基于生成词库论和论元结构理论的语义知识体系研究*

袁毓林

（北京大学中文系/汉语语言学研究中心/教育部计算语言学重点实验室，北京 100871）

摘要：本文讨论如何构造合适的汉语语义描写体系并建设相应的语义知识库，从而为文本语义的计算机自动分析提供可靠的资源。文章提出的技术路线是：在生成词库论和论元结构理论的指导下，分别描写名词的物性结构和动词、形容词的论元结构（包括物性角色或论元角色集合及其句法配置格式集合），标定名词、动词和形容词的情感评价色彩，揭示相关名词、动词和形容词的物性角色和论元角色之间的关联和推导关系，从而形成比较完整的关于名词、动词和形容词的实体指称、概念关系和情感评价等多层面的语义知识。最后，还展示了这种多层面的语义知识在语义自动计算中的运用案例。

关键词：语义描写体系；语义知识库；物性结构；论元结构；情感评价；语义关联

中图分类号：TP391　　**文献标识码**：A

A Study of Chinese Semantic Knowledge System Based on the Theory of Generative Lexicon and Argument Structure

Yuan Yulin

(Dept. of Chinese Lang. & Lit., Peking University / Research Center of Chinese Linguistics / Ministry of Education Key Laboratory for Computational Linguistics, Beijing 100871,China**)**

Abstract：This paper discusses how to construct a practical Chinese semantic knowledge system and corresponding database for the purpose of computing Chinese text meaning. It proposes the working procedure as follow: (1) Under the instruction of Generative Lexicon Theory and Argument Structure Theory, describing the qualia structure of nouns and the argument structure of verbs and adjectives, which include both the set of qualia roles or themantic roles, and the syntactic constructions constituted by such nouns, verbs and adjectives. (2) Giving the semantic orientation and sentiment polarity of the nouns, verbs and adjectives, which are indicated with the 5 points scale. (3) Revealing and connecting the correlative and inference relation of the qualia roles and themantic roles of related nouns, verbs and adjectives as a network. (4) Integrating all the meaning of entity reference, conception relation, and sentiment polarity of the nouns, verbs and adjectives into a multi-level semantic knowledge database. Finally, a case study of computing meaning with the help of the multi-level semantic knowledge is presented.

Key Words: Semantic Description System; Semantic Knowledge Database; Qualia Structure; Argument Structure; Sentiment Polarity; Semantic Correlation

1 引言：研究目标和技术路线

本文讨论的主要问题是：如何建设合适的汉语语义知识库，从而为文本语义的计算机自动分析提供可靠的基础。重点研究的内容是：如何描写汉语名词、动词、形容词等实词的语义结构以及其间的关联关系，为名词和动词、形容词设计出前后一致、互相照应的语义表示框架，形成完整的汉语语义知识体系，并且转化为结构合理、使用方便的汉语语义知识库。

* **基金项目**：本课题的研究得到国家社科基金重大招标项目《面向网络文本的多视角语义分析方法、语言知识库及平台建设研究》（批准号：12&ZD227）的资助，谨此致以诚挚的谢意。

作者简介：袁毓林（1962—），男，教授，主要研究方向为汉语语法学和计算语言学。

对于上述问题，我们采用的技术路线是：以生成词库论（generative lexicon theory）和论元结构理论（the theory of argument structure）为指导，充分研究汉语常用的名词、动词、形容词的物性结构（qualia structure）和论元结构（argument structure），揭示它们之间的搭配连接和选择限制关系；还要刻画汉语常用的名词、动词、形容词的情感评价色彩（semantic orientation and sentiment polarity），最终形成完整的汉语语义知识体系，并且转化为具有可扩展性的、面向对象（object orientation）的语义知识数据库。从而为计算机自动分析文本的语义提供充分的语义知识资源。

2 名词的物性结构知识的描述体系

对于汉语名词的语义结构，我们主要采用生成词库论的物性结构的描写框架，从形式角色（Formal role）、构成角色（Constitutive role）、施成角色（Agentive Role）和功用角色（Telic role）等多个方面，说明名词所指谓的事物的性质及其跟相关事物、事件的关系；这种概念层面上的事物或事件关系，最终在语言层面上表现为词语（名词跟名词、动词、形容词等）之间的搭配关系（即选择限制关系）。在物性角色的数量和类型上，我们根据汉语名词在真实文本中的词语搭配情况，突破了 Pustejovsky (1995)的上述四种，扩展到下列九种[1]：

（1）**形式**（formal，简写为 FOR）：用以反映名词的分类属性、语义类型和本体层级特征（semantic classes and ontological plane）。比如，“**水**”是“有形物质、液体”、“**医生**”是“人、身份、职业”，等等；

（2）**构成**（constitutive，简写为 CON）：用以反映名词所指的事物的结构属性，包括：构成状态、组成成分、在更大的范围内构成或组成哪些事物、跟其他事物的关系，也包括物体的大小（magnitude）、形状（shape）、维度（dimensionality）、颜色（color）和方位（orientation），等等。比如，“**树**”的构成是“比较高大的植物，有躯干、枝条和叶子；可以根据所结的果子、来源、用途、形状、特征等属性进行分类：果（子）、苹果、梨、橘（子）、柑桔、油棕、落叶、常青、相思、圣诞、痒痒、糖槭、庭院、道路，等等；也可以根据颜色进行分类：绿色、白色、黄色、黑色、红色、褐色，等等”；

（3）**单位**（unite，简写为 UNI）：用以反映名词所指事物的计量单位，也即跟名词相应的量词；如：“张[纸]、双[筷子]、斤[白酒]、点儿[事情]、口袋[面粉]、[看三]次[电影]、[三]天[时间]”，等等；

（4）**评价**（evaluation，简写为 EVA）：用以反映人们对名词所指事物的主观评价、情感色彩。比如，对“月亮”的评价有“洁白、皎洁、明亮、明朗、朦胧、圆圆、圆润、弯弯”，对“妈妈”的评价有“伟大、英雄、勇敢、慈祥、慈爱、无私”等等；

（5）**施成**（agentive，简写为 AGE）：用以反映名词所指的事物是怎样形成的，如创造、天然存在、因果关系等。比如，“**抽屉**”的施成是“制作、做”等等，“**细菌**”的施成是“滋生、培养、繁殖、感染（上）”等等；

（6）**材料**（material，简写为 MAT）：用以反映创造名词所指的事物所用的材料。比如，“**椅子**”的材料是“木头、竹子、藤子、木、竹、藤、钢、铁、塑料、硬板”等等，“**书**”的材料是“帛、竹、纸草、羊皮、竹皮、树叶、纸版、电子”等等；

（7）**功用**（telic，简写为 TEL）：用以反映名词所指的事物的用途和功能。比如，“**抽屉**”的功用是“盛（东西）、放（衣物）、装（文件）、搁（杂物）、藏（东西）”等等，“**医生**”的功用是“治病、治疗疾病”等等；

[1] 详见袁毓林（2012）。

（8）**行为**（action，简写为 ACT）：用以反映名词所指的事物的惯常性的动作、行为、活动。比如，“**细菌**”的行为是“繁殖、生长、死亡、吞噬、传播、散布、感染、侵染、进入、侵入、分解、腐蚀”等等，“**妈妈**”的行为是“生孩子、抚养孩子、照顾小孩、养育小孩、教育孩子”等等；

（9）**处置**（handle，简写为 HAN）：用以反映人或其他事物对名词所指的事物的惯常性的动作、行为、影响。比如，对“**眼泪**”的处置是“抹、含着、噙着、忍着、充满、擦、弹”等等，对“**意见**”的处置是“转达、转告、听到、理解、谅解、接受、无视、不理会”等等。

除了描写名词在语义上的各种物性角色之外，还要描写名词跟其物性角色在句法上的组配关系，形成完整的关于名词的句法-语义接口知识。例如：

食品 shípǐn 〈名词，积极〉商店出售的经过加工制作的食物。

〔1〕**物性角色**：

形式 FOR：有形物质、商品、可摄入物、食物；

构成 CON：食品有营养、热量等构成因素；可以根据来源、功能、加工或包装方式、期限等属性进行分类：鱼类、肉类、禽类、鸡肉、奶（类）、植物性、动物性、副、糖类、舶来、保健、营养、药用、方便、快餐、应急、生、熟、生鲜、腌制、熏制、强化、膨化、冷冻、速冻、罐头、罐装、听装、袋装、酸性、绿色、环保、有机、转基因、风味、清真、婴儿、老人、动物、军队、野战、节日、过期、隔夜、污染，等等；

单位 UNI：集合：批、包、种、部分，等等；度量：吨、公斤，等等；不定：点儿、些，等等；容器：箱、口袋、桌子、屋子、篮子，等等；

评价 EVA：新鲜、变质、腐败、美味、珍贵、廉价、传统、新颖、特殊、精细、高级、优质、变质、短缺、丰富（多样）、充足、匮乏，等等；

施成 AGE：加工、制作，等等；

功用 TEL：吃、吞食、享用、品尝、消费，等等；

处置 HAN：出售、购买、存放、冷藏、包装、运输、分发、给…消毒，等等。

〔2〕**句法格式**：

S1：__ + 有/的 + CON

如：~有营养 | ~有热量 | ~的营养 | ~的热量

S2：Num + UNI + __

如：一批~ | 一包~ | 一种~ | 一部分~| 一吨~ | 一点儿~ | 一些~ | 一箱~| 一口袋~| 一桌子~ | 一屋子~ | 一篮子~

S3：EVA +（的+）__

如：新鲜（的）~ | 变质（的）~ | 腐败（的）~| 美味（的）~| 珍贵（的）~| 廉价（的）~| 传统~ | 新颖（的）~ | 特殊~| 精细~| 高级~| 优质~

S4： __ + EVA

如：~（严重）短缺 | ~丰富（多样）|~充足 |~匮乏

S5：AGE + __

如：加工~ | 制作~ | 做~

S6：TEL + __

如：吃~ | 吞食~ | 下咽~ | 享用~ | 品味~ | 尝~| 品尝~

S7：HAN + __

如：出售~ | 卖~ | 销售~| 买~ | 购买~| 存放~ | 冷藏~ | 包装~ | 运输~

价格 jiàgè〈名词，中性〉商品价值的货币表现。比如，一件衣服卖五十元人民币，五十元就是衣服的价格；你化了多少钱买某种东西，你所化的钱的数量就是这种东西的价格。

〔1〕**物性角色**：

形式 FOR：抽象属性、经济领域、商品属性；

构成 CON：价格跟商品（x）、价值、货币数量和货币单位及货币币种（y）等概念密切相关；价格（属性名词）是一种属性名称，依附于某种商品（宿主名称），货币数量和货币单位（数量词）是价格这种属性的值；可以根据价格的宿主进行分类：商品、消费品、工业品、原材料、粮食、棉花、油料、农产品、能源、石油、黄金、白银、建筑材料、药品、土地、农资、旅游、消费、医疗、现货、期货、股票、债券，等等；可以根据价格的属性进行分类：市场、挂牌、平均、批发、零售、指令性，等等；下面的评价 EVA 是人们对于某种商品的价格与价值之间是否相当的评价，包括货币数量和货币单位（y）这种确定值和正面 vs.反面这种感性的模糊值；

单位 UNI：集合：种、类、部分，等等；不定：些，等等；

评价 EVA：是 y，高、低、贵、昂贵、不菲、过高、便宜、低廉、偏低、优惠、合理，等等；

施成 AGE：确定、决定，等等；

功用 TEL：反映价值，等等；

行为 ACT：变化、波动、异动、上涨、大涨、猛涨、暴涨、下降、跌落、大跌、狂跌、反弹、回稳、（趋于）稳定、随行就市、维持在高位/低位、偏离价值、与价值相背离，等等；

处置 HAN：监督、监测、调整、控制、调控、限制、改变、提高、降低、计算、核定、公布，等等。

〔2〕**句法格式：**

S1：x + 的 + __ + EVA

如：一桶石油的~是 120 美元 | 一斤大米的~是 12 元（人民币） | 蔬菜的 ~ 很贵 | 农产品的 ~ 十分低廉

S2：Num + UNI + __

如：一种 ~ | 一类 ~ | 一部分 ~ | 一些 ~

S3：AGE + __

如：确定 ~ | 决定 ~

S4：（x +的 +）（这种+）__ + TEL

如：~反映价值 | ~没有反映价值 | 黄金的 ~ 反映了它的价值 | 铁矿石的这种 ~ 已经充分地反映了它的价值

S5： __ + ACT

如：~变化 |~波动 |~异动 |~上涨 |~大涨 |~猛涨 |~暴涨 |~下降 |~跌落 |~大跌 |~狂跌 |~反弹 |~回稳 |~（趋于）稳定 |~随行就市 |~维持在高位/低位 |~偏离价值 |~与价值相背离

S6：HAN + __

如：监督 ~ | 监测 ~ | 调整 ~ | 控制 ~ | 调控 ~ | 限制 ~ | 改变 ~ | 提高 ~ | 降低 ~ | 计算 ~ | 核定 ~ | 公布 ~ | 分析~| 研究~

这样，通过物性角色，我们在概念层面上刻画了名词所指的事物的基本属性及其跟相关事物或事件的关系；通过句法格式，我们刻画了名词跟相关的名词、动词和形容词的选择限制和搭配关系。最终，通过名词的物性结构的描述框架，形成了比较完整的关于名词的句法-语义接口的知识。因此，这种物性结构的描述框架，可以看作是对名词的句法、语义知识的一种简略的概念建模和语言建模。目前，我们已经对常用的 1,000 多个名词进行了物性结构框架描述。接下来，将对汉语水平考试词汇表中的 4,000 多个名词进行描写。

3 动词、形容词的论元结构知识的描述体系

对于汉语动词、形容词等谓词的语义结构，我们主要采用论元结构的描写框架，对常用的 6,000 多个动词和 3,000 多个形容词的常用义项，分别建立格式一致的语义角色框架及其句法实现形式（即句法格式）。内容包括：（i）角色集合：每个谓词在某个义项下其各个论元的语义角色集合，（ii）句法格式：该谓词跟受其支配的这些论元角色在句子中的句法配置方式。其中，动词的论元角色首先分为**必有论元**和**非必有论元**两种，前者是构成意思相对完整的句子所不可缺少的，后者则用以扩充句子的意思，帮助形成意思相对复杂的句子。必有论元分为**主体论元**和**客体论元**两种，前者主要作主语，后者主要作宾语。主体论元细化为**施事**、**感事**、**经事**、**致事**、**主事**等语义角色，客体论元细化为**受事**、**与事**、**结果**、**对象**、**系事**等语义角色。非必有论元从语义上分为**依凭论元**、**环境论元**和**关涉论元**三种，它们主要作状语。其中，依凭论元细化为**工具**、**材料**、**方式**、**原因**、**目的**等语义角色，环境论元细化为**时间**、**处所**、**源点**、**终点**、**路径**等语义角色，关涉论元细化为**量幅**、**范围**等语义角色。总共为动词设立了 22 种语义角色。考虑到这种抽象的语义角色的定义难以适应到具体某个动词的某种论元，我们采用个例化的语义角色描述方法；即根据每一个谓词的特定的意义（或用法），对其所有的语义角色进行具体的语义描写。例如：

吃 <体宾动词，中性> 进食；把食物等放到嘴里咀嚼并吞咽下去。

〔1〕**语义角色**：

施事 A：吃东西的人或动物；

受事 P：施事所吃的东西；

与事 D：施事吃他东西的人；

工具 I：吃东西所用的器具，如“碗、筷子”等；

方式 M：吃东西的方式或某种伙食标准；

处所 L：吃东西的地点；

终点 GO：受事被吃后所到的地方，一般是“肚子（中）、嘴（里）”等身体部位。

〔2〕**句法格式**：

S1: A + __ + P

如：弟弟 ~ 了一个苹果。 | 咱们 ~ 烤鸭吧。

S2: P + A + __

如：苹果我 ~ 了。 | 蛋糕大家都 ~ 了。

S3: A + __ + D + P

如：他 ~ 了小李一个苹果。 | 弟弟 ~ 了我一包巧克力。

S4: A + 用 I + __ + P

如：长工们都用大碗 ~ 饭。 | 他正用刀叉 ~ 牛排呢。

S5: I + （A +）__ + P

如：这个碗我 ~ 面条。 | 这副刀叉 ~ 牛排。

S6: A + __ + I/M

如：男人们 ~ 大碗，孩子们 ~ 小碗。 | 他一直 ~ 小灶。 | 工人们都 ~ 包伙。

S7: A + 在 L + __ + P

如：学生们都在食堂 ~ 午饭。 | 他们在全聚德 ~ 晚饭。

S8: P + A + 在 L + __

如：午饭他在食堂 ~ 。 | 早饭孩子们都在家里 ~ 。

S9: P + A + __+ L

如：午饭他 ~ 食堂。| 晚饭咱们 ~ 馆子吧。

S10: A + 把 P + __了

如：你快把面条 ~ 了。| 弟弟把整块蛋糕都 ~ 了。

S11: A + 把 P + __ +（到/在）GO

如：犯人把纸团 ~ 到肚子里了。| 小猴子已经把果仁 ~ 到嘴里了。

S12: P + 被 A + __了

如：面条被他 ~ 了。| 生日蛋糕被邻居的孩子 ~ 了。

S13: P + 被 A + __ +（到）GO

如：孙悟空被铁扇公主 ~ 肚子里了。| 果仁已经被小猴子 ~ 到嘴里了。

制作 〈体宾动词，中性〉把原材料做成成品。

〔1〕**语义角色：**

施事 A：把原材料做成成品的人；

结果 R：施事所制作的成品；

材料 MA：施事制作成品所用的材料。

〔2〕**句法格式：**

S1: A + （用 MA + ）__ + R

如: 那个公司 ~ 了大量制冷发动机。| 美国大学生米勒 ~ 了能模拟原始大气的仪器。 | 先人用石头 ~ 劳动工具。 | 古人常用云母片 ~ 屏风。

S2: A + 把 R/MA + __+ 出来 / 成 R

如: 小五赶夜把专辑 ~ 出来了。 | 人们把葡萄 ~ 成葡萄酒。 | 师傅把这块木头 ~ 成了一个小凳子。

S3: MA + 可以 + __ + 成 R

如: 梨的果实营养丰富，除鲜食外，还可以 ~ 梨脯、梨汁、梨膏、梨酒等。| 这种木材可以 ~ 成高级家具、乐器和工艺品。

S4: MA + 被（A） + 用来+__ + R

如：葡萄被用来 ~ 酒。 | 玉被人们用来 ~ 艺术品。

对于形容词的各种论元，我们根据它们跟形容词在意义上的不同的关系，区分为下列 9 种不同的语义角色：**主事、感事、范围、与事、量幅、对象、系事、原因、目的**，等等。对于形容词的语义角色，我们也采用个例化的语义角色描述方法。例如：

贵 <形容词，消极> 价格高；价值大。跟“贱”相对。

（1） **语义角色：**

主事 TH：具有价格高、价值大这种属性的物体；

范围 RA：主事表现出贵这种属性的具体方面，一般是**价格**、价值等；

与事 D：主事跟它在贵这种属性上进行比较的参照物；

量幅 EXT：主事和与事在贵这种属性上的差别所达到的程度或幅度。

（2） **句法格式：**

S1：TH +(RA+)__

如：这些仪器非常~。| 友谊商店的化妆品价格很~。

S2：TH+(RA+)比 D +__（+ EXT）

如：这些仪器价格比那些仪器~。| 这台电脑比那台电脑~三千元。

S3: __ +的+ RA

如：（这么）~的价格 |（很）~的价钱

S3: (RA+)+__ +的+TH

如：（这么）~的书 |（价格）很~的一块手表

S4: (RA+)比 D+__+（EXT+）的+ TH

如：（价格）比这本书还 ~ 的书 | 比那个项链还 ~ 三百块钱的戒指

谓词的这种句法、语义描述体系，具体地刻画了动词、形容词在语义结构和句法组配方面的特点。这种个例化的语义角色知识是一种非常重要的资源，特别有利于计算机理解语句的基本的命题意义，并调用这种知识来进行有关的自动推理。这种把动词的语义角色及其句法组配相结合的描述方式，充分地表示了谓词的论元结构和语义角色关系的各种重要的信息。通过这种谓词的论元结构的描述框架，形成了比较完整的关于动词、形容词的句法-语义接口的知识；也可以看作是对动词和形容词的句法、语义知识的一种简略的概念建模和语言建模。

4 词语的情感评价色彩知识的描述体系

语言不仅有传递事实性信息的功能，而且还有表示情感性评价的功能。对于同样一件事情，可以用积极性词语进行正面肯定，也可以用消极性词语进行负面否定。比如，对于同样一辆城市越野车，喜欢它的人会说它“马力大、结实耐用”，而讨厌它的人会说它“油耗大、粗重笨拙”；对于同样一款手机，喜欢它的人会说它“功能齐全、性价比好”，而讨厌它的人会说它“功能多余、价格昂贵”。这种正面评价和褒扬背后的会话蕴含可能是推荐听话人购买这种产品，而这种负面评价和贬斥背后的会话蕴含可能是劝阻听话人购买这种产品。也就是说，人们不仅用语言来报道有关事实，而且还通过渗透在话语中的情感倾向和评价色彩来影响听话人的思想、感情和行动，劝说他人相信某种情况、甚至做说话人所希望的事情。从上面的例子可见，这种文本的情感倾向和评价色彩往往通过具有不同的情感评价色彩的词汇来实现的。因此，我们首先要研究词语的情感评价色彩。

对于汉语常用的名词、动词、形容词的情感评价色彩，我们拟采用 5 点式量表的方式，把词语的情感评价色彩分为 5 级：褒义（+2）、积极（+1）、中性（0）、消极（-1）、贬义（-2）。这样，可以克服情感评价色彩划分粒度过于粗粝的弊病，更好地反映人们对于事物情感评价的连续性和梯度性。例如：

褒义词（+2）：好事、硕果、好人、歼灭、豪饮、颂扬、勇猛、圆满、高级

积极词（+1）：婚事、成果、人民、击毙、小酌、推荐、激烈、圆通、上游

中性词（ 0）：事情、结果、人们、杀死、喝酒、宣传、猛烈、变通、中间

消极词（-1）：事件、后果、闲人、杀害、贪杯、宣扬、凶猛、圆滑、下游

贬义词（-2）：事故、恶果、坏人、屠杀、酗酒、吹捧、凶恶、油滑、低级

其中，中性词是在情感倾向和评价色彩上比较客观的词语，其他四种是在情感倾向和评价色彩上比较主观的词语；褒义词和积极词是在情感倾向和评价色彩上比较正面的词语，贬义词和消极词是在情感倾向和评价色彩上比较负面的词语；褒义词在情感倾向和评价色彩上的正面性比积极词更强，贬义词在情感倾向和评价色彩上的负面性比消极词更强。可见，这种 5 级体系可以根据应用的需要，方便地映射到正负（或褒贬）二值的情感极性空间。

此外，还要研究当这些情感色彩不同的词语跟“很、非常、太、过于、偏、稍微、不过、仅仅、不”等程度、范围和否定副词、“而已、罢了”等语气词组合以后，在情感的极性（正面 vs.反面）和强度（增强 vs.减弱）方面的变化。

对于词语的情感倾向类型，我们兼顾汉语传统的“七情”（喜、怒、哀、惧、爱、恶、欲）等通俗分类，参照情绪心理学和人格心理学等学科对于人类情感的分类，再联系相关情

感词语在句法、语义上表现出来的对立和互补特点，把情感词语分为 6 大类和若干小类：

快乐：高兴、兴奋、愉快、激动、喜悦、欢喜、宽慰、安心、平静

喜好：喜欢、喜爱、欲求、信任、相信、尊敬、赞扬、歌颂、推荐

悲哀：悲伤、失望、忧愁、哀愁、烦恼、郁闷、忧郁、内疚、后悔

惊恐：吃惊、惊讶、惊奇、慌张、惊慌、恐惧、害羞、焦虑、不安

愤怒：气愤、气恼、愤慨、恼火、发火、生气、发怒、不满、泄愤

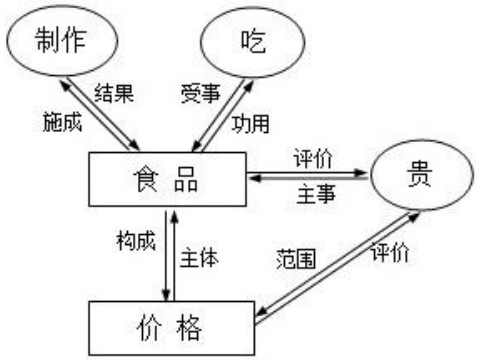
厌恶：讨厌、嫌弃、反感、怀疑、嫉妒、愤恨、批评、指责、贬斥

这方面的研究，我们还刚刚开始，希望有了具体的结果再详细讨论。

5 名词、动词、形容词的语义关联和互动推导

上文关于名词、动词和形容词的物性结构和论元结构及其情感评价的知识可以关联起来，形成以名词为中心的相关语义知识的互动和推导。比如，名词“食品”的施成角色是动词“制作”等、功用角色是动词“吃”等，这是从名词出发看名词和动词的语义关联；反过来，从动词出发看动词和名词的语义关联，名词“食品”是动词“制作”等的结果角色、是动词“吃”等的受事角色。同样，名词“食品”的构成角色是名词“价格”等、“食品”和“价格”的评价角色都是形容词“贵”等；反过来看，一价名词“价格”的主体角色是名词“食品”等，形容词“贵”的主事角色是“食品”等名词，范围角色（表示一种属性、维度）是名词“价格”等。从中可以发现，形容词“贵”既可以先评价范围角色“价格”等，再间接地评价“食品”等主体角色；也可以跳过范围角色“价格”等，直接评价“食品”等主体角色。在后面这种情况下，主事名词直接跟评价形容词组合，范围名词隐藏起来了。在文本的情感分析中，这种藏在后面的范围名词被称为“隐特征”（implicit feature）。这种隐特征的语义理解，对于人来说不成问题；但是，对于计算机而言就有理解障碍[2]。现在，我们通过对形容词的论元结构和名词的物性结构的刻画，揭示“**食品—（价格）—贵**”之间的语义关联，从而为计算机理解名词—形容词之间的语义关系，提供一种有效的知识表示。

下面是“**制作/吃—食品—价格—贵**”等几个相关的名词、动词和形容词的物性结构和论元结构及其语义关联的图示：



这样，在调查大规模真实文本语料的基础上，通过对名词、动词和形容词等实词的物性结构和论元结构的精心设计和合理描述，可以把事物和跟事物相关的事件的有关世界知识及其语言表达形式表示出来。再辅之以指针链接和知识图谱（knowledge graph）等数据表示技术和拉近—推远（zoom-in and zone-out）等便捷的呈现手段，可以有效地把相关的名词、动

[2] 这一点，承北京大学计算语言学研究所王厚峰教授告知，谨此谢忱。

词和形容词的语义关联起来，并且形成以名词（实体）为检索核心的、面向对象（object orientation）的语义知识库。

6 多层面语义及其关联知识在语义计算上的应用

一般来说，名词涉及时间、地点、人物、事物等实体指称意义，动词、形容词涉及性状、行为、联系等关系意义，并且许多名词、动词和形容词还有情感评价意义。我们的研究一方面要区分词语的所指、概念和评价意义，另一方面又要借助名词的物性结构和动词（包括形容词）的论元结构的描写框架，把这些不同的意义整合和连接起来。显然，这种融合了相关百科知识的综合性语义知识，对于计算机歧义消解是十分重要的；它使得本来人可以意会、但是难以精确地表示（以便机器调用）的知识，得到了明确的表示，并且具备了完整的描述体系。比如，一个经典的歧义例子“鲁迅的书”，对于人来说，可以轻易地解读出它至少有两种意义：（a）‘鲁迅[**拥有**]的书’（b）‘鲁迅[**写**]的书’。对于机器来说，（a）这种意义或许可以通过语义解释规则——“NP1+的+NP2”表示‘NP1+拥有+的+NP2’，当 NP1 表示人或机构、NP2 表示物品时——来表示和获取；但是（b）这种意义就不容易表示和处理。现在，有了专有名词“鲁迅”的所指意义和普通名词“作家、书”等的物性结构知识：

作家 zuòjiā 〈名词，积极〉从事文学创作有成就的人。

〔1〕**物性角色：**

形式 FOR：人、身份、职业、文化人；

构成 CON：作家可以根据其所创作的作品的体裁、题材、发表园地等进行分类：小说、散文、戏剧、专栏、影视、网络，等等；可以根据其国籍、地区、语种、人种（或肤色）等进行分类：中国、外国、英国、法语，等等；可以根据其时代、性别、年龄和身份、职业等进行分类：古代、现代、男、美女、青年、专业、业余、军人，等等；可以根据其流派或思想倾向进行分类：古典派、现代派、现实主义、浪漫主义、后现代主义、左翼、右派、学院派，等等；

评价 EVA：伟大、著名、知名、成名、杰出、优秀、代表（性）、先锋、新锐、重要、（第）一流、二流、三流、不入流，等等；

施成 AGE：当、做、成为，等等；

功用 TEL：写（书、文章、作品等）、创作（小说、诗歌等文学作品），等等；

〔2〕**句法格式：**

S1：CON + __

如：小说~ | 散文~ | 戏剧~ | 专栏~ | 影视~ | 网络~ | 中国~ | 外国~ | 英国~ | 法语~ | 古代~ | 现代~ | 男~ | 美女~ | 青年~ | 军人~ | 古典派~ | 现代派~ | 现实主义~ | 浪漫主义~ | 后现代主义~

S2：EVA +（的+）__

如：著名~ | 知名~ | 伟大的~ | 杰出的~| 优秀（的）~| 先锋~ | 新锐~ | 重要（的）~ | （第）一流（的）~ | 二流~ | 三流~ | 不入流的~

S3：AGE + __

如：当~ | 做~ |成为~

S4：__ + TEL

如：~写（书、文章、作品） | ~创作（小说、诗歌等文学作品）

书 shū 〈名词，中性〉装订成册的印刷品。

〔1〕**物性角色：**

形式 FOR：人造物、印刷品、文化用品；

构成 CON：一般由纸张、文字、图画，内容、信息等物质和文化因素组成；可以根据科目、内容或功能进行分类：语文、数学、历史、地理、化学、物理、生物、外语、必读、参考，等等；也可以根据颜色进行分类：白色、黄色、绿色、棕色，等等；

评价 EVA：大、小、好、坏、新、旧、破、淫、普通、特殊，等等；

施成 AGE：写、印、印刷、出、出版，等等；

材料 MAT：帛、竹、纸草、羊皮、竹皮、树叶、木板、纸版、电子，等等；

功用 TEL：读、念、看，等等；

处置 HAN：买、卖、收藏、拿、借、还、扔、撕、烧、焚、啃、浏览，等等。

〔2〕**句法格式：**

S1：__（上/中）+ 的 + CON

如：~的纸张 | ~（上）的文字 | ~（上）的图画 | ~的内容 | ~中的信息

S2：CON +（的+）__

如：语文~ | 数学~ | 历史~ | 地理~ | 化学~ | 物理~ | 生物~ | 外语~ | 必读~ | 参考~ | 彩色（的）~ | 红色（的）~ | 褐色（的）~

S3：EVA +__

如：大~ | 小~ | 好~ | 坏~ | 新~ | 旧~ | 破~ | 淫~ | 普通~ | 特殊~

S4：AGE + __

如：写~ | 印~ | 印刷~ | 出~ | 出版~ | 制作~

S5：MAT +（AGE +）（的+）__

如：帛/竹/纸草/羊皮/竹皮/树叶/木板（制作/印刷）（的）~ | 纸版/电子~

S6：TEL + __

如：读~ | 念~ | 看~

S7：HAN + __

如：买~ | 卖~ | 收藏~ | 拿~ | 借~ | 还~ | 撕~ | 烧~ | 焚~ | 啃~ | 浏览~

在一定的语义解释规则的指引和约束下，通过调用专有名词“鲁迅”的百科知识，得到他的身份（或职业）是作家；再调用普通名词“作家”的物性结构知识，得到其功能角色是“写（书）”；再调用普通名词“书”的物性结构知识，得到其施成角色是“写”；最后通过某种特征加权机制，就可以为“鲁迅的书”获得“写”这个隐含的释义动词（implicit paraphrasing verb）。最终，不仅完成了歧义结构的识别，而且获得了歧义结构的多种语义解读。

可见，这种多层次的语义知识体系对于信息抽取、内容计算、舆情分析、产品评论观点挖掘等多种自然语言处理和应用任务，都具有重要的资源支撑作用。

参考文献

[1] 董振东、董强《知网》；http://www.keenage.com/zhiwang/c_zhiwang.html。

[2] 袁毓林（1998）《语言的认知研究和计算分析》，北京：北京大学出版社。

[3] 袁毓林（2008）《面向信息检索系统的语义资源规划》，《语言科学》第1期，第1-11页。

[4] 袁毓林（2010）《汉语配价语法研究》，北京：商务印书馆。

[5] 袁毓林（2012）《汉语名词物性结构的描写体系和运用案例》，待刊。

[6] Baker, F. Collin, Charles J. Fillmore, John B. Lowe (1998) The Berkeley FrameNet Project, *Proceedings of the 17th International Conference on Computational Linguistics* [COLING '98] / *the 36th Annual Meeting on Association for Computational Linguistics* [ACL '98]- Volume 1, pp.

86-90, Montreal, Canada, August 1998.
[7] Fellbaum, Christiane (ed.) (1998) *WordNet: An Electronic Lexical Database*. MIT Press: Cambridge, Mssachusetts.
[8] Gildea, D., Jurafsky, D. (2002) Automatic Labeling of Semantic Roles, *Computational Linguistics*, 28:3, pp. 245-288.
[9] Palmer, M., Gildea, D., Kingsbury, P. (2005) The Proposition Bank: A Corpus Annotated with Semantic Roles, *Computational Linguistics*, 31:1, pp. 71-105.
[10] Pustejovsky, James (1995) *The Generative Lexicon*, Cambridge, Massachusetts: The MIT Press.
[11] Pustejovsky, James (2006) Type Theory and Lexical Decomposition. *Journal of Cognitive Science* 7(1): 39-76.