

文章编号: 1003-0077(2009)05-0009-10

形容词与名词的语义组合模型研究

赵春利, 石定栩

(香港理工大学 中文及双语学系, 香港)

摘要: 该文首先针对传统方法研究形名组合的不足, 提出了理解形名组合的基本语义模式, 即事物、属性值和属性域; 其次, 根据形名组合的理解模式和语料库的调查, 从哲学理论和语言事实角度, 把名词各自分成了主体、事体、物体、时空、逻辑五个次类, 把形容词分成了主体、事体、物体、时空和评价五个次类; 最后, 借助于计算语言学的研究思想和语义语法的理论原则, 构建了形容词次类与名词次类间语义匹配的形名语义组合模型。研究结果表明该形名语义组合模型能深入细致地揭示形容词与名词的组合规律。

关键词: 计算机应用; 中文信息处理; 形名组合; 属性域; 语义组合模型; 语义语法
中图分类号: TP391 **文献标识码:** A

A Semantic Construction Model between Adjectives and Nouns in Chinese

ZHAO Chunli, SHI Dingxu

(Department of Chinese & Bilingual Studies, the Hong Kong Polytechnic University, Hong Kong, China)

Abstract: Based on the analyses of the disadvantages of the traditional probing of the Adjective-Noun constructions, this paper first puts forward a basic semantic model of "thing, attribute value and attribute domain" for understanding this construction. Then, according to the basic semantic model and the investigation of the corpus, it suggests to classify the noun into 5 sub-groups of subjects, events, objects, space & time, and logical nouns, and the adjectives into 5 sub-groups of subjective, eventual, objective, spatial & temporal, and evaluative. Finally, the semantic construction model between sub-classes of the noun and the adjective is established in light of the computational linguistic scheme and the theoretical principles of the semantic grammar. The research result proves this model to be effective in revealing the rules between nouns and adjectives construction.

Key words: computer application; Chinese information processing; adjective-noun constructions; attribute domain; the semantic construction model; semantic grammar

1 理解形名组合的基本语义模式: 事物、属性域、属性值

汉语理论语言学界对形名组合的传统观点主要有三点: 一是从逻辑意义上把形名组合内部的句法结构看作是偏正结构; 二是从句法结构上看作是修饰关系, 前边的形容词修饰后边的名词; 三是从功能意义上看作是限定关系, 形容词限定名词, 但对形容词与名词的语义关联方式没有提出相应的观点, 如

表 1 所示。

表 1 形名组合的传统观点

分析视角	例子	
	伤心	孩子
逻辑结构	偏词	正词
句法结构	修饰语	中心语
功能结构	限定语	被限定语
语义结构	?	?

收稿日期: 2008-09-22 定稿日期: 2008-11-27

基金项目: 2004 年度暨南大学博士学位论文创新基金资助(52004009); 香港理工大学博士后研究基金资助(G YX0J)

作者简介: 赵春利(1972 →), 男, 博士, 副教授, 主要研究汉语语法与汉语教学; 石定栩(1949 →), 男, 博士, 教授, 主要研究句法理论。

这一高度概括而过于抽象的结论在过去几十年的母语教学中确实发挥过积极的指导作用,但汉语母语教学面对的是已经熟练掌握汉语的中国学生,他们在解读语句之前已经理解了语句,他们所学到的仅仅是利用一种逻辑工具和一套分析语言的概念来表述词语之间(包括形名组合)的句法关系。进入20世纪80年代以后,随着对外汉语教学和汉语信息处理这两项实践应用型事业的发展,传统观点所提供的语义语法信息过于“稀疏、宽泛、空洞”问题日显突出。就形名组合的研究来看,传统观点主要存在四个方面缺陷。其一,在选择关系上,无法解释哪类形容词与哪类名词能够形成合法的形名组合;其二,在语义关系上,无法解释形名组合内在的语义规则;其三,在研究方法上,单纯基于内省而非语料库的传统方法很难发现解释力较强的组合规律;其四,在研究理论上,唯意义论或唯形式论都很难提取出经得起验证的语法规则与语义信息。

因此,我们一方面要充分利用已有的理论资源和技术手段,尽管现在尚无能力考察所有形容词与所有名词如何自由组合的规律问题,但内省可为我们的研究提供语感意义上的心理事实,而语料库的统计数据则从典型性上提供了具有参照性和证据性的参数和语言事实,逻辑和哲学知识则为我们考察关系、划分类别提供某些指导性的原则。另一方面,我们还应该尽量摆脱上述传统结论的影响,尽量用“用朴素的眼光看汉语^[1]”,重新认识形容词与名词的语义关系。

我们知道:语言的意义来源于语言所反映的对象,名词的意义来自名词所指称的事物,形容词的意义则来自事物的性质和状态。因此,可从两个层面认识形容词与名词的关系,见表2。

表2 语言符号与现实指称关系

语言符号层面	名词	形容词
现实指称层面	事物	性质和状态

性状是反映事物某个领域的属性值,比如,“伤心的孩子”中的性状“伤心”反映了属性主体“孩子”在情感属性域的属性值。因此,事物、性状(属性值)和属性域(性状所指的语义特征)三者之间的关系反映了形容词与名词能否组合的基本条件,也是分析形名组合的基本语义模式。

第一,形容词所指性状对名词所指事物具有依存性。名词所指事物是自足的、独立的,具有直接的

心理现实性,如:看到“桌子、房间、姑娘、事故、战争”等词语,我们马上会将词语与一定的事物联系起来,在理解形容词所指性状的认知结构中,必然含有“关于什么”的心理探询倾向,如:看到“红、香、野蛮、容易、高兴、偶然”等词语,我们会想到是“关于某事物的”“颜色、气味、行为、难度、情感、模态”的性质,所以说,形容词所指性状对名词所指事物具有认知和语义上的依存性,形容词和名词作为语言符号形式结合起来指称外在现实世界的事物。维特根斯坦认为:“形象的要素以一定的方式结合起来,这件事情表明各物也是这样互相结合起来的^[2]”。罗素在该书导论中把它归结为:“不管语言是如何构成的,句子的结构和事实的结构必须有某种共同的东西”,对世界中的事物来说,性质和状态对事物具有依附性,如同形式对内容的依存关系一样,那么,在符号层面,形容词和名词可以组成两种基本的结构形式:AN和NA。可以说,形容词所指称的性状是一定名词所指事物的性状,反过来,名词所指事物都具有一定的性状。

第二,一个形容词所指性状只表示一种属性值,该属性值指向事物的某一方面的属性,这“某一方面”用名词表示出来就是属性域(即名词具有的语义特征),比如:颜色、性状、体积、面积、高度、温度、性格、情感、模态、难度、强度等,属性值与属性域的关系是元素与集合的关系,即:

属性域(F)

= 属性值(A1 + A2 + A3 +An - 1 + An)

比如:形容词“红”这一属性值只指称具有“颜色”属性域的事物的“颜色”属性域,“倔强”只指称具有“性格”属性域的事物的“性格”属性域,“漫长”只指称具有“时间”属性域的事物的“时间”属性域。一般来说,形容词具有单域指向性,即一个形容词作为一个属性值只指向事物的一个属性域,如果把属性域比作一个集合,那么属性值就是该集合的一个元素。一般来说,属性域则是多值性的。

第三,名词所指事物具有不同的属性域(语义特征集),每一属性域又具有不同属性值。即事物具有

有个别形容词可以与多个属性域之间存在附属关系,这样如果名词所指事物同时具有它所附属的两个以上属性域,那么,就会造成歧义,如:“老”可属于“时间”和“年龄”属性域,在“老朋友、老同学、老教师、老干部”就会造成歧解;另外,有的形容词可能指向的不是单一事物的属性,而是通过比较得出的事实间的属性域,比如:“相似”指向两个以上事物的“关系”属性域,“多、少”指向事物集合的“数量”属性域。

多属性域,每个属性域则具有多属性值。如:“苹果”具有“颜色、体积、味道、气味、质地、口感、形状、温度、鲜度、数量、质量、价值”等多种属性域,相应地每个属性域可用不同的属性值来指称:“红/绿、大/小、甜/酸、香/臭、硬/软、脆/面、圆/方、热/冷、新鲜/烂、粗糙/光滑、多/少、好/坏、便宜/贵”等不同的属性值。可以说,名词(N)所指的人事物有一个属性域集合($N_1 + N_2 + N_3 + \dots + N_{x-1} + N_x$)和对属性域单向对应的属性值集合(A),用图1表示如下。

$$N \begin{cases} N_1(A_1 + A_2 + A_3 + \dots + A_{n-1} + A_n) \\ N_2(A_1 + A_2 + A_3 + \dots + A_{n-1} + A_n) \\ N_3(A_1 + A_2 + A_3 + \dots + A_{n-1} + A_n) \\ \dots \\ N_{x-1}(A_1 + A_2 + A_3 + \dots + A_{n-1} + A_n) \\ N_x(A_1 + A_2 + A_3 + \dots + A_{n-1} + A_n) \end{cases}$$

图1 人事物、属性域集合与属性值集合关系

第四,跟属性值与属性域的关系一样,属性域与事物的关系也是元素与集合的关系。但不同的是,属性域对属性值具有一定程度的选择性,比如:属性域“颜色”可以选择“红、黄、蓝、白”等属性值中的一种或几种;而属性域对事物缺少选择性而存在着一种必然性,我们无法想象一个没有属性域“颜色、形状”的视觉事物,一个没有属性域“情感”的主体(人),一个没有“模态”的事件。正如维特根斯坦所言:“视野中的斑点不是必须红色的,但它应该有颜色,它是被所谓颜色空间所包围的。音响应该具有一种高度,触觉的客体应有一种硬度等等”^[3]。基于属性域与事物之间的这种必然性,我们可根据属性域反映的事物不同以及相关属性域的属性值所选择的事物不同,从语义上对事物(名词)进行宏观分类。

第五,同一形容词与不同类别的名词组合时,就会形成不同语义关系的形名组合。值得注意的是,这些不同类别的名词之间存在着一定程度的上下位关系或同位关系。比如:痛苦的孩子、痛苦的表情、痛苦的眼泪、痛苦的消息。

2 基于形容词和名词双向选择的形容词和名词宏观语义分类

2.1 当前词类语义分类存在的问题及其对策

众所周知,理论语言学对词类的研究主要根据意义和功能分布。前者主要基于内省和语感对词语的词汇意义而非句法结构中的组合意义进行聚合,

得出的语义分类有时难以得到形式验证;后者根据句法功能分布能够解决部分验证问题,但得出的规则过于稀疏和笼统,甚至得不到语感上的认可,也不利于计算机进行信息处理。计算语言学对词类的研究主要根据人工语义分类和统计主义,但人工分类体系目前并不完善,很难解决自然语言理解问题,而统计主义是通过词语分布环境的相关计算对词语进行分类,其优点是面对已有的语言事实,考察范围广泛,但得出的往往不是语义分类,当然不利于自然语言的理解和生成。

那么,如何基于词语组合给各种词语进行分类呢?我们认为,词语搭配在性质上是指两个或两个以上词语能够组成一个有意义的短语,并且在语感、内省或者语料事实上一般都能得到认可的组合。无论从理论上还是语言事实上看,词语之间的组合搭配具有一定的规律性和倾向性。因此,要想得到既符合语感又利于信息处理的形容词和名词分类,就应该根据形名组合的语义关系类型^[4],把理论语言学和计算语言学的方法结合起来,把规则主义和统计主义结合起来,把语义和数据结合起来。具体地说,可以从两个方面入手:一方面,从哲学理论上基于内省和语感宏观地给出名词和形容词语义分类,探讨形名组合的可能性,提炼出具有一定演绎性质的规则。另一方面,从语言事实上来验证、修正、补充、修改、细化名词和形容词的语义分类,这是以形名组合的现实性为基础借助数据统计来归纳和验证形名组合的规律,当然,这种统计尽量从形容词看名词和从名词看形容词这一双向验证视角出发。

2.2 从哲学理论上对名词和形容词的语义分类

先看名词的语义分类。名词分类就是对名词所

选择能力的大小和高度的高低会因事物的性质不同而不同,比如:自然物的颜色一般不由人的意志来决定。通常我们说“红太阳、白天鹅”而很少说“绿太阳、蓝天鹅”,自然物颜色的属性值多寡依赖于该事物自然颜色的多寡,比如“花”的颜色值就极为丰富。另外,某一事物的出现频率和典型程度也会影响属性域对属性值的选择能力。其实,从认知和语言事实上说,这反映了属性值与事物的同现频率和关联度。

不同属性域与事物之间的紧密关联的程度会因事物和属性域的不同有所不同。比如:颜色大于温度对“苹果”的关联度,而基于评价和对比的属性域(外貌、价值、隐显等)与视觉事物的关系则不如“颜色、形状”等属性域与事物的关系那么紧密。这在一定程度上反映了属性域的适用范围和组合能力。某一属性域的属性值组合范围越广(语义特征意义上的广,即能够与多种语义特征的名词组合),组合能力越强,该属性域与某一类事物的关系越不紧密,比如:属性域“价值”的属性值“好”就属此类,当然这只是一种倾向,绝非严格意义上的反比关系。

指事物的分类,因为名词的意义来自名词所指的对象事物,尽管说,“意义”不等于“事物”,但离开名词所指的事物恐怕难以讨论和理解名词的“意义”。从定义上看,语法学界只有意义派明确的提出了名词的定义,提出的有:(a)马建忠的“凡实字以名一切事物者,曰名字”^[5];(b)黎锦熙的“名词是事物的名称,用来表示观念中的实体”^[6];(c)王力的“凡事物的名称,或哲学科学上的名称叫做名词”^[7]。朱德熙总结为:“名词表示事物的名称”^[8]，“事物”就成为理解“名词”意义和进行语义分类的核心概念。那么,事物包括什么?马建忠认为:“事物二字,一切毕赅矣。在天之日月星辰,在地之河海华岳,人伦之君臣父子,物之有形者也。怪力乱神,利命与仁,物之无形者也。而所教者文行忠信,所治者德礼政刑,所得者位禄名寿,所艺者礼乐射御书数,皆事也,皆名也。凡目所见,耳所闻,口所嗜,鼻所嗅,四肢之所触,与夫心之所志,意之所感,举凡别声、被色与无声、无臭,苟可以语言称之者,无非事也,无非物也,无非名也”,可归结为:事物是指人之所感的有形的东西,人之所想的无形的东西以及人类做的事情。可细化为图2。



图2 传统语法的事物分类

从“人伦之君臣父子”和“苟可以语言称之者,无非事也,无非物也,无非名也”中可以看出,他把作为有形物体的感知者、无形物体的创造者和事件行为的活动者的“人”看作是“有形”物体了,我们认为这是不妥的。根据《现代汉语词典》的名词定义:“名词是表示人或事物名称的词”^[9],可能更为合理。因为从哲学上说,无论是唯心主义,还是唯物主义,是一元论,还是二元论,基本的逻辑基点都是把具有认识能力和实践能力的主体“人”与作为认识对象和实践对象的客观物体区别开来。基于此,我们认为对名词所指对象进行分类,第一步应该根据能动性智力特征划分为:主体和物体,这是构成世界的两大对立要素;第二步,主体和物体由自身及其相互作用所呈现出来的产生、发展、变化、消亡,我们称之为事体;第三步,时间、空间是主体与物体存在、运动、变化的基本形式,主体、物体及其运动、发展、变化所形成的事体之所以作为现象能够呈现出来并为主体所

感知和识别,必须借助这两大先天直观形式:时间和空间;第四步,具有能动性的主体还会根据自身的理解和价值趋向把主体、物体以及事体的关系连接起来,形成一个可理解并预测的逻辑体系;第五步,主体在感知、认识、改造外在对象的过程中,还能借助内在的自主意识和心理活动,形成一些认识、情感、意志等抽象的思想观念。因此,基于哲学分类给出的名词定义是:名词是指称以时空为基本存在形式、以逻辑为基本认知形式的主体、物体及其运动发展变化的事体的名称。从静态平行机械的角度看,指称这些对象的名词从概念上可分为五类:主体、事体、物体、时空、逻辑,至于思想观念作为主体的产物暂时包含在主体内部。

再看形容词的语义分类。我们知道,形名组合是一种依附性短语,一般来说,确定了名词的语义分类,结合形名组合的语义关系类型,也就在一定程度上确定了形容词的语义分类。因为形容词在语义上反映了名词所指事物的属性,名词在理论可以分为五类,那么,与之相应,形容词也可以分为五类:主体、事体、物体、时空、逻辑。根据形名之间的这种语义对应性,形名之间的组合模型就可以非常容易地建立起来,似乎可以用下列一一对应的图式表示。

形容词 A		名词 N
主体	——	主体
事体	——	事体
物体	——	物体
时空	——	时空
逻辑	——	逻辑

图3 形容词小类与名词小类直观对应表

形态派只是列举了几个标记,功能派和原型派都是从鉴别名词的标准角度解释了名词的分布和组合能力,不是对名词内涵的界定,而是对符合名词条件的功能限定,当然不属于严格意义上的定义。

在这里我们无意争论历史上的时空学说的优劣,无论是绝对时空观(牛顿),还是相对时空观(爱因斯坦),本文主要从三个角度理解时空的意义:一是从形成主体物体事体的表象的意义上,我们认为康德的时空直观说具有深刻的道理,即空间是外感官的直观形式,时间是内感官的直观形式,通过外感官感到的一切都在空间中,通过内感官感到的一切都在时间里,空间和时间是两种先天的纯粹直观形式;二是从日常经验的角度看,以莱布尼茨为代表的理性派的时空关系说也很有见地,即空间代表事物的并存关系,时间代表事物的连续关系;三是从时空与存在物的关联性上看,将黑格尔的辩证法与费尔巴哈的唯物观结合起来的马克思主义时空说具有很强的解释力,即一切存在的基本形式是空间和时间。限于篇幅我们只说出我们所认可的时空学说的合理之处,不讨论认可的理由和依据。

当然,这是一种逻辑推理,形名之间的语义组合模型决非如此简单、整齐,我们还必须从语言事实出发,基于语料库的调查来分析、补充、验证这一组合模型的可靠性和精确性。

2.3 从语言事实上对名词和形容词的语义分类

要想借助语料库的统计调查得到基于形名组合的形名分类,必须坚持语义语法^[10]的双向性原则。上述的哲学分类偏重从形容词对名词的依附性角度划分形容词的聚合类别,这是宏观分类的一个必要步骤;但是语义语法的方法论要求,应该从名词选择形容词和形容词选择名词的双向角度进行聚合分类。具体地说,就是以名词为核心,以形容词为名词的组合分布环境,可以验证并得出名词的聚合类别;以形容词为核心,以名词为形容词的组合分布环境,可以得到形容词的聚类。把这两个过程结合起来,就会形成基于形名组合的名形分类。

语义语法的这一双向思想与计算语言学界的双向分级聚类算法不谋而合。闻扬等人^[11]认为:“已经进行的工作大都从某一词性的词的分布环境出发,得到该词性词语的聚类结果,但事实上,不同词性的词语之间的聚类是有一定联系的,不能割裂开来。以汉语中的形容词——名词搭配组合即 A-N 结构为例,把形容词作为实体,名词作为其分布的特征,可以得到形容词的聚类,同样,以名词为实体,形容词为特征,可以得到名词的聚类。这两者的聚类过程是互相作用的。如果把两者的聚类过程交替进行,将有可能提高聚类的效果。”计算语言学的方法过于重视数据的精确性和语言事实的同现率,而忽略了语言是以语义为基础的,语言组合具有生成性。因此,计算语言学所得出的形名分类就会出现与语义和语感相左的现象。闻扬等人得出了 397 类形容词和 610 类名词,不仅分类过于庞杂,而且得出的类别语感很不一致。比如:就形容词看,把“红润”与“狠心”、“迫切”与“清醒”、“内疚”与“花哨”归为一类,我们认为是不妥的;就名词看,把“老板娘”与“寓言”、“盗匪”与“年头”、“敌军”与“心态”归为一类也得不到语感的支持。可见,脱离语义的单纯统计主义做法并不能得出有利于语言理解和生成的形名语义类别和组合模型。

我们先以名词为核心,以形容词为名词的分布环境,验证名词的哲学分类,然后以形容词为核心,以名词为形容词的参照对象,提炼形容词的聚合类别。

根据对 CCL 语料库的调查,我们发现,从哲学逻辑上推理得出的名词分类与从语言事实上观察到的形容词对名词的选择和聚合具有惊人的一致性。基于哲学意义上的名词定义所得出的主体、事体、物体、时空、逻辑这五种概念分类能够得到语言事实的证明。也就是说,从属性值及其所指向的属性域角度看,表示一定属性值的形容词也会聚合在具有相应属性域特征(语义特征)的名词周围,形成以名词的语义分类为核心的形名组合,从而得出相应的主体、事体、物体、时空和评价五种形容词类别,从而使相应的名词表现出各自独特的属性域和属性值。籍此,我们可以验证上述名词和形容词的语义分类是科学的。

第一类【主体名词】和【主体形容词】

有一类名词具有多种人类特有的属性域,比如:【容貌】【体态】【体质】【体感】【态度】【性格】【智力】【情感】【气质】【意志】【品质】【经济】等,相应地,语义指向该类属性域的形容词就会聚合在具有该种属性域的名词前后,具有这些属性域并能受到相应属性值修饰的名词就是主体名词。该类名词很多,如:

阿姨、爱人、伴侣、帮手、帮凶、编辑、标兵、大众、代表、党羽、党员、导演、敌人、对象、队伍、恶霸、儿女、翻译、俘虏、冠军、汉奸、后代、傀儡、懒汉、邻居、领导、领袖、流氓、旅客、民众、民族、能手、农民、奴隶、扒手、叛徒、朋友、青年、妻子、囚犯、群众、人类、人员、人质、伤员、上级、上司、少年、少女、诗人、师傅、师长、市侩、首脑、首长、天才、同伴、同胞、同行、同谋、同事、同志、统帅、外宾、外敌、外行、外寇、委员、文豪、学生、英雄、勇士、战士、战友、政客、职工、指挥、志士、主编、主管、主将、主角、主力、主人、主席、助手、专家、壮士、祖先、罪犯、罪人、作家、对手、高手、新手、老手等。

表示一定属性值的形容词就会从生理、精神、行为等角度聚集在具有【容貌】【体态】【体质】【体感】【态度】【性格】【智力】【情感】【气质】【意志】【品质】【经济】等属性域特征的【述人】或者相关语义特征的主体名词“老师、学生、姑娘、大夫、朋友、家长”等前后,从而使主体在生理、精神、行为等方面体现出自身独有的属性域及其对应的属性值,这就是主体形容词,如:

【容貌】:俊俏、俊美、俊丽、俊秀、俊逸、英俊、清

本文使用的是北京大学汉语语言学研究中心的 CCL 现代汉语网络版语料库,在此表示感谢。

癯、清秀、秀气、秀美等。

【体态】：高、矮、胖、瘦、魁梧、魁伟、剽悍、窈窕、瘦小、矮小、娇小等。

【体质】：健康、单薄、羸弱、柔弱、纤弱、瘦弱、孱弱、嫩弱、壮健、健壮等。

【体感】：疲惫、疲倦、疲乏、疲劳、疲困、困乏、困倦、劳累、麻木等。

【态度】：热情、冷淡、冷漠、诚恳、忠厚、谦虚、骄傲、认真、仔细、虔诚等。

【性格】：倔强、开朗、温顺、泼辣、活泼、内向、外向、爽朗、爽直等。

【智力】：聪明、愚蠢、精明、迟钝、老练、狡猾、狡诈、奸诈、机警等。

【情感】：伤心、高兴、激动、紧张、愉快、悲哀、兴奋、幸福、苦闷、苦恼、烦恼等。

【气质】：潇洒、洒脱、文雅、儒雅、腼腆、和蔼、不凡、飘逸等。

【意志】：冷静、沉着、果断、独断、坚强、顽固、执拗、勇敢、勤奋、懦弱、冒失等。

【品质】：高尚、卑鄙、残忍、正直、公正、礼貌、忠诚等。

【经济】：富、穷、富裕、富有、贫穷等。

第二类【事体名词】和【事体形容词】

有些名词或者名动词具有一定的事体属性域，比如：【模态】【难度】【急度】【速度】【强度】【范围】【深度】【细度】【便度】【灵度】【力度】【动貌】等，常见的事体名词主要有两类：一类是单纯事体名词，如：战争、动作、事情、事件、事儿、说话、办事、思想、认识、记忆、生活、风、雨、霜、雪、地震、海啸等；一类是名动词，可称为句法事体名词^[4]，如：旅行、讨论、工作、学习、争辩、战斗、设计、训练、调查、分析、研究、服务、照顾等。

一定属性的形容词就会从各种句法结构中聚集在这些具有【模态】【难度】【急度】【速度】【强度】【范围】【深度】【细度】【便度】【灵度】【力度】【动貌】等属性域特征的【述事】事体名词“事情、动作、战争、交通、照顾、认识、思想、研究、调查、分析”等周围，使事体名词从发生、过程、状态、性质等角度呈现出自身独有的属性域和属性值，比如：

【模态】：可能、必然、突然、偶然、意外等。

【难度】：难、困难、艰难、棘手、麻烦、容易、艰苦、艰辛、艰巨等。

【急度】：急、危急、紧急、紧迫、急迫、迫切、要紧等。

【速度】：快、慢、迅速、迅捷、快捷、缓慢、迟缓、舒缓、及时等。

【强度】：强烈、猛烈、剧烈、激烈、热烈、严重、轻微等。

【范围】：广泛、普遍、全面、宽泛、片面、狭隘、偏狭等。

【深度】：深刻、深入、深切、深广、粗浅、浅显、浅陋、肤浅、浮浅、短浅等。

【细度】：详细、详尽、详密、详明、详实、翔实、周详、粗略、简要、简略、笼统等。

【便度】：方便、便利、顺利、通畅、顺畅等。

【灵度】：灵巧、灵活、灵敏、敏捷、轻捷、熟练、巧妙、笨拙等。

【力度】：勤奋、努力、刻苦、勤勉等。

【动貌】：匆忙、仓促、从容、悠闲、慌张、慌乱、忙碌、轻盈等。

第三类【物体名词】和【物体形容词】

有一类指向物体的名词具有自身特有的属性域，像【颜色】【形状】【味道】【维度】【粗度】【硬度】【温度】【湿度】【密度】【浓度】【鲜度】【亮度】【清度】【气味】【表面】等，当然并不是每一个物体名词都具有上述属性域及其属性值，而是具有部分上述属性值，比如：“桌子”具有“颜色、形状、维度、硬度、表面等”属性域，但是没有“粗度、清度、浓度等”属性域，常见的物体主要呈现为固体、液体、气体、光线等形态，如：

固体类：石头、木头、椅子、书架、电脑、刀子、汽车、蔬菜、水果、食品、服装、道路、土地、山川、花草等。

液体类：汽水、啤酒、饮料、河水、湖泊、药水、雨水、泪水、粥、汤等。

气体类：空气、热气、水蒸气、煤气等。

光线类：灯光、阳光、火光、月光等。

表示一定属性值的形容词就会在句法结构中集结在这些具有【颜色】【形状】【味道】【维度】【粗度】【硬度】【温度】【湿度】【密度】【浓度】【鲜度】【亮度】【清度】【气味】【表面】【边端】【价格】【土质】等属性域特征的可感【述物】语义特征的名词周围，如“桌子、石头、刀子、食品、商品、水、空气、光线”等，从而使述物名词从固体、液体、气体等角度表现出具有自身独特的属性域及其属性值，如：

【颜色】：红、黄、黑、白、绿、蓝等。

【形状】：圆、瘪、弯、直等。

【味道】：酸、甜、苦、辣、咸、涩、淡等。

【维度】：大、小、长、短、高、低、宽、窄、厚、薄、

深、浅等。

【粗度】：粗、细等。

【硬度】：硬、软、坚硬、酥软、柔软等。

【温度】：热、冷、凉、炎热、燥热、闷热、寒冷、凉爽、温暖等。

【湿度】：干、湿、潮、干燥、潮湿、湿润等。

【密度】：密、稀、稠密、稀薄等。

【浓度】：稠、浓、稀、浓烈等。

【鲜度】：新鲜、鲜等。

【亮度】：昏暗、光亮、明亮、黯淡、暗、亮。

【清度】：清、透明、清澈、浑浊、明澈等。

【气味】：香、臭、腥、臊、芬芳、清香、芳香、醇香、难闻、好闻、馥郁等。

【表面】：平、粗糙、细腻、光滑、平滑、平坦、平整、崎岖等。

【边端】：尖、尖锐、锋利、快、钝等。

【价格】：贵、贵重、昂贵、便宜、廉价、优惠等。

【土质】：肥沃、贫瘠、富饶等。

第四类【时空名词】和【时空形容词】

时空名词具有一定的时空特性,如:时刻、事件、上午、星期天、六月、环境、形势、局面、场面、广场、市场、山村、天空、草原等。

一定性质和数量的形容词分别从【时间性】和【空间性】角度聚集在具有【时长】和【环境】【情境】【局势】【距离】【空间】等属性域特征的【时间】名词和【空间】处所名词周围。在语义上,由于事体与时间之间的天然关系,事体名词周围也常会聚集具有【时长】特征的名词,比如:漫长的旅行、长久的记忆、短暂的战争等(我们会把【时长】属性看作事体名词的属性域之一进行研究);除此之外,时间名词【早晨、时间等】和处所名词【环境、形势、地方】都会表现出独特的属性域并吸引相应的属性值形容词,比如:

【时长】：新、旧、早、晚、长、久、短、漫长、长久、长远、短暂、短促等。

【环境】：安静、静谧、平静、背静、清静、僻静、喧嚣、热闹、嘈杂、吵闹、拥挤等。

【情境】：繁荣、火爆、红火、荒凉、凄凉、凄惨等。

【局势】：稳定、动荡、混乱等。

【距离】：远、近、偏远、遥远、偏僻等。

【空间】：广阔、宽敞、宽大、宽阔、空旷、狭窄、窄小、狭长等。

第五类【逻辑名词】和【评价形容词】

逻辑术语,如:原因、前提、条件、证据、好处、坏处、程度、特征等,属于对处于一定时空处所中的主

体、事体、物体之间的关系进行表述的元语言,因此,附属于一定属性域的属性值形容词很少会选择缺乏独特属性域的逻辑名词作为中心语。基于语料事实的调查,我们发现,即使属于一定属性域的属性值形容词选择逻辑术语名词,该形容词所反映的属性域也不是逻辑名词固有的,如:高兴的原因,紧张的程度等,A属于【述人】名词的【情感】属性域。此外,一些基于对比和评价的形容词或者语义指向关系或数量的形容词会选择语义特征不一的名词,此时,它们也会选择逻辑名词,比如:重要的原因、突出的结果、相似的条件、充足的理由、确凿的证据等,但总的说来,逻辑名词并没有自身独有的属性域和属性值。

与没有自身固有属性域和属性值的逻辑名词一致,语料库的调查显示:有的形容词并非仅仅反映某一类名词的单一属性,而是多种名词的共同属性,我们称之为评价形容词,如:“一般、特殊、相似、神秘、重要”等,它们与主体、事体、物体、时空、逻辑五类名词都可以形成组合,见图4。

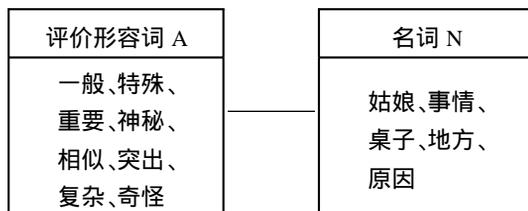


图4 评价形容词与名词组合例证表

这样的话,形容词所反映的属性是某一类名词的,还是多类名词的,将会影响到形容词与名词组合能力。从名词角度说,有的属性是某一类名词本身固有的,也是该类名词与别的名词区别开来的标记,而有的属性则是多种名词共有的,反映了多种名词的共同属性。因此,名词的语义分类与形容词的分类还存在着不一致的地方。也就是说,有些形容词的组合能力很强,能够跟多种语义特征的名词组合,这类形容词一般都是从某一视角或标准出发做出评价的评价形容词。由于评价标准不同,评价形容词也包括很多类别,常见的如:

【价值】：好、不错、出色、优秀、优异、精彩、坏、差、不好、恶劣、低劣、拙劣等。

【怪异】：特殊、奇怪、古怪、奇特、怪异、怪诞、离奇、稀奇、瑰奇、诡异等。

【审美】：靓、丑陋、潇洒、优美、美丽、优雅等。

【常态】：一般、常见、平常、普通等。

【隐显】：隐秘、显著、突出、鲜明、显眼、醒目等。

【名位】：有名、著名、关键、重要等。

【繁简】：繁琐、复杂、简单等。

【标准】：合格、合理、合法、恰当、合适、贴切等。

【异同】：相似、类似、近似、相同、相近、相仿、不同、一样、一致、相反等。

【真假】：真、假、真实、逼真、虚假等。

总之,从宏观上看,表述一定属性值的形容词对具有一定属性域名词语义类别的选择表现出强烈的一致性。也就是说,形容词在具有一定语义特征的名词周围集结时,表现出一定的倾向性,这种倾向性与名词的逻辑语义分类是一致性,这说明名词的哲学分类与通过一定属性值的形容词对一定语义特征的名词的选择聚合所体现出来的分类是相同的。因此,我们说,从名词内涵的界定以及自身是否具有独有的属性域和属性值方面,名词可以宏观地分为主体、事体、物体、时空、逻辑五大类。但逻辑名词不具有自身独有的属性,与逻辑名词同现的形容词一般也可以与别的名词类别同现,因此,形容词类别中就没有独立的逻辑形容词,而是评价形容词。这样我们可以得到形容词的五种类别:主体、事体、物体、时空、评价。

3 静态形名组合的语义组合模型

3.1 计算语言学界对形名组合模型的设想

无论是国内还是国外,理论语言学界主要研究倾向是在词类范围内独立地研究名词的分类和形容词的分类,很少从宏观组合角度就形容词和名词两大实词整体的静态语义组合模型进行研究,也没有就两大词类间的整体组合规律和搭配倾向进行探讨。

有意思的是国外和国内的计算语言学界出于语言理解和生成的需要,都对形名组合模型或相关问题进行过研究。墨西哥国家工学院计算研究中心(Center for Computing Research of the National Polytechnic Institute)的 Igor A. Bolshakov 和 Alexander Gelbukh^[12]曾以名词为核心研究其修饰语的语义分类。文章有三个亮点:首先是就名形间的语义关系和语义分类提出了两点认识,即被修饰的名词能够显示形容词修饰语的语义分类以及名词与形容词在语义分类具有相关性;其次,他们还提出了确定核心名词两条标准:使用频率高和可被多种修饰语修饰;第三,在研究方法和语料搜集上,以核心名词“人”为例,基于语料库来搜索“人”前修饰语集

合,并对其进行语义分类,这一操作性很强的方法对我们搜索形名搭配语料启发很大。其实,从研究目的上说,这篇论文并不是以建立形名组合模型为目的,而是着眼于如何以名词为核心研究修饰语的语义分类,所以它并没有提出一个系统的名词语义分类以及形容词的语义分类,也没有提出名词和形容词分类的基本原则,但对如何建立形名语义组合模型意义重大。

真正以建立形名组合模型为研究目的的学者是姬东鸿、黄昌宁^[13],他们首先按照形名组合关系建立一个分级的树状名词属性体系,该体系分为三级,第一级 12 种,其中每一种再分为二级若干,同样,二级再分为三级若干。其次,根据意义分别对名词和形容词分类,每一类按照同义词、近义词、反义词组成不同的集合。第三,建立名词属性框架,同一集合的名词具有相同的属性框架,同一集合的形容词与同一集合的名词组合时,往往给相同的名词属性赋予属性值。第四,建立基于名词属性框架的缺省继承机制、多重继承机制和基于属性激活理论的冲突消解机制。应该说,这一计算语言学研究视角从理论上对我们研究形名语义关系类型和静态组合模型都具有重要的指导意义。他们提出的基于名词属性框架的属性、名词、形容词三分法使我们建立了基本的形名语义关系类型:固有属性值和属性承载体、固有属性值和固有属性域,基于上下位原则的名词内部分类和基于意义及其与名词组合能力的形容词内部分类为我们从同位原则出发给名词形容词分类也产生了深远的影响。

但是应该看到,由于他们没有考虑到句法语义的影响,因此还存在很多缺陷。比如:他们建立的名词属性体系缺乏哲学和语言事实的支持,没有给出验证、区分并识别物名、物貌、物理、物源、物果、物用、物境、物值、物为、德才、性情、仪态这 12 种一级属性的句法语义依据(像“物貌—美丽和物境—干净”的差异),有的属性不是所有的名词都具有的(像“德才、性情、仪态”属于主体属性),如果说这些都是细枝末节的话,那么,对一些重要的属性的忽略却是不可思议的(像“时间—长远、短暂”、“距离—远、近、偏僻、偏远”、“模态—偶然、突然、可能”等)。

他们虽然建立了形名组合模型的三个基本要

【时间】【距离】并非动词独有的属性,很多名词也具有这两类属性,特别是在形名组合层面上,如短暂的生命、长久的岁月、漫长的历史、遥远的山区、偏远的农村等等。

素：形容词(属性值)、属性(属性域)、名词(载体),但借助以属性体系为基础的名词属性框架而把形名组合的语义关系简单化了。因为形名之间的语义关系深刻地反映了形名组合的论元角色的性质,比如:“聪明的先生”与“聪明的办法”其语义关系是不一样的,其缺点就在于过于依赖于缺乏句法语义基础的属性体系和名词属性框架,过于强调了形容词对名词的依附性,没有看到形容词对名词的独立选择能力,缺乏双向选择意识。正如张瑞霞、张蕾所做的“采用以语义为主、语法为辅的策略”^[14]才能有效地分析短语结构。

3.2 基于语义语法理论的静态形名组合模型

语义语法理论不仅有很深的哲学理论作为基础,而且还有较为完善的本体论、方法论,它不仅为研究形名组合语义关系类型提供理论指导,而且也作为从哲学、语言事实角度对名词与形容词的次范畴进行分类研究起着理论指导意义,由此我们提出了十种主要语义关系类型,以及基于形名组合规律研究的五种名词语义次类和五种形容词语义次类。

形容词与名词的基本语义关系是固有属性值与属性承载体的关系,名词是指称一定时空中的主体、事体和物体,形容词则反映处于时空中的人、事、物的性状,而形容词与名词语义组合的基础主要是名词具有的属性域与形容词属性值所属的属性域的一致性和关联性,因此,无论是形容词选择名词还是名词选择形容词,以属性域为基点,对形名组合语义关系的划分,既能从语义上细化对名词内部的分类,还能深入细致地揭示形容词的语义分类。我们从宏观总体上把名词划分为:主体名词、事体名词、物体名词、时空名词、逻辑名词;把形容词化分为:主体形容词、事体形容词、物体形容词、时空形容词、评价形容词。这样形名次类间的语义组合模型可以简单地通过下列组合表格表示出来,如:

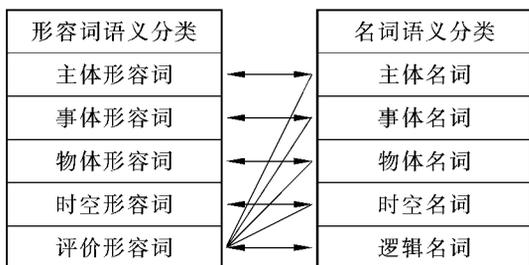


图5 形容词小类与名词小类搭配的语义组合模型

尽管说上述组合模型非常宏观,不甚细致,但它

确实能够从组合规律角度把形名搭配的倾向性反映出来,所划分的五种形名组合语义模块,为从形名语义关系角度进一步细化形容词和名词的语义分类奠定了基础。

还有几点值得说明的:

第一,这一形名组合模型是基于哲学和语义语法理论得出的,它反映出形名语义组合的宏观倾向,即主体形容词与主体名词组合、事体形容词与事体名词组合、物体形容词与物体名词组合、时空形容词与时空名词组合、评价形容词与五种名词都存在搭配可能性,但是不是说主体形容词只与主体名词组合,而是说主体形容词主要指向主体名词的属性域,因为主体形容词有可能借助主体属性域与主体名词次范畴发生关系,比如:主体形容词“激动”可以指向主体名词的属性域“情感”,组成“激动的姑娘”,还可以借助属性域“情感”,与主体名词次范畴(即下位概念)组合成“激动的表情”,这是从纵深方向对主体名词的研究,其他名词可依此类推。

第二,组合模型中评价形容词所指的评价属性值主要是根据标准视角和关系视角而不是主观视角和客观视角得出的。也就是说,评价者是根据某种标准或关系而不是评价对象或评价主体来进行评价的,这样,评价属性值既不指向评价对象,也不指向评价主体。从组合上看,评价形容词就可以与主体、事体、物体、时空和逻辑名词组合搭配,具有广泛的组合性,比如:反映人、事、物“怪异”属性域的属性值“奇怪、怪异、奇特、特别”形容词可以对各类名词进行评价:奇怪的女人/战争/箱子/晚上/地方/原因。而语义指向“异同”的评价形容词“相似”,同样可以对各种事物之间关系通过比较进行评价:相似的人物/遭遇/杯子/早晨/地方/特征。

第三,这一组合模型将名词分类、形容词分类与属性域分类紧密的结合起来,除了评价形容词之外,基本上体现了一类形容词、一个属性域、一类名词之间三位一体的对应关系,比如:形容词“俊俏、英俊、清秀”、属性域“容貌、相貌、脸庞”、名词“姑娘、小伙子、作家”就存在对应关系,可以形成固有属性值与固有属性域的语义关系(俊俏的容貌、英俊的相貌、清秀的脸庞)和固有属性值与属性承载体的语义关

张国宪曾把形容词划分为表物形容词和表行形容词两类,我们认为这是不妥的,因为主体形容词在形容词中数量很大,出现频率很高,具有非常独特的地位,其句法语义特征和组合规律非常明显。参见张国宪. 现代汉语的动态形容词[J]. 中国语文, 1995, (3):211-223.

系(俊俏的姑娘、英俊的小伙子、清秀的作家)。有的形容词通过与不同语义特征的名词组合,形成不同的语义关系,比如:聪明的老人(固有属性值与属性承载体)、聪明的大脑(固有属性值与固有属性域)、聪明的想法(固有属性值与属性呈现体),这样就可以通过语义关系扩展了形名组合研究的丰富性,也保证了正确地理解复杂形名组合的语义内涵,从而提取出更细腻的语义信息。

第四,语义组合模型中的时空类形名组合可分别归入【事体类】和【物体类】形名组合规律的研究中。这是因为:(1)语料库的调查发现,五类形名组合间区别性特征最明显的是主体类、事体类、物体类和评价类,而时空类除了逻辑和概念上存在着一定的语义基础外,其组合形式上的区别性特征并不明显,根据时空类形容词和名词的同现规律,它们容易向事体类和物体类漂移,比如【时长类】接近事体类,而【空间类】接近物体类。(2)【时长类】形名组合的特点更接近于事体类形名组合,从概念上,除了时间名词所指的事物具有时间长度属性之外,事体名词所表达的事件也具有在一维时间中持续的性质,如:

a: 漫长的时间、很长的周期、短暂的时光、短促的岁月。

b: 漫长的等待、很长的会议、短暂的交谈、短促的声音。

而且,从同现频率看,时长形容词与事体名词的同现率高于与时间名词的同现率,而且时长形容词集合的数量也很少,鉴于此,把【时长类】形名组合放在了【事体类】形名组合中更为合理。(3)【空间类】形容词主要有:大、宽、宽敞、宽大、宽绰、广阔、宽阔、开阔、辽阔、小、窄、狭小、狭窄、窄小、局促等,与它们同现的名词除了属性域“空间、领域、区域、范围”之外,主要是指称具有二维平面或者三维立体空间性质的空间名词,如:草原、水面、广场、体育场、疆土、土地、森林、沙地、田野、马路、走廊、教室、房间、礼堂、厂房、大厅、体育馆等,因为空间名词既可以与空间类形容词同现,也可以与物体类形容词同现,反映出的空间名词与形容词的组合特征非常接近于物体类形名组合的特点,比如:

a: 宽阔的街道、宽敞的教室、开阔的水面、狭窄的走廊、狭小的房间、窄小的厨房。

b: 干净的街道、肮脏的教室、很绿的水面、昏暗的走廊、明亮的房间、光滑的厨房。

可见,空间名词既可与体现自身特点的空间形容词组合,也可以与物体形容词同现,因此,可归入

物体类形名组合中。

赵军、黄昌宁^[15]曾从静态和动态角度研究了如何识别名词短语,这种思路是科学合理的。我们也从静态角度全面细致地调查了六种主体类(容貌、身体、态度、性格、智力、情感)、九种事体类(模态、难度、急度、速度、时长、强度、范围、深度、细度)、两种物体类(颜色、味道)和两种评价类(价值、怪异)形名组合的规律性,结果表明该形名语义组合模型既符合语感,也符合语言事实。另外,我们还从动态角度研究了整体形名组合在 $N1 + A + N2$ 格式、 $N1$ 有 $N2$ 很 A 格式、名形组合作状语、形名组合做状语等句法结构中的句法语义功能,研究结果表明,该形名语义组合模型能够很好地把词类与句法结合起来,揭示形名组合所提供的语法规则和语义信息对语言信息的理解和生成具有重要的价值。本项研究所涉及的形容词和名词数量统计如下:

表3 已分析的形名组合中名词形容词统计表

形名组合类型 \ 组合词类	形容词 A	名词 N
主体类形名组合	473	901
事体类形名组合	145	2 149
物体类形名组合	109	556
评价类形名组合	49	317
总计	766	3 923

参考文献:

- [1] 朱德熙. 语法答问[M]. 北京:商务印书馆, 1985:3.
- [2] 维特根斯坦. 逻辑哲学论[M]. 北京:商务印书馆, 1962:26.
- [3] 维特根斯坦. 逻辑哲学论[M]. 北京:商务印书馆, 1962:3.
- [4] 赵春利. 形名组合的静态与动态研究[D]. 广州:暨南大学 2006 年度博士学位论文.
- [5] 马建忠. 马氏文通读本[M]. 吕叔湘、王海棻编. 上海:上海教育出版社, 1986:49-50.
- [6] 黎锦熙. 新著国语文法[M]. 北京:商务印书馆, 1992:18.
- [7] 王力. 中国现代语法[M]. 北京:商务印书馆, 1985:17.
- [8] 朱德熙. 语法讲义[M]. 北京:商务印书馆, 1982:41.
- [9] 中国社会科学院语言所词典编辑室编. 现代汉语词典[M]. 北京:商务印书馆, 2005:953.

(下转第 24 页)

5 结束语

本文基于真实文本语料,利用 SVM 模型建立了介词结构的自动识别系统,尝试了语义类信息等不同的特征集合,将介词结构边界识别的准确率提高到了 90.95%。该实验结果表明,在浅层句法分析中,SVM 模型可以较好地实现介词结构边界识别任务,可以满足一些应用的需求。

文中尝试他语义类信息的使用,是对语义信息在中文信息处理中的应用的一个初步探索。在下一步的研究中,会分析实验中的错误结果,继续尝试其他可能有效的特征。

在进一步的工作中,我们将要把本文的介词结构识别软件嵌入到一个语音合成系统中,帮助判断韵律短语,以提高语音合成的效能。

参考文献:

- [1] Eneko Agirre, Timothy Baldwin, David Martinez. Improving Parsing and PP attachment Performance with Sense Information [C]// Proceedings of ACL 08. 2008: 317-325.
- [2] 干俊伟,黄德根. 汉语介词短语的自动识别[J]. 中文信息学报,2004,19(4):17-23.
- [3] 奚建清,罗强. 基于 HMM 的汉语介词短语自动识别研究[J]. 计算机工程,2007,33(3):172-182.
- [4] 王立霞,孙宏林. 现代汉语介词短语边界识别研究[J]. 中文信息学报,2005,19(3):80-86.
- [5] 李珩,朱靖波,姚天顺. 基于 SVM 的中文组块分析[J]. 中文信息学报,2004,18(2):1-7.
- [6] 黄德根,王莹莹. 基于 SVM 的组块识别及其错误驱动学习方法[J]. 中文信息学报,2005,20(6):17-24.
- [7] 俞士汶,段慧明,朱学锋,等. 北大语料库加工规范:切分、词性标注、注音[J]. 汉语语言与计算学报,2003,13(2):121-158.
- [1] Eneko Agirre, Timothy Baldwin, David Martinez. Im-
berg, 2002, (2553): 172-182.
- [13] 姬东鸿,黄昌宁. 汉语形容词和名词的语义组合模型[J]. 中文与东方语言信息处理学会通讯,1996,6(1):25-33.
- [14] 张瑞霞,张蕾. 基于知识图的汉语基本名词短语分析模型[J]. 中文信息学报,2004,18(3):47-53.
- [15] 赵军,黄昌宁. 基于转换的汉语基本名词短语识别模型[J]. 中文信息学报,1999,13(2):1-7.

(上接第 18 页)

- [10] 邵敬敏. “语义语法”说略[J]. 暨南学报,2004(1):100-106.
- [11] 闻扬,苑春法,黄昌宁. 基于搭配对的汉语形容词-名词聚类[J]. 中文信息学报,2000,14(16):45-50.
- [12] Igor A. Bolshakov and Alexander Gelbukh. On Semantic Classification of Modifiers[C]//Lecture Notes In Computer Science, Natural Language Processing and Information Systems, Springer Berlin / Heidel-