

# 在反思中前进，在碰撞中成长\*

**摘要** 本文是《HNC(概念层次网络)理论》的续篇，概述了HNC理论探索近五年来的基本成果，对HNC的某些传统论述进行了反思与改正，同时给出了HNC未来15年探索的素描。本文正式将HNC的探索目标定位于交互引擎的研发，将语言概念空间描述的原五模式说改为概念基元空间、句类空间、语境单元空间和语境空间的四模式说，给出了这四个模式的数学表示式；将交互引擎的基本模块凝练成句类分析、语境单元萃取和语境生成的三部曲或三个基本环节。为达到交互引擎的彼岸，本文阐释了这一科学征途的三项理论工程、四项技术工程和HNC研发平台的概貌。为说明这一彼岸的理论可行性本文诠释了HNC的两个基本假设和“概念无限而概念基元有限、语句无限而句类有限、语境无限而语境单元有限”的基本论点。

## 1 引言

HNC理论基本构架的形成已经整整10年了。

HNC这10年的历程是一个幸运的历程。

小托尔斯泰有一部不朽名著叫《苦难的历程》，该名著三部曲的名称是《两姐妹》、《一九一八年》和《阴暗的早晨》。HNC的幸运历程也有一个三部曲，可类比定名为《两仙翁》（按：指语法与逻辑或语言

本文发表在《第二届HNC与语言学研讨会论文集》，海洋出版社，2004年。

学与哲学)《一九九七年》和《明媚的晨曦》。

小托尔斯泰完成那“苦难历程”的写作时间是在小说描述事件的20年之后，如果以这个时间为参照那么 HNC 幸运历程的描述放到2023年以后比较适当，是今天座中年轻才俊进入“知天命”或“从心所欲”的时候，你们到那个时候看着办吧。

但是我今天还是要说一点，HNC 幸运历程的最大幸运是什么？那就是遇到了众多的伯乐。由于近年我成了一个经常阅读圣经的无神论者所以他们成了我心中的科学而非神学的弥赛亚。现在让我们以热烈的掌声向这些科学弥赛亚表示我们的无限敬意与感谢。有两位科学弥赛亚——陈力为院士和林杏光教授已经离开我们了，让我们以肃穆的沉思表示我们永恒的缅怀。

中国历史最辉煌的年代是唐朝，唐朝最伟大的文豪兼哲人是韩退之 愈 先生 退之先生最富于哲理的名篇之一是《进学解》，《进学解》里最著名的警句之一是“业精于勤 荒于嬉 行成于思 毁于随。”从某种意义上说本文是学习《进学解》的一点体会。

我的体会之一是退之先生说的“思”就是哲学意义的反思，“随”就是现代意义的拒绝创新与改革。“随”的对偶概念是“撞”无“思”无“撞”就不能前进与成长。所以本文定名为《在反思中前进，在碰撞中成长》。

世俗意义下的反思仅对内而不对外，碰撞仅对外而不对内。本文的反思和碰撞则兼对内外，且都以对内为主。

不过本文并不直接以反思与碰撞为中心展开论述，而是把两者融合到关于自然语言之电脑理解的以下三个基本问题的阐释里。这三个基本问题是：

(1) 为什么已经具有许多奇妙智能的电脑到今天还不能理解自然语言？这里的根本障碍到底是什么？怎样克服这一障碍？

(2) 假定电脑能够理解自然语言，那么让它成为一个“一目千行，过目不忘”的超级才子 语言超人 存在什么障碍 怎样克服这一障碍

(3) 假定电脑能够理解自然语言，那必然有一个从低级到高级的成长过程，那么这一过程主要阶段或基本步调是什么？我们怎样推动这些阶段或步调的实现？

围绕着第一个问题的阐释本文将着重介绍 HNC 理论规划中的三项理论工程，简称三部《手册》——《HNC 概念基元符号体系手册》、《句类知识手册》和《语境单元知识手册》。

围绕着第二个问题的阐释本文将简要介绍 HNC 技术规划中的四项技术工程——句类分析技术 SCA(Sentences Category Analysis)、语境单元萃取技术 SGUE(Sentences Group Unit Extraction)、语境生成技术 ABS(ABSTRACT) 和翻译引擎技术。

围绕着第三个问题的阐释本文将简略说明 HNC 的一项特定计算工程——计算机自然语言理解度的测试与学习平台，这个平台的首要目标是进行电脑理解度的量化与类化研究，将简称 HNC 研发平台。

如果把这三项阐释所范定的研究目标的实现比做一次重大战役，那么第一项阐释关系到该战役的战略方针的确定，第二项阐释关系到主战场或主攻方向的选择，第三项阐释关系到关键信息的搜集与分析、先进武器的供应与保证。毋庸论证这三者对于赢得一场战役的胜利具有决定性意义。

HNC 通过语言概念空间研究语言现象。语言概念空间是存在于人类大脑之中的一个符号体系，这个符号体系既是人类进行语言思维的载体，又是人类进行语言交际的引擎，这个符号体系将简称交际引擎。计算机要获得理解自然语言的能力就必须也拥有一台在功能上类似的引擎，否则计算机就永远不可能成为名副其实的电脑。这个引擎将叫做交互引擎。交互引擎是交际引擎的仿制，仿制就必然存在简化与不足。交际引擎由千亿数量级的神经元构成，交互引擎将由当前电脑的二进制数字符号构成。HNC 的基本定位就是催生交互引擎的问世，这是天方夜谭式的神话么？HNC 不持这种悲观态度。HNC 已经看到如果能够实施上述三项理论工程、四项技术工

程及其特定计算工程，则完全有可能在 15 年左右的时间里使交互引擎在中华大地诞生。

## 2 关于交互引擎的三项理论工程——三部《手册》

本章是本文的重点。将分六节进行论述。2.1 节对 HNC 探索的重要先行者作简要回顾；2.2 节介绍 HNC 理论的基本定位及其两个基本假设；2.3 节论述 HNC 探索预定的三部《手册》，不仅是三项理论探索，而且是三项巨大的理论工程；2.4 节是本章的第一重点，介绍第一部手册——《HNC 概念基元符号体系手册》的要点，重点阐释了 HNC 概念基元符号体系设计五项基本原则的延伸原则；2.5 节介绍第二部手册——《句类知识手册》的要点，对句类概念的 HNC 传统论述中的失误作了系统的反思；2.6 节是本章的第二重点，提出了交际语境和交互语境的概念，阐释了交互引擎研发的基本思路与框架，对 HNC 理论体系的整体描述进行了反思，将 HNC 理论体系的原五模式说改成四层级与三提升说，列举了语境单元萃取和语境生成研究的基本课题，概述了 HNC 理论体系四组表示式——(HNC1)、(HNC2)、(HNC3) 和 (HNC4) 的物理意义。

### 2.1 对 HNC 先驱者的简短回顾——交互引擎的背景说明 1

已经具有许多奇妙智能的电脑到今天还不能理解自然语言，为什么？这个问题的答案很复杂，但是又似乎很简单，有一个简单答案是这样的：要理解人类语言就必须拥有世界知识，而世界知识浩瀚无垠，需要软件工程师在语言学家的配合下一点一滴地教给电脑，这个工程太浩大了。面对这一浩瀚无垠的困难，人们思考着并实际采取过许多对策，这些对策可以概括成三种基本类型：最大简化策、愚公移山策和最大似然策。

最大简化策的基本思路是：设定一个受限环境，让这个环境的世

界知识十分有限，且语言表达十分规范，此策的代表人物之一是维诺格拉德教授。

愚公移山策的基本思路是：将世界知识划分出一定的类型和层次，用谓词逻辑形式或结构化数据形式写出来交给电脑，前者的代表人物是美国 CYC 计划的主持者 Lenat 教授，后者则代表人物众多。

最大似然策当前最走红，领唱者甚多，自称语料库学派。其核心思想是把言语当做一个随机过程，依托大规模真实语料制造一个计算机的进化环境，让计算机在这个环境里以大体类似于宇宙进化的方式自行进化。

前两种对策基本以所谓理性主义为依托，第三种对策基本以所谓经验主义为依托。

三种对策的基本思路朴素而简明，都具有朴素性的固有力量。

但是对于“已经具有许多奇妙智能的电脑到今天还不能理解自然语言”这样重大的科学问题仅仅依靠朴素性的固有力量是不够的，需要进行多侧面多层次的反思。我们至少需要反思以下三个基本问题：第一为什么开始只有极少世界知识的 1~5 岁正常儿童都是学习自然语言的天才？第二为什么儿童在完成这一学习的过程中既不需要学习语法课程，也不需要学习逻辑课程？第三那高深而且越来越高深的语法学和逻辑学与儿童在潜移默化中获得的基本语法和逻辑知识是否存在某些奇妙的区别？

有两位美国人——乔姆斯基先生和山克先生在 HNC 之前进行过这一反思，还有一位欧洲人维特根斯坦先生在电脑尚未出现之前就进行过这一反思，并为此写了《哲学研究》这一不朽巨著。

乔姆斯基先生对上列三个基本问题的前半作过刺猬式反思，山克先生则对三者都进行过狐狸式反思。乔先生的答案是：儿童的语言天赋是由于人类大脑里存在一个经过百万年进化而形成的东西，他把这个东西叫做普适语法 (UG)。随后他用毕生精力研究这个东西，在 20 多年间写下了参考文献中列出的八部专著<sup>[1~8]</sup>。山克先

生的答案是：语言就是用概念基元写出的脚本，语法不过是一件人为的装饰品。这位先生以典型的狐狸研究风格搞了一种概念基元和若干个小脚本之后就转向其他相关领域了。

上列三位先行者的创新都体现了西方文化的优势特征：熟谙分析与归纳的逻辑。但是理论创新还需要综合与演绎的逻辑，这四位最高理性法官携手合作才能做出重大的创新，近代和现代物理学的重大创新都是四位理性法官携手合作的光辉典范，而我国的两部古老名著——《老子》和《孙子兵法》则是古代的光辉典范。

一个极为常见的认识误区是这四位最高理性法官需要一位老大。这个老大观念在东方和西方文化里都具有挥之不去的魔性。实际上这四位最高理性法官需要轮流坐庄，不能由某一位理性法官永远充当老大，就像桥牌游戏的东南西北四家不能由某一家老是坐庄一样。在 HNC 的形成过程中演绎起过特殊重要的作用，但绝不是永远的老大，可是 HNC 经常流露出演绎老大的意识，这是首先需要反思的。因为 HNC 已经进入不再是演绎坐庄的新时期了，八项工程的实施都需要四位最高理性法官的通力合作。

## 2.2 HNC 的两个基本假设与电脑未能理解自然语言的症结——交互引擎的背景说明 2

自然语言理解是一个从自然语言空间到语言概念空间的映射过程，两个空间各有自己的一套符号体系。语言交际过程、即交际引擎的运作过程实质上是这两种符号体系相互映射的过程。说者 // <sup>①</sup> 写者将语言概念空间的符号映射成语言空间的符号，这是语言生成过程 听者 // 读者将语言空间的符号映射成语言概念空间的符号，这是语言理解过程。

语言空间符号体系千差万别，据说当今世界上还存在 6 000 种

<sup>①</sup> 符号 // 表示“和”或者“或”的情况 全文同。

之多，但语言概念空间符号体系只有一个，全人类的语言概念空间具有同一性。这是一个假设，是 HNC 理论的第一假设。这个假设一定会引起质疑，因为不同文化背景和不同民族的语言概念空间必然存在差异。这里的质疑者不了解理论探索的基本特征，那就是必须有所舍弃——舍弃枝节，抓住要害，没有舍弃就不会形成任何理论。这就是黑格尔先生所说“哲学的开端就是一个假设”的真谛。拙著<sup>[9]</sup>里写了一段不符合科学论文规矩的话，但由于关系到 HNC 第一假设是否合理这一根本问题，这里仍然全文引录下来供本文读者思考。

关于这个模型问题，可以说存在两种态度，一是得过且过，在短语结构模型的基础上修修补补，不去触动它的根本缺陷，希求通过受限的约束避开语言的种种不规范现象，也就是避开对语言本质的探索。另一种是乔姆斯基关于自然语言是一个 ill-defined 的东西的说法，脑子里存在大量比喻的和夸张的、乡土的和诗歌的、儿童的和怪诞的例句，并为之困扰而不知自拔，不相信对自然语言的表述可以出现牛顿力学对力学现象或麦克斯韦方程对电磁现象的突破，但是他们不曾想过，如果当年牛顿不是专注于天体的运动，而是专注于羽毛在狂风中的飞舞，麦克斯韦不是专注于电磁场在自由空间中的一般规律，而是专注于方孔的衍射，他们也将一事无成。在建立自然语言模型这一重大探索中，必须谨记有所为和有所不为的辩证法，并深思康德的下列两段名言：

理性必须一手拿着原则，拿着那些惟一能使符合一致的现象成为法则的原则，另一手拿着自己按照那些原则设计的实验，走向自然，去向自然请教，但不是以小学生的身份，老师爱讲什么就听什么，而是以法官的身份，强迫证人回答他所提出的问题。

自然的最高立法必须是在我们心中，即在我们的知性之中，而且我们必须不是通过经验，在自然里面去寻求自然的普遍法则；而是反过来，根据自然的普遍合法则性，在存在于我们的感性和知性里面的经验可能性条件中去寻求自然。

HNC第一假设关系到对语言本体的认识，也关系到对“自然语言理解”的理解。依据这一假设语言的本体就不是语法，而是语言概念空间。语言概念空间就是拙著<sup>[9]</sup>里所说的概念联想脉络或语义网络。乔姆斯基先生把他本人精辟阐释的那个人类百万年进化的大脑产物叫做普适语法并不妥当，因为语法这个概念还是停留在语言空间，没有升华到语言概念空间。语言空间和语言概念空间是相互依托但又毕竟具有本质区别的两类空间，语言空间依托于语音和文字而存在，语言概念空间依托于概念联想脉络而存在。语音、文字、概念联想脉络都是符号体系，语音符号体系具有可听性（现代技术还能使之可见、可存储、可复现）文字符号体系具有可见、可存储、可复现性，概念联想脉络符号体系既不可听，亦不可见。脑科学研究的现代技术虽然正在朝着使之可观测、可分析的方向努力，但远没有达到像基因复制或蛋白质合成机制研究的技术水平，目前还得主要依靠演绎式探索“鸣锣开道”。

对 HNC第一假设需要作两点重要补充，两者是对下述两个基本问题的回答。

第一语言概念空间是概念空间的全部么？回答是：否！语言概念空间只是概念空间的一个子空间。概念空间的其他子空间还有形象概念空间、情感概念空间、艺术概念空间、科学概念空间等。

第二语言概念空间具有历时性么？回答是：可以忽略！当然某些概念子空间具有非常鲜明的历时性特征。如科学概念空间在以牛顿和爱因斯坦为标志的先后三个历史时期就具有天壤之别，人类社会物质文明的发展主要依靠这一“天壤之别”的出现。但语言概念空间的主体是稳定的，其历时性远不像语言空间那样突出，仅有某些局部呈现出极为缓慢的历时性。

在 HNC 第一假设的基础上，还需要引入第二假设——关于语言概念空间结构特征的假设。这一结构特征必须假设，因为如同上面所指出的脑科学的当前研究水平还不能提供足够的启发性知识。

HNC的第二假设是：语言概念空间是一个四层级——基层、第一介层、第二介层和上层——的结构体。基层对应着语言概念基元符号体系，也称概念基元空间；第一介层对应着句类符号体系，也称句类空间；第二介层对应着语境单元符号体系，也称语境单元空间；上层对应着语境符号体系，也称语境空间。这四层级符号体系及其对应的世界知识将分别简称概念基元世界、句类世界、语境单元世界和语境世界。空间对应着相应科学对象的数学描述，世界对应着相应科学对象的物理描述。这里有一个新术语——语境单元，这里的语境也有它特定的意义，都将在 2.6 节说明。

这个四层级概念空间和概念世界共同构成语言思维（不是全部思维）的载体，这个语言思维载体的名称就是概念层次网络，简称 HNC。为什么使用这么一个人们难以习惯的奇特名称呢？因为语言概念空间的基本特征是它的层次性和网络性。为了便于与传统语言学接轨，我们曾把概念层次网络叫做语义网络；为了便于一般读者的理解，我们也曾把它叫做概念联想脉络。

HNC第二假设隐含着下述三项“无限与有限”的假设，这就是：“语言概念无限而语言概念基元有限，语句无限而语句的概念类型（句类）有限，语境无限而语境单元有限”。

如果在上述两项基本假设的基础上仅探讨概念层次网络的基本理论特征，那就是狐狸式研究，如果锲而不舍地去具体构造出概念层次网络的四层级符号体系，那就是刺猬式研究。显然概念层次网络的研究需要这两种研究方式的紧密配合，刺猬需要狐狸的智慧，而狐狸需要刺猬的执著。

语言概念空间的上述四层级符号体系一旦构造出来，而且该符号体系能够直接用二进制数字符号加以表示，那么交际引擎的仿制（即交互引擎的研制）就不再是一个炼金术的幻想，而是一项意义深远的科学工程了。

结论是一清如泉的：仿制交际引擎的关键举措就是首先把体现

语言思维载体的上述四层级符号体系装进电脑，并在电脑中形成一个四层级概念世界。这个载体当然有低级、中级与高级之分。人类大脑里存在着该载体的最完善、最高级形态，而现在的电脑连该载体的最简陋雏形都没有，这就是电脑至今不能理解自然语言的症结所在。HNC从1998年开始试图向电脑装入一个最低级的语言思维载体，2001年基本完成，2002年显示出初步效果，但其喜人之处不在于这些效果本身，而在于他//她所预示的成长潜力。今天晚上晋耀红博士将向大家展示这一婴儿的憨态，明天晚上苗传江博士将向大家介绍这一婴儿的知识结构，张全博士则将介绍对该婴儿的托儿所教育设想。

### 2.3 三部《手册》是一项巨大的理论工程 而不只是一个理论体系

为了叙述的方便下面将把《HNC概念基元符号体系手册》、《句类知识手册》和《语境单元知识手册》分别简称《手册1》、《手册2》和《手册3》。《手册1》将描述概念基元符号体系及其相应的世界知识；《手册2》将描述句类符号体系及其相应的世界知识；《手册3》将描述语境单元符号体系及相应的世界知识。三部《手册》所描述的符号体系及其相应的世界知识相当于概念基元世界、句类世界和语境单元世界的“数字地图”。

概念基元世界大体对应语言空间的词语，句类世界大体对应语言空间的语句，语境单元世界大体对应语言空间的句群。这里的对应不可能是严格的，因为概念基元、句类和语境单元这三个世界所拥有的自身符号体系并非与语言世界的符号体系一一对应，而且对应性的具体表现与具体语言世界（语种）密切相关。“大体对应”是一个极为粗略的描述，确切的描述是三部《手册》的重要内容之一。

我在文献10中曾用“7-2-1”、“57-3192”和“7-57-7”这三组数字对上述基层与介层概念世界的宏观特性作了一个最高层次的概括。这里首先需要对这三组数字进行反思，本文将把这组数字改成下面

的描述：“8-2-1”、“2-8-57-3192”和“10-57-7”。下面三节将分别说明对这三组数字变动的反思。

这里需要着重指出的是：三部《手册》不只是对基层与介层概念世界的理论阐释，还要描述对应于概念基元世界的“典型”词语、对应于句类世界的“典型”例句、对应于语境单元世界的“典型”句群描述捆绑于每一概念符号的世界知识。《手册 1》将重点描述捆绑于每一延伸概念的世界知识，《手册 2》将重点描述捆绑于每一基本句类和常用混合句类的世界知识，《手册 3》将重点描述捆绑于每一领域句类的世界知识。

上列“典型”的含义是指典型语种，不只是汉语，还包括其他重要语言，首先是英语。对基层与介层概念世界的世界知识捆绑不能只是自然语言的叙述，还必须使之符号化和结构化，形成相应的数据库和知识库。

由此可见三部《手册》不是单纯的理论性专著，而是三项巨大的理论工程。

## 2.4 概念基元符号体系手册——《手册 1》

本节将分三小节进行论述，2.4.1 节论述概念基元世界的基本特性，2.4.2 节论述《手册 1》的理论攻关，2.4.3 节论述《手册 1》的工程使命。

### 2.4.1 概念基元世界的基本特性

概念基元世界的基本特性可以浓缩成一句话来描述，那就是概念范畴的有限性。HNC 把这句话作为一项公理来看待，其整个理论体系就是建立在这一公理的基础上。上述“三无限、三有限”假设实质上是这一公理的推论，其中“概念无限而概念基元有限”的假设更是这一公理的直接推论。

这里的概念基元当然是指语言概念的基元，指概念子空间之一的语言概念空间的全部概念基元，但并不囊括其他概念子空间的概

念基元。下文都省略“语言”二字，上文已使用了这一省略。这一申明是必要的，否则容易造成一些毫无意义的误解。

概念范畴的有限性是公理，公理不需要也不能证明。但体现这一公理的概念基元有限性需要通过穷举方式加以描述，这样才能为交互引擎提供一个可操作的基础符号体系。

HNC 首先将概念分成三种基本范畴：抽象概念、具体概念和两可概念，后者兼备抽象与具体的双重特性。概念基本范畴的这种划分必然会引起质疑：概念本来就是抽象的，何来抽象与具体之分？质疑者的第一句话没有错，但质疑者的“何来”就有点冒失了。抽象本身具有不同的类型与层次，概念的三种基本范畴正是概念抽象的三种基本类型。所谓具体概念是指对物 // 人的直接抽象，例如植物和动物是对两类生命体的直接抽象，人是对一种特殊动物的直接抽象，这些概念的外延或所指联系于一定的具体“东西”——物 // 人。抽象概念不同，它不是对物 // 人的直接抽象，而是对物 // 人之表现 // 特性的抽象。例如作用、过程、转移、效应、关系、状态这六个概念就不直接联系于一定的具体东西——物 // 人，因为任何东西都具有这六方面的表现，如果说这些概念的外延或所指的话那么可以说它们既无所不延，又无所不指。这类概念就是 HNC 所定义的抽象概念之一。交际引擎最基本的联想就是把这类抽象概念与上面定义的具体概念区别开来，交互引擎当然也需要这一基本区分。

上面说到的具有“无所不延、无所不指”特性（可简称“两无”特性）的六个概念包含万事万物，是最大的“两无”，它们将构成概念基元符号体系的核心，并命名为主体基元概念。但多数抽象概念并不具有这一最大“两无”特性而仅具有一定条件（范畴）的“两无”特性，例如描述人与社会共性的概念就属于这一类，人必有情感、思维与理念等，社会必有政治、经济与文化等。这些概念可以与具体东西相联系，但又与那些直接联系于具体东西的具体概念有所不同。HNC 把这类概念也纳入抽象概念，并命名为扩展基元概念。

概念基元空间的数字描述“8-2-1”就是指八类抽象概念、两类具体概念和一类两可概念。

八类抽象概念的名称、符号和意义如表 1 所示。

表 1 抽象概念基元总表

编号	名称	映射符号	类型符号	意义
1	主体基元概念 (作用效应链)	(0-5)y	$\varphi$	描述万事万物的基本规律
2	第一类扩展基元概念	(71,72,8,a,b,d)yy	$\psi//\varphi$	描述历时性较弱的人类活动
3	第二类扩展基元概念	(q6-q8)yy	q// $\varphi$	描述历时性较强的人类活动
4	基本概念	(j0-j8)y	j	描述哲学所关注的基本课题
5	语言逻辑概念	(l0-lb)y	l	描述语言生成所必需的逻辑约定
6	语习概念	(f0-fb)y	f	描述语言生成所必需的非逻辑约定
7	基本逻辑概念	(jl0-jl1)y	jl	描述作为一切判断基本前提的逻辑
8	综合概念	(s1-s4)yy	s	描述兼有基元概念、基本概念和逻辑概念综合特征的概念

这一总表体现了下列四项反思：

(1) 被统称为基元概念的三类概念基元 ( primitive ) 仍被统一赋予类型符号  $\varphi$  ( $\varphi$  也可单用于主体基元概念的描述)，但第一和第二类扩展基元概念又被另行赋予类型符号  $\varphi$  和  $q$ 。约定类型符号  $\varphi$  和  $\psi$  都一律不进入映射符号，但  $q$  可以进入，也可以被数字符号  $6//9//c$  替换。符号  $q$  以前曾使用过符号  $\lambda$  和  $y$ ，这里正式宣告作废。

(2) 基元概念虽然可以分为主体与扩展两类，但分为三类更恰当，这就是将概念基元世界的总体性数字描述从“7-2-1”改成“8-2-1”的缘故。

(3) 概念的高层层次数只有 2 与 3 之分。主体基元概念  $\varphi$ 、基本

概念 j; 语言逻辑概念 l; 语习概念 f 和基本逻辑概念 jl 的高层层次数都是 2 第一类扩展基元概念  $\psi$  第二类扩展基元概念 q 和综合概念 s 的高层层次数都是 3。以前有关“ $\psi$ 7 特殊、高层为 4”的说法是错误的, 因为  $\psi$ 71 和  $\psi$ 72 实质上是两个特殊符号, 分别表示心理活动和精神状态, 不是通常意义下的层次延伸。心理活动与精神状态密不可分, 但又具有本质区别。两者必须放在一起, 但又不宜设置两者的共性概念节点  $\psi$ 70。以前用心理活动的复杂性来解释这一符号表示的特殊性, 乃思考不够透彻的典型表现。

(4) 调整了八类概念基元的排列顺序, 将语习概念 f 从第八位调整到第六位, 紧随语言逻辑概念 l 以后 这样综合概念 s 就顺理成章地成为殿后者。这一排序更为协调。

两类具体概念的定义是:

(1) 基本物 jw 描述宇宙构成的基本物 并以地球构成的基本物为主, 这里的基本物不仅指物质, 还包括所有自然之物, 并将生命体和人都纳入自然之物。

(2) 挂靠物 w//p 这里符号 w 表示物, p 表示人, 统称挂靠物。这两个符号可形成各种组合形式, 如 pw、gw、rw 等等, p 还有 p-、pe、pj1\*、pj2\* 等特殊符号约定 (pe 的 e 是数字 14 不是字母) 其意义可参看文献 [11] 以前将 w//p 统称挂靠具体概念, 当然无可非议, 今后也使用挂靠物统称之。

两类具体概念的基本区别是: 基本物具有自身的数字符号体系, 而挂靠物不具有, 其数字符号体系直接依附于抽象概念, 故“挂靠者依附也”。这里需要再次强调的一点是, 挂靠物这一概念的引入有利于具体概念与抽象概念之间相互联想的激活。

两可概念就是物性, 以符号 x 表示。物性 x 就是物之属性的简称。凡带有 w (包括 jw) 的具体概念符号如果将 w 替换成 x 那么该符号就变成相应物的物性表示。例如生命体有雌雄之分, 其符号是 jw6ye2m, jw6ye21 表示雄性生命体 jw6ye22 表示雌性生命体。将上

面的符号变成  $jx6ye2m$  ,它就代表“雄”与“雌”的概念。如果将其中的高层变量符号  $y$  用数字  $3$  替换,它就代表“男性”和“女性”的概念。物性这一概念本质上都具有两可性,它既具体依附于某一类特定的物,又必然具有游离于该特定物之外的抽象意义,这是所有物性概念的共性。

对概念基元世界的“8-2-1”描述概括了语言概念空间基层的基本特性:抽象概念可穷举为八类;具体概念可穷举为两类——基本物  $jw$  和挂靠物  $w//p$ ;两可概念则仅有物性  $x$  这一种类型。《手册 1》将通过它的理论和工程描述证实这些论点,下面给出《手册 1》各章的篇名,从它可以窥见概念基元世界所面临的基本科学问题。

第一章	主体基元概念
第二章	第一类扩展基元概念
第三章	第二类扩展基元概念
第四章	基本概念
第五章	语言逻辑概念和基本逻辑概念
第六章	语习概念
第七章	综合概念
第八章	具体概念与两可概念
第九章	HNC 符号体系
第十章	概念关联性表达
第十一章	抽象概念的五元组特性和语言的词类
第十二章	映射库与反映射库
第十三章	概念知识、语言知识与世界知识
第十四章	语义块、句类与语境单元

#### 2.4.2 《手册 1》的理论攻坚——实践延伸原则

我在文献 [12]里论述过 HNC 概念基元符号体系设计的四项基本原则:尔雅原则、语境原则、关联原则和延伸原则。这一论述存在一个不可原谅的疏忽,那就是忘了句类原则。这就是说 HNC 概念基

元符号体系设计实际上遵循着五项、而不是四项基本原则，那就是：尔雅原则、句类原则、语境原则、关联原则和延伸原则。在文献 [ 9 ] 里对这五项基本原则都有所阐释，但分散在不同的处所。因为该书是在特殊情况下出版的一本“散文”集，而不是一部专著，这在该书的“后记”里有明确交代。

文献 [ 12 ] 还指出延伸原则是 HNC 理论的灵魂。因为概念联想脉络网络两项基本特性（层次性与网络性）的表达难易程度有很大差异，层次性比较简单，网络性极为复杂。后者主要依靠延伸原则来体现，延伸原则的实践就是《手册 1》的理论攻坚。

三部《手册》都存在攻坚课题但《手册 1》的攻坚具有奠基意义，因为句类世界的根主要在主体基元概念（作用效应链），句类原则即导源于此；语境世界的根主要在扩展基元概念，语境原则即导源于此。

延伸原则怎样体现概念联想脉络的网络特性？一句话靠符号的再抽象原则。“符号再抽象原则”是什么意思？那就是用一个数字来表达概念的复杂组合性，文献 [ 10 ] 曾借用“节日”这个概念演示过符号再抽象原则的具体运用。

为什么符号再抽象原则就能体现概念联想脉络的网络特性？因为所谓网络性主要指概念之间的显式关联性，这种显式关联性总可以通过各项概念之间的逻辑组合加以表达。但这一表示式通常都十分冗长（如“节日”所示），不利于概念关联性的高效计算，而再抽象的简化表示反而有利于这一计算，再抽象的实质是通过隐去某些东西以便更有效地凸现另外一些，这一阐释借鉴了心理学关于“注意”的著名论点——通过“舍掉某些东西以便更有效地处理另外一些” [ 13 ]。HNC 猜想交际引擎的最大奥秘也许就在于它对符号再抽象原则的充分运用，而不仅是认知心理学所描述的并行运算 [ 14 ]。至于交互引擎它必须也只能主要依靠符号再抽象原则以获得概念联想脉络的快速激活能力。

延伸原则实质上就是符号再抽象原则的并行与交替使用。而延伸原则的可实现性又取决于概念基元的有限性。从这个意义上说《手册 1》的理论攻坚也可以理解为对“概念无限而概念基元有限”这一基本假定的具体求证，下文将把这一求证简称 1 号求证。

上面的抽象概念基元总表（表 1）穷举了抽象概念的有限范畴，1 号求证必须从这里起步。换句话说表 1 就是 1 号求证过程的基础或第一步，接下来的第二步就是把这些范畴进一步分类，并把基本物也包括进来，从而形成表 2 所示的概念树总览。

表 2 概念树总览

概念林名称	概念树符号	概念树类型数
作用效应链	0-5	6
心理活动及精神状态	710-714, 720-722	8
思维活动	80-84	5
专业及追求活动	a0-a8, b0-b4	14
理念活动	d0-d2	3
第一类劳动	q60-q64	5
业余活动	q70-q74	5
信仰活动	q80-q85	6
基本概念	j0-j8	9
语言逻辑概念	l0-lb	12
语习概念	f1-fb	11
基本逻辑概念	jl0, jl1	2
综合概念	s1-s4	4//15
基本物	jw0-jw6	7
		$\Sigma = 97 // 108$

如果说数字“8-2-1”是对概念范畴的穷举，表 1 是对抽象概念范畴的穷举，那么表 2 就是对概念树的穷举。这里应该指出两点：第