

HNC 理论全书

第三卷 语言概念空间总论 — 第五册

论语言概念空间的总体结构 图灵脑理论基础之五

黄曾阳 / 著



科学出版社

HNC 理论全书

第三卷 语言概念空间总论 — (第五册)

论语言概念空间的总体结构 图灵脑理论基础之五

黄曾阳 / 著

科学出版社

北京

内 容 简 介

本书是《HNC 理论全书》的第五册。HNC 理论以自然语言理解为其核心探索目标,试图为语言理解的探索开启一条新的途径,以语言概念空间的符号化、形式化为手段,实现人类语言脑的纯物理模拟。

本书围绕“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、显记忆无限而隐记忆有限”这 4 项 HNC 理论的基本定理,集中论述语言概念空间的概念基元、句类、语境单元和记忆四级结构,并对语言信息处理中的具体问题作了深入剖析。

本书适合对自然语言理解、人工智能、认知科学、脑科学、语言学等感兴趣的读者,特别适合语言信息处理方面的研究者及学生参阅。

图书在版编目(CIP)数据

论语言概念空间的总体结构 / 黄曾阳著. —北京: 科学出版社, 2015. 2

(HNC 理论全书)

ISBN 978-7-03-043161-5

I. ①论… II. ①黄… III. ①系统科学—研究 IV. ①N94

中国版本图书馆 CIP 数据核字 (2015) 第 018890 号

责任编辑: 付 艳 王昌凤 / 责任校对: 朱光兰

责任印制: 肖 兴 / 封面设计: 黄华斌

联系电话: 010-6403 3934

电子邮箱: fuyan@mail.sciencep.com

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

中国科学院印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2015 年 2 月第 一 版 开本: 787×1092 1/16

2015 年 2 月第一次印刷 印张: 32 1/2

字数: 725 000

定价: 138.00 元

(如有印装质量问题, 我社负责调换)

本书得到下述项目资助

国家高技术研究发展计划（“863”计划）、“十二五”项目课题“基于云计算的海量文本语义计算框架与开放域自动问答验证系统”（2012AA011104）

作者的话

本书是《HNC 理论全书》的第五册。《全书》共三卷六册，第一卷三册，第二卷一册，第三卷二册。第五册也就是第三卷的第一册。

HNC 理论以自然语言理解为其核心探索目标，试图为语言理解的探索开启一条新的途径。HNC 认为：语言理解的奥秘，是大脑之谜的核心，也是意识之谜的核心。这个谜团的探索不是当前的生命科学可以独立完成的，需要哲学和神学的参与。故《全书》之“全”是一个“三学（科学、哲学与神学）协力”的同义词，非 HNC 理论自身之“全”也。HNC 理论充其量是一名语言理解新探索的侦察兵，从这个意义上说，《全书》之“全”应看作是一种期待，一声呼唤。

《全书》的初稿是半成品，是 HNC 团队的内部读物。原定以十年（2006~2015 年）为期，完成初稿。不意十年未竟，推动者和出版者联袂而至。他们深谋远虑，要把一个半成品升级为成品，把一个内部读物正式出版，其间所展现出来的非凡胆识、灼见与谋划，居功至伟四字，不足以表达笔者心中感受之万一。

《全书》结构庞大，体例繁杂，带大量注释。结构方面，分上层与下层，上层分“卷、编、章”3 级，以汉字表示顺序，汉字“零”表示共相概念林或共相概念树。在某些编与章之间，还插入篇。下层分“节、小节、子节”3 级，子节之后，可延伸出次节，每级之内，可派生出分节。体例方面，主体文字之外，安置了大量预说和呼应。注释方面，分两类编号：数字与字母。前者是对正文本身的注释，后者是对正文背景的注释。数字和字母都放在方括号内，如[*01]和[*a]。其中的星号“*”可以多个，如同宾馆等级的标记。两星[**]以上的注释比较重要，表示读者应即时阅读。

《全书》常相互引用，为标记之便，采取了[k₁k₂k₃-m]简化表示，其中的“k₁”表示卷，“k₂”表示编，“k₃”表示篇，无篇取“0”。“m”也是一个数字序列，依次表示章节序号。例如[210-0.2.1]和[210-1.2.1]分别表示第二卷第一编第零章和第一章的第 2 节第 1 小节。

《全书》使用了大量概念关联式。概念关联式是语言理解基因的重要组成部分，也是隐记忆的重要组成部分。每一个概念关联式总是联系于特定的概念基元、句类或语境单元。概念关联式分为无编号与有编号两类，无编号的表示尚待探索，《全书》只是给出了若干示范，为上下文引用方便给出的临时数字编号也属于这一类；有编号的统一使用“——（编号）”或“——[编号]”，前者表示内使，后者表示外使。概念关联式编号

区分普通与重要两级，后者加“-0”区别，若“-0”插入编号中间，表示不同文明对此有共识，而后缀于编号，则表示特定的文明视野。有编号的概念关联式都有牵头符号，代表着该概念关联式的重要性级别，目前主要用[HNC1]符号牵头。

在撰写初稿期间，池毓焕博士一直是我的学术助手。在本书出版期间，池博士一直是我个人的全权代表。科学出版社以付艳、王昌凤编辑为主的有关同志，为初稿的升级付出了巨大的辛勤与智慧，其审校之精细，无与伦比；池博士的配合，力求尽善。笔者的钦佩与感激之情，难以言表。

老子曰：“天地万物生于有，有生于无。”伟哉斯言。

黄曾阳

2014年9月22日于北京

引文出处缩略语对照表

《理论》/ 《HNC 理论》	黄曾阳. HNC (概念层次网络) 理论[M]. 北京: 清华大学出版社, 1998
《定理》	黄曾阳. 语言概念空间的基本定理和数学物理表示式[M]. 北京: 海洋出版社, 2004
《全书》	即本丛书——HNC 理论全书, 共有三卷六册, 各册书名如下: 第一卷 第一册 论语言概念空间的主体概念基元及其基本呈现 第二册 论语言概念空间的主体语境基元 第三册 论语言概念空间的基础语境基元 第二卷 第四册 论语言概念空间的基础概念基元 第三卷 第五册 论语言概念空间的总体结构 第六册 论图灵脑技术实现之路
《导论》/《苗 著》/《HNC 理 论导论》	苗传江. HNC (概念层次网络) 理论导论[M]. 北京: 清华大学出版社, 2005
《转换》	张克亮. 面向机器翻译的汉英句类及句式转换[M]. 郑州: 河南大学出版社, 2007
《变换》	李颖, 王侃, 池毓焕. 面向汉英机器翻译的语义块构成变换[M]. 北京: 科学出版社, 2009
《现汉》	中国社会科学院语言研究所词典编辑室. 现代汉语词典(第3版)[M]. 北京: 商务印书馆, 1996
《现范》	李行健. 现代汉语规范词典[M]. 北京: 外语教学与研究出版社/语文出版社, 2004

第三卷 卷首语

本卷是本《全书》的最后一卷，阐释 HNC 理论的总体构架。笔者很希望撰写方式能有所改变，以减少给读者带来的不便，但惯性的力量或许将使这一希望化为泡影。

为什么要搞这样一个 HNC 理论？为了揭秘语言脑，为了开发语言超人（简称“语超”）。

语超将是一个纯物理系统，不涉及化学和生命科学，仅仅是对人类语言脑^[*01]的计算机模拟。但语超将具有比人类语言脑远为强大的功能，因为它可以轻而易举地做到“一目万千行，过目不忘”，是传统中华文明意义上才子（可“一目十行，过目不忘”）的超级形态，不是智力、体力或机械意义上的超人，而是一种语言超人。

语超可以有各种各样的类型，最简单的语超将是未来人类的得力助手。当然，语超必然是一把双刃剑，既可以成为人类的朋友，也可能成为人类的敌人。这是一个与本卷有关的重大话题，但将完全回避；大脑奥秘的探索应该以语言脑为突破口，而不是像当下流行的大脑研究那样，对“大脑的五大构成”不细心加以区分或无视这一区分，这也是一个与本卷有关的重大话题，也将基本回避。本卷仅致力于回答这样一个科学话题：人类语言脑的纯物理模拟可以实现么？

HNC 理论把这个科学话题分解成下列四项：

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；

显记忆无限而隐记忆有限。

这是 HNC 理论的四项基本命题。如果这四命题是正确或基本正确的，那么语超就有了比较坚实的理论基础，它的实现就不应该存在不可逾越的技术障碍。语超就是图灵脑，而图灵脑的诞生将不仅是科学的大事，也是哲学和神学的大事，关于意识与存在之间的无穷论争至少可以因此而告别一个阶段了。

这四命题也直接关系到对大脑之谜的探索，因为四者的初级技术转化产品将类似于一个婴儿型的语言超人，简称微超。微超的研发曾被命名为“图灵之战”，后续的语超

研发曾被命名为“垓下之战”，这两个命名今后可能还会使用，因为它们意味着一种憧憬。微超本身不单是语超的过渡或雏形产品，也是大脑探索的有力模拟工具，故曾被冠以语言脑“电子白老鼠”的美名。

在 HNC 理论探索的历程中，《理论》阶段仅涉及前两项与三项命题，那时，前两项是明确的，但第三项是懵懂的；《定理》阶段才扩展到全部四项命题，第三项趋于明朗，但第四项依然比较懵懂，趋于明朗乃《定理》以后的事。这些命题初期仅名之假设，后期则名之公理。假设与公理有天壤之别，但请不要太较劲，先看作是揭秘语言脑的需要吧。

这四项话题将构成本卷前四编的名称：第一编，论概念基元；第二编，论句类；第三编，论语境单元；第四编，论记忆。后续四编的名称依次是：第五编，论机器翻译；第六编，微超论；第七编，语超论；第八编，展望未来。

就是说，本卷共计八编，前五编“论”字在前，接着两编“论”字在后，为什么要作这样的区分？第六编会作交代。

第一卷曾采用“编—篇—章—节—小节”的五级结构，第二卷无“篇”，采用“编—章—节—小节”的四级结构，本卷也无“篇”，主要采用“编—章—节—小节”的结构，但也有采用“编—章—节”之简化结构和“编—章—节—小节—分节”之扩展结构的情况。

本卷的许多要点，在前两卷有不同程度的预说。基于以防万一的考虑，有些预说接近正式论述，这使得本卷的相应论述沦为一种呼应性的东西。对于这种情况，将尽可能以注释的方式予以说明。

注释

【*01】语言脑是人类大脑进化历程的第四次演变，是人类正式诞生的基本标志。此前的三次演变分别是图像脑、情感脑和艺术脑的相继诞生，此后的一次演变是科技脑的诞生。这些演变的基础是生理脑。图像脑、情感脑、艺术脑、语言脑和科技脑简称“五脑”，也称“大脑的五大构成”。在大脑里，它们各自占有自己的区块吗？答案似乎是否定的，否则，脑科学早就应该发现了。但是，尚未发现并不等于要否定，因为脑科学的研究可以说是“生不逢时”，一方面不得不跟随心理学的深厚形而下传统，另一方面又得不到形而上思考的滋润，后者的社会背景前文有过多次论述。

当然，图像脑、情感脑、艺术脑、语言脑和科技脑的各自区块绝不可能像“地球村”的五大洲那样边界分明，“五脑”之间必然呈现出比“犬牙交错”更为复杂的景象，这是大脑迷宫无与伦比的独特性。这一独特性对“五脑”各自独立存在的态势必将产生巨大的掩盖或混淆作用，而不能成为否定其存在性的证据。

关于图像脑、情感脑、艺术脑、语言脑和科技脑乃相继诞生并各自独立的证据已经太多了，有志于此的读者完全有条件写出一本精彩的科普读物。这里要告诉这位未来作者：HNC理论仅关注其中的语言脑。考虑到语言脑与情感脑的交织性最为密切，本《全书》在第一类精神生活的论述中（见第一卷第二编）对情感现象给予了一定关注。

目 录 | contents

作者的话

引文出处缩略语对照表

第三卷 卷首语

第一编 论概念基元 1

第一章 概念基元总论 3

第1节 概念基元有限性的形而上思考 5

第2节 概念基元的抽象与具体两分 15

第3节 抽象概念基元的范畴三分 18

第4节 基元、基本和逻辑概念的“6、2、4”分 23

第5节 广义作用效应链的“二生三四” 28

第6节 语境基元 29

第二章 五元组 31

第1节 五元组与词类的异同概述 33

第2节 五元组主要是对基元概念形态世界的描述 37

第3节 五元组的组合特性 39

第4节 五元组特定元素的强存在性 44

第三章 概念树 49

第1节 概念树内容世界回顾 51

第2节 概念树形态世界素描 53

第四章 概念延伸结构表示式 61

第1节 概念延伸结构表示式简介 63

第2节 横向延伸的齐备性探求 91

第3节 纵向延伸的透彻性探求 91

第4节 挂靠的哲理描述 92

第五章 概念关联式	93
第1节 概念关联式的类型描述	95
第2节 概念关联式的编码及标记	97
第六章 语言理解基因	99
第二编 论句类	103
第一章 语块与句类	105
第1节 语块类型与要素	107
第2节 主块构成	113
第3节 基本句类与广义作用效应链（基础概念）	116
第4节 基本句类表示式与句类代码	120
第5节 混合句类	124
第6节 句蜕与块扩	132
第二章 广义作用句与格式	139
第1节 回顾与反思	141
第2节 广义作用句的类型描述	147
第3节 格式与块扩是广义作用句的专利	152
第4节 格式与自然语言的分类	155
第三章 广义效应句与样式	157
第1节 回顾与反思	159
第2节 广义效应句的类型描述	167
第3节 样式的此山景象	169
第四章 句类空间与句类知识	173
第1节 句类空间与概念基元空间的对应性	175
第2节 句类空间类似于一座城市的地理描述	180
第3节 主体句类知识就是句类表示式与块码 关联式的联姻	182
第4节 句类知识类似于一座城市的社会人文描述	187
第三编 论语境单元	199
第一章 语境、领域与语境单元	203
第1节 语境与语境基元	205

第 2 节	领域 DOM	218
第 3 节	语境单元 SGU	222
第 4 节	语境单元 SGU 是领域 DOM 的函数	232
第二章	领域句类与语言理解基因 (SCD, LUG)	251
第 1 节	领域句类 SCD 与情景 SIT	253
第 2 节	语言理解基因 LUG 与领域意识 DOMC	269
第三章	浅说领域认定与语境分析 (DD; SGA; BACE+BACA)	305
第四章	浅说语境空间与领域知识 (SGS, DOMC)	317
第四编	论记忆	323
第一章	记忆与领域	327
第 1 节	同一领域的记忆分置于不同位置是不可思议的	329
第 2 节	不同领域记忆区块之间的信息传递	333
第 3 节	不同领域记忆区块之间的缓冲区	339
第二章	记忆与作用效应链 (ABS, XY)	343
第 1 节	作用与效应侧面的记忆 以殊相记忆为主——(XYD, XYN)	346
第 2 节	过程与转移、关系与状态侧面的记忆以共相 记忆为主——(PT, RS)	353
第 3 节	广义作用效应链之判断侧面在记忆中的角色	369
第三章	记忆与对象内容 (ABS, BC)	395
第 1 节	具体概念的记忆唤起	397
第 2 节	抽象概念的记忆唤起	400
第 3 节	记忆索引小议	411
第四章	动态记忆 DM 与记忆接口 (I/O)M 浅说	413
第 1 节	动态记忆 DM 浅说	415
第 2 节	记忆输入接口 IM 浅说	422
第 3 节	记忆输出接口 OM 浅说	474
第五章	广义记忆 (MEM) 杂谈	477
第 1 节	关于广义记忆的补充说明	479

第2节 记忆特区杂谈	480
第3节 拷贝记忆杂谈	487
术语索引	489
人名索引	495
《HNC理论全书》总目	501
第1章 绪论	501
第2章 语言与思维	501
第3章 语言与心理	501
第4章 语言与社会	501
第5章 语言与艺术	501
第6章 语言与科学	501
第7章 语言与宗教	501
第8章 语言与政治	501
第9章 语言与法律	501
第10章 语言与医学	501
第11章 语言与军事	501
第12章 语言与外交	501
第13章 语言与教育	501
第14章 语言与文学	501
第15章 语言与历史	501
第16章 语言与地理	501
第17章 语言与气候	501
第18章 语言与生物	501
第19章 语言与化学	501
第20章 语言与物理	501
第21章 语言与天文	501
第22章 语言与地文	501
第23章 语言与气象	501
第24章 语言与海洋	501
第25章 语言与大气	501
第26章 语言与空间	501
第27章 语言与时间	501
第28章 语言与质量	501
第29章 语言与数量	501
第30章 语言与温度	501
第31章 语言与湿度	501
第32章 语言与风速	501
第33章 语言与气压	501
第34章 语言与降水	501
第35章 语言与日照	501
第36章 语言与霜冻	501
第37章 语言与降雪	501
第38章 语言与结冰	501
第39章 语言与融化	501
第40章 语言与蒸发	501
第41章 语言与凝结	501
第42章 语言与降水	501
第43章 语言与日照	501
第44章 语言与霜冻	501
第45章 语言与降雪	501
第46章 语言与结冰	501
第47章 语言与融化	501
第48章 语言与蒸发	501
第49章 语言与凝结	501
第50章 语言与降水	501

第一编

论概念基元

编首语

本编分 6 章，各章编号、汉语命名和相关 HNC 符号如下：

第一章	概念基元总论	CP
第二章	五元组	(v,g,u,z,r)
第三章	概念树	CT
第四章	概念延伸结构表示式	(CT:(OD//ED) :)
第五章	概念关联式	(CPm,L,CPn)
第六章	语言理解基因	LUG

其中，HNC 符号对应的英语、汉语如下：

CP	concept primitive	概念基元
CT	concept tree	概念树
OD	ontology description	本体论描述
ED	epistemology description	认识论描述
,L,	,logic link,	逻辑连接
LUG	language understanding gene	语言理解基因

这些 HNC 符号前两卷都使用过，直接阅读本卷的读者可能很不习惯，这没有关系，本卷下文都有交代。本《全书》篇幅过于庞大，笔者不仅希望每一卷可供读者独立阅读，甚至希望每一编以至每一章都能提供这种便利。这是前两卷进行了许多插写和预写的基本缘由，但希望归希望，结果很可能是事与愿违。

第一章

概念基元总论

引言

本章的地位等同于前两卷的第零章，是关于概念基元共相的论述，随后各章是其殊相的论述。这些话本应该放在卷首语里叙说，告知读者本卷的一项特殊约定：各编都不设置第零章，但第一章可能拥有前两卷第零章的地位。这毕竟属于技术性细节，所以就后移到这里了。

前两卷的章对应于概念林，节对应于概念树，这种对应性在本卷形式上不复存在，但实质上依然存在。这里的“实质上”是指 HNC 特别关注的所谓“体说、面说、线说和点说”^[*00]，编或篇大体对应于体说；章大体对应于面说；节大体对应于线说；小节大体对应于点说。但是，“体-面-线-点”是相对的，不是绝对的，四者之间“你中有我”，“我中有你”。汉语“面面俱到”不仅是体说的通俗解释，也是后续三说的通俗解释，“面面”者，齐备也；“俱到”者，透彻也。请把这两对“者，也”当作是“面面俱到”的真解吧。

本《全书》卷、篇、编、章、节、小节的设置与论述都依据上述“面面俱到”原则，本卷亦不例外，本章的下列 6 节就是实践该原则的产物：

第 1 节，概念基元有限性的形而上思考（语言概念空间基本景象的描述）。

第 2 节，概念基元的抽象与具体两分。

第 3 节，抽象概念基元的范畴三分。

第 4 节，基元、基本和逻辑概念的“6、2、4”分。

第 5 节，广义作用效应链的“二生三四”。

第 6 节，语境基元。