

第四届HNC与语言学研究学术研讨会论文集

HNC与语言学研究

(第4辑)

HNC YU YUYANXUE YANJIU

朱小健 张 全 陈小盟 主 编

2009



北京师范大学出版集团

BEIJING NORMAL UNIVERSITY PUBLISHING GROUP

北京师范大学出版社

第四届HNC与语言研究学术研讨会论文集

HNC与语言研究

(第4辑)

HNC YU YUYANXUE YANJIU

朱小健 张 全 陈小盟 主 编

2009



北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
北京师范大学出版社

图书在版编目(CIP)数据

HNC与语言学研究.第四辑 / 朱小健、张全、陈小盟主编.
—北京: 北京师范大学出版社, 2010.7
ISBN 978-7-303-10771-1

I. ①H… II. ①朱… ②张… ③陈… III. ①语言
学—学术会议—文集 ②自然语言理解—学术会议—文
集 IV. ①H0—53 ①H087—53

中国版本图书馆CIP数据核字(2010)第014560号

营销中心电话 010-58802181 58808006
北师大出版社高等教育分社网 <http://gaojiao.bnup.com.cn>
电子邮箱 beishida168@126.com

出版发行: 北京师范大学出版社 www.bnup.com.cn
北京新街口外大街19号
邮政编码: 100875

印刷: 北京京师印务有限公司
经销: 全国新华书店
开本: 184 mm × 260 mm
印张: 29.75
字数: 723 千字
版次: 2010年7月第1版
印次: 2010年7月第1次印刷
定 价: 66.00 元

策划编辑: 杨 帆 责任编辑: 杨 帆 郭 瑜
美术编辑: 毛 佳 装帧设计: 天泽润
责任校对: 李 菡 责任印制: 李 丽

版权所有 侵权必究

反盗版、侵权举报电话: 010-58800697

北京读者服务部电话: 010-58808104

外埠邮购电话: 010-58808083

本书如有印装质量问题, 请与印制管理部联系调换。

印制管理部电话: 010-58800825

第四届 HNC 与语言学研究学术研讨会

第四届HNC与语言学研究学术研讨会的举行及本书的出版得到“十一五”国家科技支撑计划“中文信息处理应用研究与系统开发”项目（编号：2007BAH05B00）的资助。

会议时间：2009年11月14日至15日

会议地点：北京师范大学

主办单位：北京师范大学中文信息处理研究所
中国科学院声学研究所
北京大正语言知识处理科技有限公司

名誉主席：许嘉璐

大会主席：朱小健

程序委员会

主 席：张 全

委 员：朱小健 陈小盟 晋耀红 苗传江

组织委员会

主 席：陈小盟

秘书长：丁紫惠

委 员：刘智颖 王宏丽 汪 涓 杨素华

前 言

本书是第四届“HNC 与语言学研究学术研讨会”的论文集。

HNC 是概念层次网络理论的简称,该理论由中国科学院声学研究所黄曾阳先生创立,它的基本特征是通过构建语言概念空间来探索和模拟大脑的语言理解过程。

“HNC 与语言学研究学术研讨会”(简称“HNC 研讨会”)最早由语言学家邢福义先生和黄曾阳先生倡议,它有三个基本宗旨:一是搭建语言信息处理界的学术交流平台,以碰撞促发展;二是创造有利于原创性新思想发展的学术环境,鼓励和发扬创新;三是理论探索与应用实践并重,促进学术界和产业界的沟通与互动。第一届 HNC 研讨会于 2001 年 4 月在华中师范大学举行,第二届于 2003 年 9 月在中国科学院声学研究所举行,第三届于 2005 年 12 月在北京师范大学举行。

第四届 HNC 研讨会于 2009 年 11 月 14 日至 15 日在北京师范大学举行,本届研讨会留给与会者的感受可以概括为四点:第一,HNC 的基础研究和应用开发与四年前相比又取得了很大的进展,HNC 团队进一步壮大;第二,学术界和产业界对 HNC 理论与技术有了更多的关注和了解;第三,中文信息处理的各个方面都在不断向前发展,新的思想、理论和技术不断产生,各个流派之间的交流不断扩大和加深;第四,论文作者大多是年轻的生力军,说明我国语言信息处理的发展后继有人。

前三届 HNC 研讨会的论文集,书名分别为《HNC 与语言学研究》《第二届 HNC 与语言学研究论文集》《中文信息处理的探索与实践》,本届 HNC 研讨会论文集的书名定为《HNC 与语言学研究(第 4 辑)》,以后将固定采用这一名称,只是改变括号中的数字,以与研讨会的届数对应。

本书收录了第四届 HNC 研讨会上宣读的 74 篇论文,分为 6 个部分:

1. 基础理论和语言本体研究(14 篇)
2. 知识库资源建设及其应用(12 篇)
3. 基础技术与开发(16 篇)
4. 应用研发之一:机器翻译(14 篇)
5. 应用研发之二:信息检索和抽取(11 篇)
6. 其他应用技术和系统研发(7 篇)

其中第四部分集中反映了 HNC 的最新进展,即在机器翻译方面的基础研究和应用实践。

本书出版时间比较仓促,编者水平也有限,如有疏漏和不妥之处,恳请读者和论文作者批评指正。

编 者
2009 年 12 月 1 日

目 录

第一部分 基础理论和语言本体研究

把文字数据变成文字记忆·····	2
黄曾阳	
关于汉语语义概念的一点思考·····	12
陆汝占	
概念基元空间的下沉及与词义结构的语言对接·····	18
萧国政	
现代汉语空间短语的基本类型·····	25
陈传海	
组织名在主流媒体上的分布信息及其价值·····	29
陈 慧	
从 HNC 看“从……到……”的语义角色及其判定·····	36
陈锡华	
HNC 与本体概念网的比较研究·····	42
胡金柱 舒江波 齐一真	
面向汉语框架语义知识库的动词“感觉”分析·····	49
刘鸣洋 由丽萍	
面向自然语言处理的人机语义研究·····	56
欧阳晓芳 萧国政	
基于 HNC 理论的“是……的”结构研究及教学·····	61
谭晓平 王红娟	
谈 HNC 的知识表示方法·····	68
王慧兰	
面向自然语言处理的语言本体研究·····	74
姚爱钢 易绵竹	
标点句的独立性判断·····	81
张瑞朋	
网络流行语的叙事特点及功能·····	88
邹春燕 张茂元	

第二部分 知识库资源建设及其应用

语义互联网与《新编同义词词林》·····	94
冯志伟	
事件描述块句法语义标注库的构建·····	101
郭嘉伟 周明海 刘金凤 亢世勇	

释雅例句检索工具的需求分析·····	108
雷 静	
动词词典在汉语词汇教学中的应用·····	113
刘宁静 翟保军 赵 星	
略论基于 HNC 的学习型电子词典编纂研究·····	118
刘宁静	
HNC 语义标注语料库查询工具的用户需求·····	123
刘智颖	
句群处理突破的关键:HNC 领域句类知识库·····	130
缪建明 张 全	
汉英双语多元逻辑组合标注的基本内容·····	136
钦丽丽	
法治活动的领域句类表示式和知识描述·····	143
宋培彦 刘宁静 赵 星	
基于本体的多媒体教学资源管理及检索研究·····	150
王 冠 张克亮	
语言资源建设的理论与实践初探·····	156
曾小兵 邱丽娜	
跨越语句理解的知识(库)建设·····	163
张 全 袁 毅 缪建明 韦向峰 吴崇斌	
第三部分 基础技术与开发	
无头迭句与花园幽径句的辨识与处理·····	172
池哲洁 李 颖	
英语特征语义块构成分析及计算机处理·····	179
董凌冲 李 颖 池毓焕	
结合领域句类知识的省略恢复方法·····	186
黄 娴 张克亮	
面向大句处理的句类分析系统的设计与实现·····	193
晋耀红	
英语非限定形态动词分析处理·····	202
李 颖 马海慧 董凌冲	
宪法句间回指模糊的消解研究·····	209
宋培彦	
汉语理解处理中的动态词及其组合模式·····	216
唐兴全	
句处理与句群处理·····	222
韦向峰	
利用读音信息的词语切分处理·····	229
吴崇斌	

现代汉语文本中四字语的识别·····	234
武慧敏	
英语主辅语义块辨识初探·····	238
翟保军	
中文机构名识别需要使用的知识资源·····	244
张冬梅	
现代汉语数量短语及其自动识别的研究·····	250
张学青	
HNC 符号的解析及应用 ·····	255
赵 星	
中文信息处理中兼类词研究困难的原因·····	262
周明海	
HNC 理论的句类假设检验知识与现有句类分析系统 E 语义块识别的检测分析 ·····	269
朱 筠	

第四部分 应用研发之一:机器翻译

形如独立小句的原型句蜕之汉英变换·····	276
池毓煊 李 颖 连巍魏	
从 HNC 机器翻译评测看机器翻译发展 ·····	284
丁泽亚 张 全	
从一项实证研究看机器翻译的路线之争·····	290
李 青 张克亮	
面向汉英机器翻译的格式自转换研究·····	297
连巍魏 张克亮	
关联于是否判断句的迭句转换研究 ——以汉英机器翻译为视角·····	304
连巍魏	
英语方式动词在英汉机器翻译中的处理分析·····	311
刘 然 刘亦娴	
汉英机器翻译中的辅块序位研究·····	318
王立霞	
英汉机器翻译中“OF”结构的处理 ·····	324
杨 森 胡富茂	
基于 HNC 理论的时间辅语义块的汉英机器翻译初探 ·····	331
臧翰芬	
块扩句的汉英句类及句式转换·····	338
曾 维 张克亮	
汉语小句向英语辅块的变换研究·····	345
翟保军 宋培彦 赵 星	

汉英机器翻译中“以……为……”句式的理解处理	352
张 玲 刘建勇	
统计机器翻译的句法调序与 HNC 句式转换	359
张霄军	
简析机器翻译软件的需求	364
张玉玺	

第五部分 应用研发之二:信息检索和抽取

中文文本中企业并购信息的自动抽取	370
韩伟峰	
农业信息检索中含单动词和双名词的词语串之语义理解策略	376
刘鹏元 杨建艳	
针对 HNC 理论的特殊语序结构的抽取策略 ——基于足球比赛报道新闻文本的抽取任务	381
罗雯涛	
求职简历信息的自动抽取	388
任 宁	
基于规则的人物关系抽取所需的语言知识研究	395
佟佳丽	
搜索引擎的个性化和智能化	400
吴 艳	
农业信息智能检索的知识需求分析	405
杨建艳	
基于 HNC 理论的精确检索方法的设计与实现	409
袁 毅 张 全	
中国奥运冠军动态信息抽取技术研究	414
张 芳	
专利文献检索再分类的语言知识	417
郑 婧	
网络文本中对企业产品和服务评价倾向的表达方式分析	423
周静静	

第六部分 其他应用技术和系统研发

基于 HNC 的网络信息舆情分析系统应用	430
陈威烈	
HNC 知识检索系统在专利文献检索中的应用	435
李长宏 刘 庆	
HNC 在农业垂直搜索中的应用	440
刘秋明	

HNC 在 Web 日志管理中的应用探讨	445
龙华仁	
HNC 在网站客服系统中的应用探讨	448
龙华仁	
基于向量空间模型的话题主动发现和跟踪	452
王智超 晋耀红 张 舵	
汉语网络词典的现状和发展方向浅析	459
望小玲	



第一部分

基础理论和语言本体研究



把文字数据变成文字记忆

黄曾阳

中国科学院声学研究所 北 京 100190

hy@mail.ioa.ac.cn

摘 要：本文论述了把文字数据变成文字记忆的三点基础性思考。第一点是关于语言记忆与语言理解关系的略说，以记忆公理和理解基因的概念为立足点，仰望了记忆与理解的互动性；第二点是关于记忆模式和记忆样式的略说，以显记忆—隐记忆—动态记忆和共相记忆—殊相记忆等概念为立足点，仰望了语言记忆的两种模式和两种范式；第三点略说了文字记忆技术实现的三大战役——机器翻译、一目千行和智力培育。上述一系列作为立足点的概念是 HNC 理论必然衍生出来的新概念，形成已久。不过，多数是第一次公开出现，但在 HNC 团队内部，曾以相同或不同的名称使用多年了。相对于第三届 HNC 与语言学研讨会而言，这次仰望是悲观一些还是乐观一些？应该是兼而有之吧。

关键词：理解基因，隐记忆，记忆公理

1 引 言

1.1 信息转换的低级与高级形态

在 20 世纪众多的科技明星中，粉丝最多的一位叫“数字化”。数字化就是把各种形态的信息统一转换成数字形式。数字化是一位女明星，数字化是她的艺名，其学名应该是“信息转换的低级形态”或“低级信息形态转换”。

这位女明星尚未出嫁。这里预告一声，其未来夫婿的艺名叫“智力化”（包括智能化）。“智力化”的核心科学问题之一就是把数据变成记忆，这也是信息形态的一种转换。但这一转换的层次显然远高于数字化转换，故名之“信息形态的高级转换”或“高级信息形态转换”。

我们已经看到，“低级信息形态转换”（数字化）已经造就了多少壮丽的信息产业数字化辉煌，那么，“低级信息形态转换”与“高级信息形态转换”（智力化）的结合（结婚）将会造就什么样的信息产业辉煌呢？能不能说，这是两种具有天壤之别的辉煌呢？

1.2 大脑硬件的基本特征

表面上看，似乎大脑的硬件（大脑皮层的生理结构）已经搞得相当清楚了，大脑的软件（大脑皮层的智力运作过程）也有诸多发现^①。然而，问题在于大脑之谜的探索需要硬件与软件并举的思考，而这种思考是很欠缺的。

大脑的硬件似乎可以同计算机进行类比，划分为 I/O（接口）、CPU 和 MEM（内存与外存）这三个基本环节。但是，计算机的 CPU 目前根本不存在语言、图像、艺术、科学和情感的范畴区分，而大脑的 CPU 肯定存在着这样的范畴区分，这区分体现了大脑 CPU 的基本特征。这个命题很重要，可惜还没有获得应有的重视。

^① 主要指酶分泌、神经胶质细胞、传导液等的发现。

1.3 关于大脑软件基本特征的设想

大脑的软件似乎也可以同计算机作某种类比,划分为自我操作和对外服务两个侧面。前者密切联系于CPU+MEM,后者密切联系于(I/O, MEM, CPU)。

大脑的操作软件有什么不同于计算机的本质区别?

HNC的答案是:大脑操作软件存在着智能与智慧的基本差异。这里不介绍智能与智慧的定义及其HNC符号表示式,只用一个例子来表明两者之间的巨大不同。柏拉图与孔夫子、恺撒与拿破仑都智力超群,但两位夫子和两位将军的具体智力表现可大不相同,能不能说“两位夫子智慧超群而智能平平,两位将军智能超群而智慧平平”呢?好像是可以这么说的,可见,智力存在着智能与智慧的本质区分。孔夫子正是由于智能平平而于最近引出了“孔子,丧家犬”的命题,这命题受到许多学者的盛赞。但问题不在于盛赞或反对,而在于对先哲的基本态度。柏拉图经历过与孔子极为类似的境遇,但希腊人和西方人绝不会把丧家犬之类的侮辱性描述加到柏拉图身上。尼采先生确实喊出过“上帝死了”,但绝不会超出这个限度而高喊“打倒上帝”。

数据和记忆是两种性质截然不同的信息载体。数据仅涉及信息的量与形式,不涉及信息的质与内容,无关于智能与智慧的差异;记忆则不仅涉及信息的量与形式,更涉及信息的质与内容,有关于智能与智慧的差异。可见,记忆所要求的信息形态转换,其难度必远大于数据。作为一项科技课题,“高级信息形态转换”的命运非常奇特,很像那位出塞前的王昭君。

1.4 准备迎娶“昭君”

可是,命运类似于王昭君的“高级信息形态转换”并不是“昭君”,而是“昭君”的未来夫婿。

上面说到的那位数字化美人才是“昭君”,她还没有找到如意郎君。上帝似乎在刻意安排一场年龄差距破历史纪录的姐弟恋。那位美人的芳龄已经超过了30岁,可是那位未来的新郎还没有出世。他以胎儿的形态已经存在多年了,但还没有降生。最近的检查表明,胎儿发育正常,主要问题是母亲营养不良。预产期还没有完全确定,乐观的估计是2012年,在座诸君应该都能赶上。不过,能在这个研讨会上想象一下从“昭君出塞”到“迎娶昭君”的历史巨变,已经是足够欣慰的事了。

2 语言记忆与语言理解

2.1 语言记忆公理

记忆公理1:记忆必须以某种符号形式存在于大脑的某一特定区域(例如海马),语言记忆使用的符号绝不是语言符号的拷贝,而是语言符号的某种变换形态。

记忆公理2:语言记忆是语言理解的前提与结果,没有语言理解就没有语言记忆,没有语言记忆也就没有语言理解。故婴儿没有语言记忆,听不懂的话语和看不懂的文字不会形成内容记忆,但可能形成某种对象记忆^①。

“理解与记忆”悖论——蛋与鸡的悖论。

^① HNC理论对“对象与内容”、“具体概念和抽象概念”都给出了特定的定义,这些特定定义对语言的分析、理解与生成可成立竿见影的效果。在语句分析阶段(层面),“对象与内容”密切联系于句类;在语境分析阶段(层面),对象仅联系于具体概念,内容仅联系于抽象概念。这时,对象记忆是关于特定人与物的记忆,而内容记忆是关于特定事件的记忆。

“语言学与公理1”悖论^①。

2.2 自然语言符号体系和语言概念空间符号体系——为什么训诂学大师们看不起《马氏文通》

自然语言符号体系的地位和价值被索绪尔先生的崇奉者过分抬高了,因为自然语言符号体系不过是语言概念空间符号体系的殊相表现,而且它只关系到大脑(交际引擎)的接口,根本无关于大脑(交际引擎)的CPU与MEM。

语言概念空间符号体系才是大脑CPU(智力)与MEM(记忆)实际运用的符号体系,承载这个符号体系的硬件(大脑皮层的神经系统)是人类百万年进化的产物,其发育成长过程的基本定型大约要5年左右的时间,运行这个符号体系的软件则需要至少20年的“编程”与“试运行”时间。这一点,奥古斯丁早在1600年前就讲得比较清楚了^②,理论语言学和心理学的研究似乎都没有充分参照奥古斯丁的重要思考。

训诂学大师们与奥古斯丁“心有灵犀”,他们之所以看不起《马氏文通》,要害就在这里。

让我们把上面的话语从交际的角度重复一下:语言交际绝不是一个单纯运用自然语言符号体系的过程,而是自然语言符号体系和语言概念空间符号体系两者交替运用的过程。这里存在两个基本过程:表达过程和理解过程,表达过程最终必须通过输出接口使用自然语言符号体系,但理解过程仅使用语言概念空间符号体系,自然符号体系的作用仅限于通过输入接口激活理解基因。

语言概念空间符号体系是内在物,是自我的终极载体之一,是交际引擎的中枢;语言符号体系不过是外在物,不是自我的终极载体之一,而只是自我表现的手段之一,仅有关于交际引擎的对外接口。

如果我们只关心语言交际,那么我们也许可以只关注自然语言符号体系,而不必去关心那个看不见、摸不着的语言概念空间符号体系;但是,如果我们要去探索人类思维的奥秘或大脑运作过程的奥秘,那就必须同时关注两个符号体系,而不能只关注一个符号体系了。不仅如此,还必须把探索的重点转移到语言概念空间符号体系,因为这个符号体系才是思维的存在之本。

这个答案是不是一个老掉牙的答案呢?难道科学如此昌明发达的西方专家不知道这个答案吗?难道西方的相关学界没有思考过语言概念空间符号体系吗?人工智能学界没有思考过吗?计算语言学界没有思考过吗?认知心理学界没有思考过吗?语言哲学界没有思考过吗?《皇帝新脑》一书的作者没有思考过吗?乔姆斯基先生以其毕生精力研究过的普适语法UG难道不属于语言概念空间符号体系吗?

这些问题的答案不过是一层窗户纸,但笔者历来采取“走为上”的态度,不去碰它。

^① 语言学关于“语言与思维的经典陈述”里存在原则性的不妥论断,例如下面的论述:“从功用来说,语言又是人类赖以思维的工具,人进行思维,思考问题,都必须依附于某种具体的语言,所以语言一向被认为是思维的物质外壳。”这可视作“语言学与记忆公理1”悖论的典型代表。

^② 奥古斯丁在《忏悔录》里有一系列关于思维与语言的精彩论述,这里仅摘录其中的一段:“我开始学语了,并不是大人们依照一定程序教我言语,和稍后读书一样;凭仗你,我的天主赋给我的理智,用呻吟,用各种声音,用肢体的种种动作,想表达出我内心的思想,……但不可能表达我要的一切……为此,听到别人指称一件东西,或看到别人随着某一声音做某一种动作,我便记下来;我记住了这东西叫什么,要指那件东西时,便发出那种声音。……这是各民族的自然语言:用面部表情,用目光和其他肢体动作,用声音表达内心的情感,或为要求,或为保留,或为拒绝,或为逃避。这样一再听到那些语言,按各种句子中的先后次序,我逐渐通解它们的意义,便勉强鼓动唇舌,借以表达我的意愿。”(注:加黑处是笔者加的)

2.3 语言概念空间符号体系与语言理解基因

语言概念空间符号体系是大脑里的五类概念空间符号体系之一，另外四类概念空间符号体系的名称分别是：图像、艺术、科学和情感。

每一个符号体系都拥有自己的理解基因。理解基因主要是智力基因，而不是生理基因。大脑之谜实际上主要是五类理解基因之谜，而不仅仅是生理基因之谜。当然，理解基因与生理基因是相互交织的，但交织度应该存在巨大差异：语言、图像、科学和艺术理解基因的交织度比较弱，情感理解基因的交织度比较强。五类理解基因的迷雾度存在着巨大差异：图像与艺术理解基因的迷雾度最浓，科学次之；情感之谜的迷雾度似明似暗，在那里，理解基因的作用应小于生理基因。

以上论述是为了烘托出这样的命题：语言理解基因的迷雾已基本洞开。

2.4 语言理解基因的自然语言表述

笔者经常思考，什么样的东西可以用来描述语言理解基因的结构与功能呢？神经元显然是不合适的，因为单个神经元不过是理解基因的一个元件，连组件的资格都不具备，而理解基因必然是许多神经元的复杂组合或连接。笔者偶然联想到印度的种姓制度，觉得它是一个非常合适的描述样板。请看下表：

印度的种姓	HNC 理解基因的范畴
婆罗门：精神生活(文化、教育)	三类精神生活
刹帝利：政治(军事、法律)	主导性第二类劳动
吠舍：经济、医疗	基础性第二类劳动
首陀罗：体力劳动	第一类劳动
贱民：低级体力劳动	本能活动

此表说明，古老的印度种姓制度最接近于 HNC 对理解基因的第一级(范畴)划分，这“最”是相对于希腊—罗马文明、古老中华文明而言的。前者有“贵族—平民—奴隶”的划分，后者有“士农工商”的划分，这两种划分也考虑到了社会的结构性和功能性，但并没有将两者融为一体，种姓制度则体现了一种彻底的融合，用哲学的术语来说就是：本体论与认识论的彻底融合。现代印度正在逐步消解这种古老印度式融合的非正义性或非人道性，不过，如果单就融合本身来说，应该说它体现了一种高级的综合与演绎智力，而现代文明则过于强调分析与归纳智力了，20 世纪后期的解构主义更是把这种强调推向了极端。笔者对解构主义的浪潮深感恐惧：第一，各学科都由于这一浪潮的推动而走上了置综合与演绎于不顾的可怕地步；第二，任何信念、理念或主义一旦被解构，就会演变出极度畸形的价值观。

理解基因就是智力基因的别称，是智力结构与功能的完美结合，是智力本体论与认识论的完美结合，这种完美结合是全能上帝的创造或“专利”，开普勒的惊叹是有道理的。但人们往往只注意到开普勒惊叹的话语，而忽略了它的实质：原来如此复杂的物理现象竟然可以用一个如此简明的符号体系来加以描述，这一点既是绝大部分科学实际走过的历程，也是所有科学的基本共识。HNC 的使命就是对语言概念空间给出一个符号体系的描述，而且顺便把这个符号体系数字化(为了迎娶数据数字化这位“昭君”)。但是，该符号体系只追求康德所说的透彻性和齐备性，也就是说，只追求智力结构与功能的某种融合，而不追求完美性，因为

如上所述,完美性是上帝的“专利”。某种者,局限于语言概念空间也。

语言理解基因的基础符号体系就是[HNC-1],其符号表示式比较复杂,这里就不写了。但是,其总体设计思路十分简明,那就是下面的语言表述:

理解基因: = 范畴表示 + 结构与功能的各级综合表示

范畴与“各级”者,层次也;结构与功能者,网络也,局部网络也,非全局网络也。

这里有一项关键性的技巧,那就是对结构与功能分别赋予不同的约定数符号。

语言理解基因的全貌也是可以用语言来描述的,那就是下面的六句话:[HNC-1]是语言概念空间符号体系的数字化表示;部分[HNC-1]是语言理解基因的基础结构;[HNC-2]和[HNC-3]是语言理解基因的上层建筑;概念关联式是语言理解基因的主体信息渠道;语言理解基因主要靠词语直接激活,这种激活是大脑输入接口的基本转换功能之一;语言记忆就是从[HNC-3]到[HNC-4]的转换。

2.5 语言记忆贯穿于语言理解过程的始终

这个问题需要从记忆公理2谈起。

记忆公理2说“语言记忆是语言理解的前提与结果,没有语言理解就没有语言记忆,没有语言记忆也没有语言理解”。这个论断里,记忆和理解都出现了3处,3处理解具有相同的内涵与外延,但3处记忆则具有不同的外延。为叙述便利,下文将把这3处记忆依次简称记忆1、记忆2和记忆3。

先说记忆2。记忆2对应于通常所说的记忆,即“从[HNC-3]到[HNC-4]的转换”所形成的记忆,上文曾名之语言记忆,也就是HNC命名的语境生成或语境,是语言记忆的可自感(可回忆)部分,可名之显记忆;记忆3则由稳定记忆与动态记忆两者构成,稳定记忆包括语言理解基因的基础结构、上层建筑、主体信息渠道和输入接口转换器,是记忆的不可自感(不可回忆)部分,可名之隐记忆。动态记忆大体相当于心理学的工作记忆,它又分为两部分,一是语境生成过程的过渡信息,最终不纳入显记忆,二是从已有显记忆里临时调用的相关记忆片段。

记忆1是记忆2与记忆3的总和,可名之广义记忆,包含显记忆、隐记忆和动态记忆这三种记忆类型。

HNC给出了显记忆和隐记忆的符号表示式,但尚未给出动态记忆的符号表示式,这属于记忆工程第三战役——智力培育的任务。

记忆力是指显记忆的能力,它强关联于一个人的知识面,弱关联于一个人的智力。智力主要决定于隐记忆和动态记忆的能力,从这个意义说,智商是一个有待改进的概念,因为它既未作显记忆、动态记忆、隐记忆这3种记忆类型的区分,也未作智能与智慧的区分。而这样的区分对智力的研究或所谓脑力的开发至关重要。

隐记忆能力是大脑智力的核心指标,是大脑软件“编程”与“试运行”进度的基本考核指标,对这项指标的研究需要新的思路,HNC团队有责任推进该思路的酝酿与完善。这些话实际上已经是“事后诸葛亮”了,可当今的诸葛亮们,在功利主义和解构主义全球浪潮的冲击下,基本处于昏迷不醒的可怜状态。所以,这里不得不说不一声:各位同仁要提高警惕啊!

3 语言的记忆模式与记忆样式

3.1 语言记忆两模式：范畴—领域模式和对象—内容模式

HNC 理论“三‘无限—有限’说”的第三说是：语境无限而语境单元有限。这里的语境其实就是指显记忆，语境单元其实就是指隐记忆，也就是语言理解基因。请记住下面的两个基本命题吧：

命题 1：语境单元有限 =：理解基因有限

命题 2：理解基因有限 $\equiv$ 语言概念空间的范畴—领域有限

（“=：”者，等同也，非完全等同也；“ $\equiv$ ”者，强关联也，互为因果者也。）

上列命题的推理是：显记忆（语境）是理解基因运作的结果。理解基因是依据范畴与领域排序的，那么，显记忆跟着“如法炮制”就是最自然不过的“理所当然”了。“如法炮制”的学名就是语言记忆的范畴—领域模式。

但是，范畴—领域记忆模式显然不利于动态记忆的形成，动态记忆需针对特定的对象或内容，这些特定信息必然散布在不同的范畴—领域。从大的方面说，人都会有三类精神生活和两类劳动的个人范畴经历，从小的方面说，专业人士不但会有不同专业的个人领域经历，也会有各种精神生活的个人领域经历。因此，必须配置另外一种语言记忆模式以便于针对特定对象或内容的搜寻。这另一种语言记忆模式就是对象—内容模式。[HNC-3]和[HNC-4]分别以范畴—领域(DOM)和对象—内容(BC)为纲，道理就在这里。

范畴—领域模式和对象—内容模式对应于两种索引方式，这有点类似于汉语词典的笔画索引和拼音索引，《现代汉语词典》就配置了这两种索引，但两者有虚实之别。实者，拼音索引也；虚者，笔画索引也。

大脑语言记忆两种模式的存在乃联想逻辑的简明推论，可当作记忆公理 3 来对待。但 HNC 还引入了一项基本假设：范畴—领域记忆模式是实模式，拥有相当于《现汉》的拼音索引，对象—内容记忆模式是虚模式，拥有相当于《现汉》的笔画索引。后者与语言输入接口直接连通。

3.2 语言记忆两样式：共相样式和殊相样式

语言记忆两模式只面向显记忆；语言记忆两样式则是面向广义记忆。

广义记忆存在两种样式：共相样式和殊相样式。共相和殊相是一对哲学概念，意思同共性与个性“差不多”，可以说，它就是共性与个性的哲学表述。就语言记忆来说，如果使用共性样式和个性样式的术语也未尝不可，但笔者还是觉得共相样式和殊相样式的表述要传神得多。为什么？这需要从语言记忆公理 2 说起。

语言记忆公理 2 的论断之一是“没有语言记忆就没有语言理解”，此论断里的“记忆”不是人们常说的记忆，而是前已指出的隐记忆和动态记忆。这个记忆应该处于大脑皮层的“中央”区，而不应该处于大脑皮层的边缘区——海马。爱因斯坦说过“上帝是奇妙的”，这里也许可以补充说“上帝的第一奇妙就是把隐记忆和动态记忆安置在大脑的中央区，把显记忆安置在大脑的边缘区”。这样的安置或设计才符合智力效率原则，最近有人说，“大脑如同城市”^①，此话多少有点形而上的味道，因为智力效率原则的空间特性应该与城市效率原则相类似。

① 大脑如同城市的说法见《参考消息》2009 年 9 月 9 日第 7 版。