

基于类义抽象的汉语复合词义的求解模式探索

宋春阳¹ 陆汝占² 方向红¹

¹上海师范大学语言研究所 上海 200234 ²上海交大计算机科学与工程系 上海 200030

E-mail: songchy63@163.com

摘要: 文章以内涵逻辑理论为指导, 吸取训诂学精华, 通过提取词的抽象类义来把握词的语义结构, 以达到使汉语语义分析和解释形式化的目的。文章首先分析了未登陆词语义的求解困难, 由此提出运用抽象类义求解复合词义的策略, 论述了抽象类义的原则、操作方法。最后通过词语“关”验证了由抽象类义求组合义的模式。

关键词: 现代汉语, 复合词义求解, 类义抽象, 内涵逻辑

Research on the Interpretation Pattern of Compound-word Meaning by Generalized Meaning Abstract

Song Chunyang*

Lu RuZhan **

Fang Xianghong*

* Linguistic Institute of Shanghai Normal University, Shanghai 200234

** Department of Computer Science and Engineering of Shanghai JiaoTong University, Shanghai 200030

E-mail: songchy63@163.com

Abstracts: Based on the intensional logic theory, absorbing the delicacy of traditional Chinese lexicology, This paper intends to grasp semantic structure of words by generalized meaning abstract, so as to formalize Chinese semantic analysis and interpretation. After introduce the difficulty on the formalization of semantic analyses and interpretation of newcome words, we lodge our solution on the formalization of compound meaning by generalized meaning abstract and discuss its principle and operating methods. In the end, we test and verify its models by the example word “guan”.

Key words: modern Chinese, generalized meaning abstract compound meaning, formalization of Chinese semantic analyses and interpretation, intensional logic

1 引言

不管词库有多大，总是有限的，总会有一些未登陆词在真实的文本中出现。计算机在理解这类未登录词的语义时往往就会显得无能为力。目前处理这类问题较普遍的方法是通过相似度比较，求出最优解，因为这是一个基于统计规则的算法，得出的最优解往往却不是正确解，所以人们开始寻求通过语素库加上构词方式来求出新词的意义道路，也就是传统语言学界所说的“语素构词”或“生词熟字”说在机器上的实践。本文就是这方面的探索，希望能为最终在机器上解决复合词义的理解提供一些支持和帮助。

关于依据语素理解词义，历来有两种意见。一是否定说，认为词义不是组成成分意义的简单相加（范可育 1993），不能简单的“见字知义”。二是赞成说，认为词、短语和句子构造是一致的，可以从语素推知词义（徐通锵 1998，苏宝荣 1999、2002）。我们认为两种意见都有一定道理，但都有片面性，都是各自抓住了一部分词的语义构造现象而做出了片面的结论。简单地说“见字知义”可以不可以已没有意义，应该做的是首先对词进行分类，看哪些是可用规则进行描述即可计算的，哪些是不能描述不能计算的。

如果我们相信有一部分词是可以通过对语素组合规则求解出来的，那么这里就有两个问题第一、哪些可以计算，哪些不能计算？判定的标准或依据是什么？第二、如果可计算，那么计算的规则又是什么？对于第一个问题我们判定可计算的标准就是叠置原理。对于第二个问题，我们主要采用类义抽象的方法。

2 组合义的求解的困难及抽象类义的提出

2.1 组合义的求解的困难

$A + B = C$ ，这个公式表示由 A 和 B 两个单位可以生成更大的单位 C。语法单位有形式和意义两个方面，造句规则在说明怎样由 A 和 B 的形式得出 C 的形式时，同时也要说明怎样由语义|A|和|B|得出|C|，二者缺一就不成其为用词造句的规则。

$C = A + B$ （句法）和 $|C| = |A| \cdot |B|$ （语义）同构是我们确定是否符合叠置原理的一个原则。但如何能够从句法形式复合 $A+B$ 求出语义成分 $|A| \cdot |B|$ 的组合呢？现在问题是：动名、名名搭配能力是由词汇意义决定的，凭意义就可预测一个词语所能搭配的词，这种看法从词汇学、语义学的角度应该具有一定合理性，似乎复合义应该能从词典中单个词的义项中推测出来。但是一个字往往代表几个同形词，一个词又有多个义项的，我们根据什么确定在复合结构中取哪一个义项呢？这是复合义求解的困难所在。

2.2 抽象类义的目的

机器工作的原理是由句子→短语→词→语素→原子，最终为机器所理解；相反则是由基本原子一步一步合成句子。抽象类义的目的是化简，主要服务于构词。

$$C = A+B,$$

$$A+B\{B_1, \dots, B_n\},$$

$$A\{A_1, \dots, A_n\}+B\{B_1, \dots, B_n\},$$

组合情况 $n \times m$ 。

以上符号表示的是词义 C 由语素 A 和 B 语义合成而来，但是 A 和 B 这两个语素各自都有多个义位（义项），那么其组合结果就有 $n \times m$ 多种可能，这样计算机就无法知道正确的组合。我们的做法是不但对词义加载语境义、情景义，相反要对现有词典义项“减负”，“抽象”。

通过类义的抽象将上述组合改进为： $A_0+B\{B_1, \dots, B_n\}$ project, $j=1, \dots, k$ 。（project

投影，j 代表足码）通常约简为 $A_0+B\{B_1, \dots, B_i, (I \leq \min(m, k))\}$

其结果：削减叠置中的歧义数。

抽象类义的理论意义在于词汇义项选择的尝试方法和改进释义。因为词汇义项注释、判定本身是一大难题，抽象类义的方法在于主张 AB 组合中 A 的义项选择不决定自身，取决于 B，应选择将 B 作为坐标参数。

3 抽象类义理论依据、定义、操作原则和方法

3.1 抽象的理论依据

词的意义与词的运用或功能是两个不同的概念，词义是使用语言的集体语言心理的共同“约定俗成”产物，词的基本义即概念义是相对固定的信息内容。相对稳定的意义核心才使得词典编写成为可能，词典中词的释义就是脱离具体运用的稳定意义。如“女人”这个词，其基本的概念义“女性的成年人”，在不同环境中可表示一个对象，如“这是他的女人（女人=妻子）”；可表示一个类，如“女人不得入内”；也可以表性质“她毕竟是女人”；也可以表示各种感情、评价意义“女人嘛，能有什么好主意”。词的用法和功能与词的意义有内在的联系，受其制约，同时又与词义是两种不同性质的现象。如果将词的用法和功能都归进词义，认为词义就等于这些用法和功能，那么“意义”概念本身就含混不清。目前不少词典都将词的用法列为义项，所以说从词的多个义项中提取抽象类义应该成为可能。

3.2 类义抽象的定义

抽象类义不同于词典中的本义，本义往往是以物质性、物理性的外延指称义来定义的，引申义是由本义联想、引申来的。“抽象‘类义’是在本义、引申义的中舍弃了个别的、非本质的属性，抽出共同的、本质的属性，最终提出本义、引申义的共同‘因子’”（陆汝占 2002）。因为抽象类义是抽象的，所以它属于高层次或深层的语义，这种抽象义只有和具体场景和上下文中才会实现为表层的、实在的意义。比如后文的“关”抽象类义是“使不通过”，只有“关”和其他的词素或词组成词或短语时，其抽象义才会实现为具体意义。我们知道“心”是一个场所，“关”和“心”结合，即“把某物关进‘心’里”就是“使某物或某事不通过‘心’这一场所”。“关心”也就是“关注、留意、留心”的意思。

类义是抽取一个词不同义项（义位）的共同义素或语义特征。它是抽象的元语言，抽象类义不同于传统的几个词共同的义类如“液体类、食品类”等，也不等同于本义，不等同于语素义。类义抽象是一个过程，抽象类义则是类义抽象的结果。一个义位就等于“抽象的类义+环境义素”，一个义位就是一个概念，作为一个概念就有内涵特征，外延指称。概念和

概念的结合有内在的理据——内涵特征。

3. 3 抽象类义的操作原则

1 尽可能多地摘录著名权威词典、字典的释义。

2 枚举所有义项，目的是涵盖意义，使得类义具有完整性。

3 排斥无法解释的义项——干扰项，如一些借代用法、理据不清的义项。

4 排斥不合理、不合格的义项——择善而从。如：词典中解释“木”时，列出“棺木”的义项，其实这个义项只是出现在“行将就木”这个成语中，因此不如在词表中加入“行将就木”；在语义词典中删去“棺木”这个义项。

5 按字提表，不能按类提表。即对每个字的语义抽象类义进行提取，并填制出相应的特征表，而不是按照语义类制定特征表。但要总结一类词的相同语义组合的方式和关系。

6 压缩码要解码，不能简单替换。如跑百米，其中“百米”不能用“十米”替换。“跑第一棒”中“棒”不能用“棍子”替换。字面是“字”，底层是一个句子，不能随意替换。两个语素合成词，是两个小句——压缩的小句。有时连接和添加的部分都省略，要理解就要加标记，必需展开。

7 要注意词典中义项的使用率，如果只是偶尔出现的某一义项，则可略去。其组合词语项列入词表。

3. 4 抽象类义操作方法

抽象就是抽出单音节汉字的类义，进而建立汉语语义基本单位和构造规则，只要是语义可计算的就不列入词库，可计算就是符合叠置原理。

1 对一个词的多个义项进行归并，然后进行类义抽象。

2 根据抽象类义出现的环境或投影组合义。由A求B，也可由B求A。

3 利用统计的方法先搜集整理有关复合词的语料，然后分类。

4 确定中心词及中心词与成分之间的语义关系。

5 确定中心词与附加成分的具体语义关系，分析产生上述关系语义上的原因。

6 建立语义特征的组合模型。

7 在句子中验证所描述名词语义特征正确与否。

验证就是通过已知的成分义和结构义推知一个复合词语的意义，从而验证所设参数和规则是否正确即考察召回律的情况。这是不同于先知道复合词语的整体意义，再推定两个成分的意义和结构关系义，从而得出成分义与结构整体意义的理据关系的方法。前者是部分到整体，后者是从整体到部分，前者是演绎，后者是分析。这是两个不同的研究方向和路线。

3. 5 “关”的类义抽象及相关复合词义的求解

3. 5. 1 对相关辞典解释进行收集

《现代汉语词典》第三版对“关”的解释：关（1）使开着的物体合拢：～窗户|把抽屉～上。（2）使机器等停止运转；使电器装置结束工作状态：～机|～灯|～电视。（3）放在里面不使出来：鸟儿在～笼子里|监狱是～犯人的。（4）（企业等）倒闭；歇业：有一年，镇上～

了好几家店铺。(5)古代在交通险要的地方或边境出入的地方设置的守卫处所:~口~防山海~|嘉峪~。我的职责就是不让废品混过~去。(6)城门外附近的地区:城~北~|~厢。(7)门栓:门插~儿|斩~落锁。(8)‘关上’的简称。(9)货物出口和入口收税地方:海~|~税。(10)比喻重要的转折点或不容易度过的一段时间:难~|只要突破这一~,就好办了。(11)起转折关联作用的部分:机~|~节|~键。(12)牵连;关系:这些见解很~重要|此事与他无~|交有~部门去处理。(13)发放或领取(工资)~饷。(14)(Guān)姓。

《高级汉语大辞典》对“关”的解释:关:(形声。从门,𠔁(gu)声。金文作“門”,象门里有门闩形,本象形字。小篆繁化,变成了形声字。本义:门闩)

3.5.2 对辞典的义项的归并

由以上辞典解释可知“关”的语义发展过程,可对其进行归并。

本义:门闩(插关儿/门插关)儿 → (功能:不使通过)关闭 关上 关押 关张

开关 关门

(具有阻隔的作用的东西) 牙关 关节; 城关 机关 海关 边关 鬼门关 难关 雄关 大关; 关键 关山 关隘 关塞 关卡 关碍 关防 关厢; 关口 年关 关头 关子 关书; 关东 关外 关内 关中; 关说 关税

(放在心里) 关心 关意(留意) 关怀 关注 关照 关爱 关切

(起连接作用) 关系 有关 关于 双关 无关 交关 相关 关联 关涉 关乎

3.5.3 抽象义提取, 操作命令形式化

关: 本义——门闩 → 抽象类义——功能: [闭]=不使 x 通过 y)

门闩的功能	关闭
关闭的对象	门
起关闭作用的场所或机构	海关
起关闭作用的人体部分	关节
关闭在心	关心
不使通过(阻止)的对象	关犯人
具有关闭连接作用	关系

g 抽象义=不使 x 通过 y, x={x1, x2, x3, x4}, y={场所/机构}

g. 1.0 使开的 y 闭合/ v。

g. 1.1 功能提升, 切换状态——使运行的 y {机构} 处于终止状态/v。

g. 1.2 摹状: 使开的 y 闭合的东西。即 LxV(x) (命题、动词 → 实体) “门栓/n”。

g. 1.3 隐喻: 起阻隔作用的地方/n。

g. 1.4 转喻: 阻隔的地方功能就是连接和联系 n/v。

(“L”读埃欧塔, 表示“具有 v 作用的”, “→”表示引中)

“关”的抽象义： $(-V_{\text{使通过}}(x, y))$ 即“不使 x 通过 y 。” x =对象 $x \in \{x_1, x_2, x_3\}$ y =场所 $\in \{y_1, y_2, y_3, y_4\}$

由以上词典义项的归并可知“关”具有的五个投影或相位(g1.0, g1.1, g1.2, g1.3, g1.4) 投影 1=g1.0

当“关”的后字是 $\{y_1, y_2\}$ 或 $\{x_1, x_2, x_3, x_4\}$ 时，“关”取“不使 x 通过 y ”义。

$y_1 \in$ 建筑部分： $\{门、窗、闸\} \rightarrow y_2 \in$ 房屋： $\{监狱、笼子、房间\} \rightarrow x_1 \in$ 人/动物： $\{犯人、鸟、孩子、\dots\}$ $x_2 \in$ 电力： $\{动力、电流\} \rightarrow y_3 \in$ 机器： $\{机、电、车、灯\}$ 。

$y_2 \in$ 心： $\{心、怀、意\} \vee V_{\text{关}}(x, y) \wedge V_{\text{在}}(x, \text{心})$ 。关爱=关心+爱护，关切=关心+切合，关注=关心+注意，关照=关心+照顾，这些都是复合命题。

词义的引申反映了人的认知规律，是一种层层转喻或转指关系。人们看到[部分]联想到它的[整体]，由[容器]联想到[内容]，由[具体物]联想到[抽象物]，再由[内容]联想到[容器]，这都符合人的认知规律。“关”作用的对象变化过程就是： y_1 “门” $\rightarrow y_2$ “房屋” $\rightarrow x_1$ “犯人” $\rightarrow x_2$ “动力” $\rightarrow x_5$ “机器”。

投影 2=g1.1 $LxV(x)$

当“关”与 $\{x_5, x_6, x_7, x_8\}$ 组合时，“关”的语义是“起阻隔作用的地方/n”

$x_5 \in$ 起阻隔作用的场所： $\{口、隘、卡、塞\}$ ，“关”与 x_5 组合时， x_5 中的词与“关”同义，“关”与 x_5 组合义取<同>。

$x_6 \in$ 场所： $\{海、城、边、鬼、门\}$ ，“关”与后字 x_6 组合，组合义取<加>。

$x_7 \in$ 方位： $\{东、中、内、外\}$ ，“关”与后字 x_7 组合，组合义取<加>。

$x_8 \in$ 部件： $\{头、键、机\}$ ，“关”与 x_8 组合，组合义取<同>。

“关税、牙关、关子、年关”等组合是非频繁结构可以单独列入词表。从句法信息中我们可以知道“税、牙、年”都属名词，“子”是名词词缀，所以可知以上的词都是名词。“关税”的语义可以从“税”的内涵特征中算出。（“关”的其他投影的计算略）

从内涵特征的角度解释，就是一个实体或抽象类义有多个特征或投射角度，当与其他词语项组合时就会取相应的不同特征，或者说使某一个特征得到突显，从而实现为义位，也就是多义词在上下文中使抽象的类义得到具体的实现，从而得到一个具体的概念，在某一可能世界中指称某一实体。“关”与后字“门”结合，就有“使 X 不通过门”的语义。

参 考 文 献

- [1] 范可育：“从‘生词熟字说’看词义和构词语素义的关系”，《语言文字应用》1993年第一期。
- [2] 利奇：《语义学》，李瑞华等译，上海外语教育出版社，1987年。
- [3] 吕叔湘：《汉语语法分析问题》，商务印书馆，1979年。
- [4] 沈家煊：“转指和转喻”，《当代语言学》，1999年（1）
- [5] 苏宝荣：“汉语语素组合关系与辞书释义”，《辞书研究》1999年第四期。
- [6] 苏宝荣：“词义研究与汉语的‘语法—语义结构’”，《语言教学与研究》，2002年第一期。
- [7] 王宁：《训诂学原理》，中国国际广播出版社，1996年
- [8] 徐通锵：《语言论——语义型语言的结构原理和研究方法》，东北师范大学出版社，1997年