方式来刻画”,而不仅仅是传统意义上的面向人的描写)。而且,在已有的模型中,因为各个“语义范畴”并非直接建立在基于形式特征的定义基础上,因而难以满足内部一致性的要求,这就使得其实用价值大打折扣。详细讨论请参见文献[17]。

概括说来,原则三提示一个语义知识系统的构建者思考语义范畴如何设置,即可形式化或可计算性(computability)的问题。努力的目标是,尽量以形式化的方式来整理表述已知的语义知识,同时尽可能去发明更多的基于形式特征定义的“语义范畴”,来覆盖更广的可观察的语言事实。简而言之,原则一引导我们考虑语义范畴是否够用,原则二、三则要求语义范畴尽量好用。

## 2 语义范畴的相对性

### 2.1 语义范畴的概括程度和精确程度是相对的

一个语义范畴可以看作是从某些形式特征的角度对语言成分的性质或语言成分间关系的性质进行的概括。概括一定有程度大小的差异。还拿上文的例子来说,“我很冷”跟“水很冷”中的“我”跟“水”可以区分为前者是“感事”,后者是“当事”,这是相对精确的一种概括结果,或者说是概括度低,精确度高,因为跟不区分“感事”和“当事”(管它们叫“施事”或“主体”)比起来,区分了“感事”和“当事”,实际上就意味着对“动名”语义关系的描述,可以在一个相对更精确的水平上操作。这就好像实词词类可以分成“体词”跟“谓词”,“谓词”又可分成“动词”跟“形容词”一样,分类的层级性等价于逐级精确性,同时这也就意味着各个类之间的相对性。如何去评判这种相对性的语义范畴呢?答案正是上文提出的三条原则,也就是够用、好用的实用主义评价标准。

不过,实用主义的评价标准在用的时候并不见得就真的“实用”。由于语言事实本身的充分复杂性,人们到目前为止,还没有找到一套普遍适用的办法来评价什么叫“好用,够用”。这也就造成现有的语义知识体系在确立语义范畴时相当的不一致性。从本文附录中表一和表二对几个语义知识体系的比较可以清楚地看出,语义知识库目前的现状正反映了语义范畴的相对性,这就如同各家的词类体系也不尽然相同一样(只不过因为词类的形式验证相对清晰,因而不同词类体系的差异也相对比较小)。

### 2.2 具体词语语义范畴信息描述的相对性

语义范畴的相对性还体现在:不同的语义知识体系在对具体词语的语义信息进行描述的时候,处理上可以是相对的,并不存在一个绝对的答案。

比如我们要描述“渴望”这个动词的语义信息。就可以不止一种方式。比如有下面这两种描述的可能:

I. 渴望 :- [论元数:2] {感事:[语义类:人],内容:[语义类:事物]}

II. 渴望 :- [语义类:心理活动,论元数:2] {施事:[语义类:人],受事:[语义类:事物]}

表面上看,上面对“渴望”这个动词的语义信息描述有很大差别,但实际上差别只是表面的、相对的。根据上文 1.2.1 节的分析,如果是按照自然语言理解的目标来要求的话,无论按照 I、II 哪种描述方式,基本都可以达到同样的效果。甚至 I、II 两种描述方式的信息都是可以相互转化的。很明显,只要在 II 中增加一条用于范畴转义的规则说明就行了:

IF V. 语义类=心理活动 THEN V. 施事=V. 感事, V. 受事=V. 内容 ENDIF

这个形式定义的含义是：如果一个动词是“心理活动”类动词，那么，它的“施事”论元可以转义为“感事”论元，“受事”论元可以转义为“内容”论元。

再看一些在配价数目上有争议的例子。比如有人认为“死”是一价动词，有人认为“死”是二价动词；有人认为“扔”是二价动词，有人认为“扔”是三价动词；有人认为“买”是三价动词，有人认为“买”是四价动词；此外还有关于汉语中是否有零价动词的讨论（有人认为汉语中有零价动词，如“刮风”、“例如”之类，有人认为汉语中没有零价动词）；等等。孤立地看，似乎这些观点之间水火不相容，答案非此即彼。其实不然。一个具体的语义知识表述系统可以根据需要把“买”处理为三价动词，也可以处理为四价动词。到底怎么样处理，只受两个因素制约，也就是（1）我们前面提到的目标驱动和实用主义原则（怎么处理好用）；（2）语义知识系统的整体性考虑。当把“买”跟“偷”都处理成三价动词的时候，可能在“配价数”这个语义范畴上无法区分“买”跟“偷”表现出来的形式差异，但可以在语义知识系统中另立范畴去刻画“买”跟“偷”的差别，比如就用“代价”这个范畴去区分“买”跟“偷”，“买”可以描述为：{代价:[语义类:货币|时间]}；“偷”可以描述为：{代价:[语义类:时间]}。有了这样的语义信息，照样可以说明 A “他十块钱买了我三个西瓜”可以接受，B “他十块钱偷了我三个西瓜”不能接受，同时也可以说明 A 跟 C “他半小时偷了我两车西瓜”之间的某个方面的“意义”差异和形式变换共性（比如都能变换成“np1+用+np2+vp+np3+np4”这样的格式）。

一个动词的价数与其说是该动词“本来”的属性，不如说是构造语言模型者赋值给它的。从这点上说，语言学中的“价”的性质更接近逻辑学中的谓词主目（argument）和数学中的函数变元（variable）概念（而不是接近化学中的“价/valence”）。一个函数有几个变元，完全取决于设计者的用意。当你需要处理像“张三、李四、王五各自买了桃、梨、苹果”的时候，你会把“买”设计成几价动词呢？我们的答案仍然是，处理成三价也好，四价也好，五价也好，六价也好，都有可能。不过要记住，当你选择某种处理方式的时候，你原来构造的整个语义知识系统就会发生改变，你必须做出相应的调整。

### 3 余论：关注语义研究中真正的问题

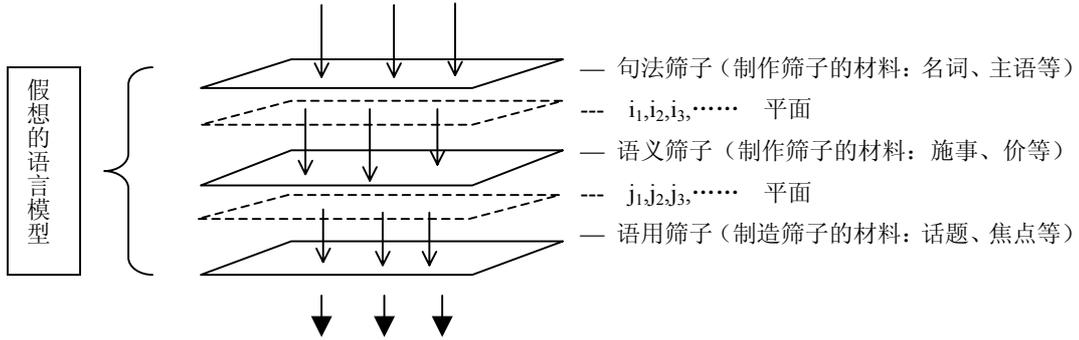
3.1 本文讨论的“语义范畴”含盖的范围包括一般谈到的“配价”、“论元”、“题元”、“论旨角色”、“上下位关系”、“同义反义关系”、“部分-整体关系”、“蕴含关系”、“因果关系”、“致使关系”、……等等诸如此类的概念（“语义范畴”可以看作是这些概念的总称）。对这些似乎跟“语义”沾边的概念的反思在汉语语言学界突出地表现在“配价”上。

3.2 汉语语言学界近 20 年来对“配价”理论情有独钟，特别是在二十世纪九十年代以来取得许多具体的研究成果，这些成果都可以对中文信息处理系统中构建语义知识库提供直接的支持。但是，围绕汉语“配价”理论研究所进行的有关一些比较基础的争议问题的一些讨论，却似乎并没有触及问题的关键，而是一些似是而非的观念的堆积。比如“配价是句法价”、“配价是语义价”、“配价是句法语义价”等等先后出现过的论点，就属于这种性质的认识。人们在热衷于争论“配价是个句法范畴、语义范畴、还是个句法语义范畴”的时候，为什么不反思一下，如果“配价”被称为是“句法语义范畴”的话，“名词”、“补语”等等概念难道不可以被称作是一个“句法语义范畴”吗？

3.3 所谓“句法”、“语义”，到底是什么“意思”？我们可以用下面这个我们称之为“语言学的筛子模型”的图示来描述这些概念的内涵。通过这个模型，我们可以知道语言学家面对纷繁复杂的“语言”对象时所应该扮演的角色。

图示：语言学的筛子模型<sup>5</sup>

可观察的语言/言语 1 { S1 : 所有言语成分任意组合的集合 }



可观察的语言/言语 2 { S2 : 正确的言语表达式的集合 }

在这个模型下，人们用自己“构造”出来的“句法”、“语义”、“语用”等筛子去筛“语言符号可能存在的任意组合”（集合 S1），筛过之后，人们希望得到的结果是所有的“正确的语言表达式的集合”（集合 S2）。

人们用“名词、主语、……”等“材料”（范畴）去编制“句法”筛子；用“施事、受事、……”等材料（范畴）去编制“语义”筛子。当集合 S1 中的符号表达式试图通过“句法”筛子时，像“一张木头椅子”这样的表达式就能透过筛眼儿，而像“一椅子木头张”就透不过去。不仅如此，“句法”筛子一般还要想办法让“一张木头椅子”跟“两名执法人员”从同一个筛眼儿透过去。而“语义”筛子的网眼儿一般相对“句法”筛子的网眼儿要小一些，它的任务是将“一张木头椅子”跟“两名执法人员”分开，并且将“一张木头椅子”跟“一张用木头做的椅子”这两个表达式以一定的方式关联起来，将“两名执法人员”跟“两个从事执法工作的人”这两个表达式以一定的方式关联起来。

为了比较方便和有效地干上面这些事情，人们选择的做法是做好几个筛子来完成任务。现在人们通常做“句法”、“语义”、“语用”三个筛子。不过只要你愿意，可以再多做几个筛子，并给他们其名字叫作“i”筛子，“j”筛子，“k”筛子，……（如果觉得这些名字不太好听，可以管它们叫“句法语义范畴”、“语义语用范畴”什么的）。如果有个“超人”觉得做这么多小筛子太麻烦，他用一个大筛子也能干同样的活，谁会反对他只用一个大筛子（比如就叫“语言平面”）干活呢？

3.4 以上述筛子模型来看所谓“句法范畴”和“语义范畴”，很容易形成下面这样的观念：

（1）把“配价”归到“句法范畴”还是“语义范畴”，并不是问题的关键，因为语言学家的任务是把筛子尽量做得好用一些，管用一些。要达到这个目标，主要是去思考如何编排那些做筛子的材料（各种范畴），至于把这些材料叫什么，并不重要。（2）一个人可以选择把句法筛子做得复杂一些，把语义筛子做的简单一些，也可以相反（这跟具体的语言相关），甚至可以选择不去区分什么句法筛子和语义筛子，只做一个筛子来完成任务（当然经验告诉我们“普通人”这样做不可取）。（3）不应该孤立地评价一个筛子中某一个网眼儿的好坏，评价只能从模型整体的设计着眼。因此，孤立地争论“买”是二价，或是三价、四价都没有太大的意义。因为我们的最终目标并不真的关心“买”到底是几价，而是关心“买群众”是否可被说汉语的人接受；关心“买冬装的钱”跟“100块钱买一件冬装”之间可能存在的形式变换关系；关心“张三买了两条裙子”跟“张三至少拥有两条裙子”之间可能存在的逻辑推导关系，等等诸如此类的问题。两个人可以设计出差别很大的筛子，但都可能筛出人们想要的东西。另一方面，有人设计出一个筛子是为了干一件事情 A，而你不应该因为他设计的筛子干不了事情 B 而指责他。

3.5 要设计出管用、好用的筛子，就要求我们去关注那些真正应该关心的问题。对于做一个“语义”筛子而言，真正的问题可以大致概括为，如何构造一个好的语义范畴体系，去

消解透过句法筛子后剩下的那些语言符号组合的歧义。为了回答这个问题，恐怕主要是做两件事情，一是以“V+N”（述宾结构）、“A+N”（定中结构）、“N+N”（联合/定中结构）、“V+A”（述补结构）等等这样的基本结构作为分析对象，提出有效的语义范畴来对这些基本结构进行细分。细分之后，我们有可能说明“警察叔叔”可以接受，而“外行亲戚”不能接受，以及说明“高个子律师”跟“医生律师”之间的差别，等等；二是以像“VAN”（踢碎热水瓶/踢新球）、“ANN”（大眼睛姑娘/大钢铁公司）、“VN的N”（批评张三的老师）、“NV的N”（汽车抛锚的原因/学校留级的学生）这样的基本歧义格式（参见文献[19]）作为分析对象，提出有效的语义范畴来消解这些格式的歧义。

对上述问题，语言学家解决得越多，能为中文信息处理提供的真正支持就越大。

附录：

表一：几个语义分类体系对名物类词语的处理比较

分类体系		905 语义工程（北京语言文化大学、清华大学等）	北大计算语言所 - 中科院计算所	知网（HowNet）
名物类 词语分 类树	节点数	152	39	137
	叶节点数	110	30	103
	最大深度	9	6	8
抽象事 物类名 词分类 树	节点数	45	11	38
	叶节点数	37	10	28
	最大深度	4	2	5
	示例	<pre>           抽象物          /     \         对象 事理 原因   结果 方法   伦理 ...                        意识 ...                        信息 ...           </pre>	<pre>           抽象事物          /     \         事情 事理   外观 性格   意识 信息                        ...           </pre>	<pre>           精神          /     \         信息 语言 语文—新闻   符号                        音乐                        情感                        事件          /     \         事情 事务   问题                        现象 ...           </pre>

表二：动作行为类词语的语义信息描述体系比较

语义体系		知网（HowNet）	北大计算语言所 - 中科院计算所	鲁川体系
动词语 义分类 树	节点数	67	14	156
	叶节点数	55	12	122
	最大深度	5	3	4
	示例	静态->状态->精神状态->感知状态->有知 行动->变状态->变本体->恢复	运动->行动->搬移	谓词->转移->搬移->置放
动词论 元角色 类型	节点数	53	4	35
	叶节点数	52	3	26
	最大深度	2	2	3
	示例	动态角色->合作施事	论元->主体	周边角色->主体->施事

需要说明的是，上述语义知识体系都处在动态更新过程中，表中数字仅供比较参考之用。有关各语义知识体系的更详细情况可参考文献[3, 4, 6, 7, 8, 18]。

#### 附注：

<sup>1</sup> 这里是用自然语言的表达方式来说明这条规则，在计算机使用时，为了操作方便，通常是以一定的符号化（形式化）的手段来表达的，有关说明可参见文献[19]。

<sup>2</sup> 所谓语言形式之间的“变换”，可以通过基于特征结构（Feature Structure）的集合运算关系来进行形式化的定义，限于篇幅，本文不展开讨论。

<sup>3</sup> 比如国外学者提出的“框架语义学”和国内学者完成的“知网”语义工程实际上都已经是篇章级构造语义知识模型的（参见文献[7,22,23]），或上网到 <http://icl.pku.edu.cn/doubtfire/translations/framesem.htm> 和 <http://www.icsi.berkeley.edu/~framenet/> 查询。

<sup>4</sup> 鲁川等（2000）也明确指出，“谓词配价必须落实到句法平面上”，详见参考文献[3]。

<sup>5</sup> 这个模型是在和北京大学计算语言学研究所于江生博士讨论的过程中形成的。在此谨表谢意。

#### 参考文献：

- [1] 白硕（1996）《语言研究中的实用主义》，载罗振声 袁毓林主编《计算机时代的汉语和汉字研究》，第46-50页，清华大学出版社。
- [2] 白硕（1998）《从信息技术看语言研究》，载《语言文字应用》1998年第1期。
- [3] 鲁川等（2000）《交易类四价动词及汉语谓词配价的分类系统》，载《汉语学习》2000年第6期。
- [4] 鲁川等（2000）《现代汉语基本句模》，载《世界汉语教学》2000年第4期。
- [5] 鲁川（1998）《汉语的意合网络》，载《语言文字应用》1998年第2期。
- [6] 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。
- [7] 董振东 董强（1998）How-net 网站上关于 HowNet 的文献介绍。网址：<http://www.keenage.com>
- [8] 陈小荷（1998）《一个面向工程的语义分析体系》，载《语言文字应用》1998年第2期。
- [9] 吴蔚天（1999）《汉语计算语义学》，电子工业出版社。
- [10] 沈阳 郑定欧（1995）编《现代汉语配价语法研究》，北京大学出版社。
- [11] 袁毓林 郭锐（1998）编《现代汉语配价语法研究》（二），北京大学出版社。
- [12] 沈阳（2000）编《配价理论与汉语语法研究》，语文出版社。
- [13] 吴继光（1999）《动作方式和动作凭借》，载邢福义编《汉语法特点面面观》北京语言文化大学出版社。
- [14] 谭景春（1995）《工具宾语和材料宾语》，载《汉语学习》1995年第6期。
- [15] 徐烈炯 沈阳（1998）《题元理论与汉语配价问题》，载《当代语言学》1998年第3期。
- [16] 沈家煊（2000）《句式和配价》，载《中国语文》，2000年第4期。
- [17] 詹卫东（1997）《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）。
- [18] 詹卫东（2000）《基于配价的汉语语义词典》，载《语言文字应用》2000年第1期。
- [19] 詹卫东（2000）《面向中文信息处理的现代汉语短语结构规则研究》，清华大学出版社，广西科学技术出版社2000年版。
- [20] R.M.W.Dixon (1991) A New Approach to English Grammar, on Semantics Principles, Oxford, Clarendon Press.
- [21] Miller, George A.(1997) ed. WordNet: An Electronic Lexical Database, MIT Press.
- [22] Fillmore,C.J. (1982) Frame semantics, In Linguistics in the morning calm, The Linguistic Society of Korea ed. Hanshin Publishing Co. Seoul, PP111-137.
- [23] Bake, C.F., C.J. Fillmore and John B. Lowe (1998) The Berkeley FrameNet Project. In Proceedings of COLING'98, PP86-90.