

HNC 理论的五元组与词性¹²

唐兴全

北京语言文化大学文化学院 100083

E-mail:tang_xq@263.net

摘要: HNC 理论认为, 抽象概念具有复杂的多元性表现, 可以概括为五元性, 用五元组表示。五元组是词性的基元, 它以基元和基元复合的方式表达抽象概念的多元性, 从语言深层阐释并解决了词性问题。HNC 的概念矩阵与同行优先准则比词性搭配规则对词语搭配的约束更加充分而必要。五元组是语句中的概念联想脉络激活的重要工具。

关键词: HNC 搭配 抽象概念 基元 词性 五元组

The Quintuple of HNC theory and part of speech

Tang Xingquan

Language Information Processing Center, Beijing Language and Culture University, Beijing 100083

E-mail:tang_xq@263.net

ABSTRACT: HNC theory thinks that abstract concepts have complicated multiplicity, which was generalized as Quintuple. The Quintuple, which is the primitive of part of speech, shows the multiplicity of abstract concepts by primitives and compound primitives. It elaborates and solves the problem of part of speech on semantic level. Compared with the rule of collocation of part of speech, HNC's matrix of concepts and the principle of same-row-privilege are more sufficient and necessary to bind the collocation of phrases. The Quintuple is an important tool to activate conceptive association vein.

Keywords: HNC, Collocation, Abstract concept, Primitive, Part of speech, Quintuple

1 引言

中科院声学所黄曾阳先生创立的面向整个自然语言理解的理论框架 HNC (Hierarchical Networks of Concepts 的缩写) 理论, 以语义网络、五元组和概念组合结构建立了自然语言词汇层面的知识表述体系。五元组为 HNC 理论独创, 形式表示为: $\{v, g, u, z, r\}$, 专门用来描述抽象概念, 是抽象概念五个侧面的表现特征。

¹ 本文得到 973 项目“G1998030506”的支持。

² 本文在写作过程中得到苗传江老师的指导与帮助, 在此表示作者诚挚的谢意。

2 抽象概念的多元性表现

任何一个概念都需要从不同的侧面、不同的角度予以表达，我们将这一现象称为概念的多元性表现。具体概念的多元性十分复杂，比较难以给出规范化的表述。抽象概念则有所不同，其多元性表现在自然语言中有明显的迹象，这就是词性现象。（黄曾阳《HNC（概念层次网络）理论》第20页，以下简称《专著》）

印欧语言中的词有很多形态变化，其中一种形态变化是，词根或具有词根特色的词，可以加上不同的后缀分别构成名词、动词、形容词和副词，如：wonderful、wonderfully、development、important、importance 以及 realize、really、reality 等都是词根 wonder、develop、real 分别加上不同的后缀 ful、ment、ance、ize、ly、ity 等构成的词。HNC 理论认为，这种词根相同词性不同的词是对同一概念不同侧面的表达，也就是说，它们是语义网络中“同一”（相同或近似）概念节点在语言中的反映，只是具体表现有所不同，有表达动态的如 develop，有表达静态的如 importance，有表达属性的如 wonderful，这是抽象概念多元性的生动表现。汉语缺乏形态变化，因此，抽象概念的多元性表现在汉语的词语上则不像印欧语言那样有相应的形式标志，而是同一个表达抽象概念的词往往具有名词、动词、形容词和副词中的两个或多个特征，即每一个抽象概念节点（意味着词的基本含义相同）所对应的词有可能同时从两个或多个方面描述属性，这就是汉语中词的“兼类”现象。例如“繁荣”“完善”都同时具有动词、形容词的词性，“代表”“爱好”兼有名词和动词的词性，“及时”“均衡”兼有形容词和副词的词性，“真实”兼有形容词、副词和名词的属性等，这同样是抽象概念多元性的生动表现。

印欧语言并没有把词根方式全部规范化，其中也有兼类，如“report”“interest”都兼作名词与动词；汉语也并非一律采用兼类，也有词根方式，如词尾“然”是形容词的标志，“性”是最常见的名词词尾，词尾“化”是动词的标志等。但无论如何变化，它们都是抽象概念多元性的表现形式，词根就意味着兼类，兼类说明有“词根”，仅此而已。可见，词根也好，兼类也罢，从中我们很容易做出抽象概念多元性的判断与概括。

3 五元组是词性的基元

抽象概念的多元性表现的“多”是一个模糊的“多”，还是一个确定的“多”？或者说，这种多元性表现是否存在某些基元呢？答案是肯定的，HNC 把抽象概念的多元性概括为五个侧面：动态、静态、属性、值和效应，提出了五元组，即{v, g, u, z, r}。HNC 理论认为，五元组是表达任一抽象概念的五个基本侧面，是描述一个抽象概念的充要条件，换句话说，任何一个抽象概念都需要从这五个方面来表达。五元组是抽象概念外在表现的基元，称之为抽象概念的五元性。从多元性到五元性，是一种概括，也是 HNC 独有的发现。在 HNC 的知识表示体系中，抽象概念的外在表现——五元组符号，与抽象概念的内涵表示——概念层次网络符号，组成了词语的 HNC 符号。其中，五元组，如果用语言学的说法，是抽象

概念的形态，因此，五元组是抽象概念的形态类别基元。

五元组里的动态、静态和属性（v, g, u）来源于语法学，其中 v 是动态表示，g 是静态表示，u 是属性表示。值 z 来源于明斯基（Minsky）的框架理论，表达概念可量化计算的性质。r 表示效应，来源于 HNC 理论的“作用效应链”思想。

v 是动态表示。在语言表层上，大体相当于语言学上的动词。其中，可以充当句子的语义核心，表达句子行为的动词，即述语动词，对句子的句类起决定作用。概念有高、中、底层之分，一些高层的 v 概念不一定能充当述语动词，如“善于”“试图”“予以”“给以”等。

对效应 r 这一侧面的发现是 HNC 的新贡献。有作用必有效应，每个抽象概念都有相应的效应，这正是“作用效应链”的思想（《专著》第 29 页）。g 与 r 都是静态表达的名词，分别代表因果两极，区分二者，有助于概念的局部联想。黄曾阳先生在其专著的论文系列 Paper1 中详细阐述了 g 与 r 的区别。我们看下面的几个例子：

成功 (vg930a1)	业绩 (r30a1)
楷模 (jg737)	表率 (jr737)
行为 (gd0)	威望 (rd001)
竞争 (gvb30)	胜利 (rvb30)
心思 (g714)	表情 (r714)

从概念内涵上说，“成功”与“业绩”的内涵是相同的，差别只在于外在形态表现。但从词性上说，它们都是名词，但仅用“名词”这个概念显然不能表示它们之间的差别。上面的五对词语都是这种情况。引入类别符号 r，就能清楚指明：业绩是成功的效应，表率是楷模的效应，威望是行为的效应，胜利是竞争的效应，表情是心思的效应。可见，r 在概念表达中具有 g 不可替代的作用。而已有的各种现代汉语词性分类系统均未注意到这一点，原因在于这些分类系统都没有深入到语义层面，进行概念局部联想的考虑。

某一抽象概念节点应具有自己特有的属性特征，u 就是这一属性特征的表示。u 概念类同语言学中的形容词和副词，但不区分二者。形容词作状语时仍是形容词还是已变成了副词？我们认为，这种讨论只是对词语位置的简单命名，是“依句辨品”论的必然结果。u 概念的本质是，它既是 v 类概念、又是 g、r、z 类概念的属性说明，如下面的例子：

非法 (ua59)	深刻 ((u811; u8108) +ju60c44)
重要 (jgu725)	美丽 (u51+j831)
坚定 (vu72015)	完善 (vu500511)
再次 (luva0100a)	陆续 (juv008)
非常 (juu60c43)	完全 (juu501)

这里以“深刻”为例说明：

这件事给克林顿留下深刻的印象。

每件东西都令人印象深刻。

会议要求全党深刻地理解科技是第一生产力的马克思主义观点。

方言无法表达深刻的思想。

上述四个句子中“深刻”一词的位置不同，句法功能不同，所修饰的词词性也不同，但“深刻”只有一个 HNC 符号(u811;u8108)+ju60c44，这里只不过是修饰 g、v 还是 r 的不同。无论在什么位置，它都是 8 行概念的属性说明（当然，这种说明也适用于与概念节点 811 和 8108 有交式关联的概念）。

作用、效应、属性应有其值，因此提出 z 概念。z 概念用以表达概念的可计算性。对 z 概念，语言学上也多以名词命名之。如“速度(z10090)”、“损耗(vz3b2)”、“体积(xjz20-)”和“质量(jz51)”等。可见，HNC 分语言学上的名词为 g, r, z，对名词在语义层面上进行了全新的分类。g, r, z 的区分是必须的和天然的，是概念层面基本特征的必然表现，将“力量(gz00)”、“力度(z00)”、“力(g008)”和“功能(r00)”一概称之为名词，其区别何在？

现代计算语言学为适应工程需要，已经对现代汉语的词性作了十分精细的划分，有的多达 100 多类。各类有相应的形式标记。五元组是抽象概念外在表现的基元，是对抽象概念多元性的概括，那么 HNC 理论是怎样表示抽象概念复杂的多元性表现的呢？HNC 把抽象概念外在形态表现概括为五元组，由它们可构成形态复合基元，构成方式就是五元组字母的连用。五元组符号可以多级连用，这时表示的是词性的兼类。如 gz, gv, vr, rv, vuz 等等。五元组复合的表示，在字母排列顺序上优先性强者在前，弱者在后。如用 gv 表示兼有 g 与 v 的词性，优先 g，如“叛乱(gvd002c33)”；用 vr 表示兼有 v 与 r 的词性，优先 v，如“破灭(vr352)”；用 rz 表示兼有 r 与 z 的词性，优先 r，如“效益(rz30)”。

出于工程的需要，HNC 还设计了一系列组合符号。在五元组的设计里，形容词和副词的概念是通过组合来体现的，gu 和 ug 都含有形容词的词性，如“中等(gu56c23)”和“访问(v9201,v2199)”；uu, uv 和 vu 三者都表示含有副词的词性，区别在于，uu 表示一般副词，uv 表示专用于修饰动词的副词，vu 表示兼有动词和修饰词（形容词或副词）的词性，如“非常(juu60c43)”、“反复(uv100a)”和“抽象(jvu746)”；zz 和 zzv 都含有量词的词性，如“公斤(zzj518)”和“来回(zzv22b0)”；vv 表示一定位于 v 前充当 EQ 的动词，如“善于(v72108)”。各种组合都有不同的意义，是对句法功能的体现，但本质不变。uu 的本质还是 u，vv 的本质也还是 v。在词语知识库的建设中，可用“HNC 符号”和“概念类别”两栏配合使用。概念的多元性表现一律放在知识库框架的“概念类别”一栏里。概念本身的 HNC 符号表示式只使用类别基元或下面要介绍的类别复合基元如 gv、gz 等。例如“领导”一词兼有名词和动词的词性，还可以是“领导者”的简称。这时 HNC 符号可表示为“vc441”，而在“概念类别”中，则可以写成“p, v+g+ug”。

长期以来，汉语词性问题的研究一直存在诸多争论，建立在印欧语形态变化基础上的词性迟迟无法落实于汉语。那么，词性问题的根本是什么？HNC 理论的五元组从语言深层——语义层面阐释并解决了词性问题。我们认为，汉语词性研究的根本问题是未区分抽象概念与具体概念，把对抽象概念与具体概念的描述一概而论。词性问题的难点是对抽象概念的表达，描述具体概念的词是没有兼类现象的。“词性”的本质内容是抽象概念的五元组特性在自然语言词汇上的体现，是描述抽象概念的必要而不充分条件，兼类是抽象概念基本特征的必然表现。任何概念都具有五元组特性，即都需要从五个侧面加以表达，不过，对某个

抽象概念各个侧面的表达，自然语言中未必有相应的语词，而且不同语种之间也存在个性差异，如汉语中的“兴趣”无动态表达，而英语中“interest”则兼作名词与动词。反过来，自然语言中的一个表达抽象概念的语词必定是从五元组的某个或某几个侧面来表达某个抽象概念。如作用（gv00）、功能（r00）、力（g008）和力度（z00）就是分别从gv、r、g和z几个侧面对同一概念加以描述的。五元组仅用五个字母，以基元和基元复合的方式表现了抽象概念形态表现的复杂的多元性。所以说，五元组是词性的本质内容，是词性的基元。

4 概念矩阵与同行优先

五元组的思想深化了语法学上的词性搭配关联性（《专著》第319页）。自然语言的一大缺陷是，意义相关的词语形式上却往往无共同之处。例如，“轮船”与“航行”、“斗志”和“昂扬”等大量概念上具有天然联系的词，形式上却毫不相干，表达它们之间的关联性成为一个大问题，而这种关联性对于计算机的自然语言理解又是至关重要的。HNC是如何解决这个问题的呢？

HNC用五元组和语义网络分别表达抽象概念的外在表现和内涵。抽象概念表达的五个基本侧面是自然语言概念关联性的基本表现之一（另外还有概念的中层特性即对比性、对偶性和包含性表示）。对于内涵或内核相同或相似的概念，表达这些概念的自然语言符号可以有很多，但我们都赋予它们相同或相似的层次符号串，即数字串。从概念矩阵上，它们属于概念矩阵的同一行。这样，以概念节点的五元组属性为列，以概念节点的层次符号为行形成概念矩阵。这一矩阵有效表示出概念节点之间的关联关系。一部分语义距离的计算问题就变成了对数字串的逐层比较问题，有利于计算机把握相同或相似的层次内涵（相同或相似的层次网络符号）所反映的不同词语之间的关联性（即搭配的可信性），这就是“广义同行优先”准则。所谓“同行优先”准则，是对层次网络符号天然属性的一种简明陈述，即：同行的五元组概念及挂靠的（w，p）类概念优先相互搭配。如：同行的（u，z）与（v，g，r）构成最优先的偏正搭配；同行的v与（g，r）构成最优先的述宾搭配；同行的（g，r）与v构成最优先的主谓搭配等。下面举例说明：

迅速 u1009c21	发展 gv10a
圆满 u30a	完成 v930a8
召开 v3959+va01	会议 gc3959
思考 v810	问题 gr53821
客轮 pw22b	启程 (v111,v22b)

语法学对词语的搭配给出了词性约束规则，比如形容词与名词可以搭配、副词与动词和形容词可以搭配、数词与量词可以搭配，而且给出了词性搭配的条件。但是，词性约束的规则仍然太宽泛，因为词语搭配不仅是语法问题，而且是语义问题。无论是“牛吃草”还是“草吃牛”所体现的都是“名+动+名”的词性搭配，而后者显然是不合逻辑的。概念矩阵与同行优先，在某种意义上是对上述词性约束规则的扩展与深化。扩展表现在它力图表述语义块之间或语句要素之间的约束，深化表现在它力图尽可能给出条件（如交式关联、同行信息）。

5 概念激活与词性标注

词性标注是指对在文本中出现的词的词性进行标注。词性标注是对词在句子中的位置的命名，是语言表层的。五元组是概念的基本特征，是语言深层的。

语言学上，曾对句子中动词出现在“的”后时的词性有过讨论，比如“这本书的出版”“这个作品的发表”，其中的“出版”与“发表”应是什么词性呢？有人认为这里的“出版”是动名词，或者称为动词的“名物化”。类似的还有形名词或者形容词的“名物化”的说法。HNC理论认为，这种追究徒具形式，而无本质意义，“理解的根本问题是把握概念之间的关联性，主要是语义问题”（《专著》第283页）。无论“出版”在“的”前还是“的”后，“这本书”是“出版”的效应对象的这一概念关联性的本质不变，也就是说，它们之间的概念联想关系没有变化。因此，这里的“出版”“发表”仍然是v概念。基于处理的需要，HNC提出了句蜕的思想，像上面的“的”加动词的两个句子是典型的v要素句蜕。

我们这里引入黄曾阳先生在《句类分析的20项难点》中的一段论述，以期使读者对上段的理解更深一步，

“句蜕的概念是对子句和‘句子成分与词性对应’这两个经典概念的极为重要的发展。……句子成分与词性对应的概念，只适用于简单构成的语义块，不适用于复杂构成的语义块。语义块的本质是复合的。主语、宾语、定语、状语中出现动词，谓语中出现名词，是语言的正常表现或现象。然而，这一语义块的正常本质现象长期以来却被‘句子成分与词性对应’的经典概念掩盖了”³

同样，定中结构的中心语的位置上可以有表达v概念的词语，并且不会因为它在“的”后充当中心语而改变其概念类别。

大脑的感知应该是对概念之间相互关联性的把握。如果说，西语是以形态变化来表现词的句法功能的话，汉语就是以词义组合的方式来探求它的句法功能。HNC理论的方法是，以HNC符号的主体——层次网络符号——作为概念的激活点，以五元组作为局部概念联想激活的重要工具，使句子分析能很快绕过词性这些表层现象，直接进入语义层面，通过概念关联来探求它的句法功能。HNC理论认为，人对语言的理解实质上是一个对概念的激活、联想、浓缩、存储的过程，体现在句子分析中就是通过概念激活语句的深层结构，即句类。HNC的句类是以它在概念层面的知识表述体系为基础，HNC的自然语言词汇层面的表述模式以作用效应链来分类，作用效应链同时是HNC句类的划分标准，因此，概念与句子成分之间有一定的对应关系，通过概念的激活就可以得到句类，从而也就抓住了语句的深层结构。这种概念激活的方式彻底解决了汉语“词无定类”，“类无定职”的困扰。

参考文献

- [1] 黄曾阳《HNC（概念层次网络）理论——计算机理解语言的新思路》清华大学出版社1998
- [2] 苗传江《简论黄曾阳先生创立的HNC理论》《中文信息学报》1998

³ 黄曾阳《句类分析的20项难点》第22页