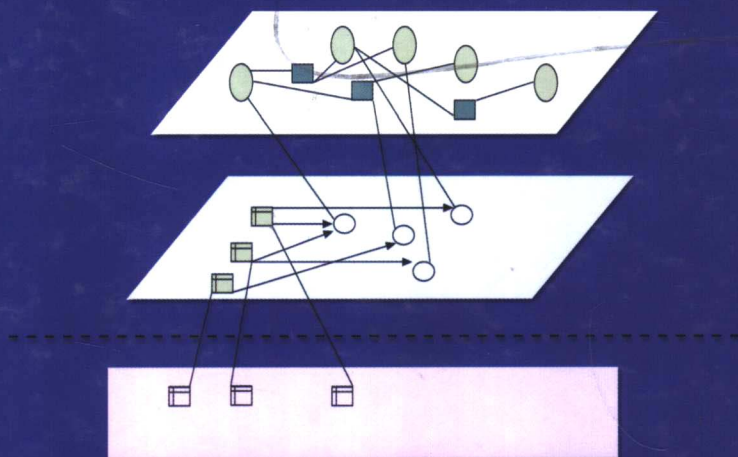




智能科学技术著作丛书

HNC(概念层次网络) 语言理解技术及其应用

晋耀红 著



科学出版社

www.sciencep.com

>> 本书特点

- >> HNC是概念层次网络 (hierarchical network of concepts) 理论的英文简称, 它是一个以概念联想脉络为主线, 融语义、语法、语用为一体的自然语言理解的理论体系。本书全面系统地介绍了基于HNC理论的语言理解技术, 特别是HNC语句理解系统的基本思路、处理策略、相应算法的设计以及试验结果, 简明扼要地说明了几个利用HNC技术开发的应用技术 (包括语境单元萃取技术、语义立场判断技术等), 以及HNC语言理解技术在文本特征提取、文本相似度计算、文本分类以及信息检索、信息过滤中的应用。本书内容在机器翻译、智能检索、自动文摘以及知识挖掘、人机对话、个性化服务等领域有广阔的应用前景。

智能科学技术著作丛书

智能无线传感器网络系统 于海斌 等

► HNC(概念层次网络)语言理解技术及其应用 晋耀红

脉冲耦合神经网络原理及其应用 马义德 等

免疫优化计算、学习与识别 焦李成 等

协同进化计算与多智能体系统 焦李成 等

高级人工智能(第二版) 史忠植

实时嵌入式系统 王飞跃 等

ISBN 7-03-016190-4



9 787030 161901 >

ISBN 7-03-016190-4

定价: 75.00 元

销售分类建议: 通信技术; 信息处理

H087

7

智能科学技术著作丛书

HNC (概念层次网络)

语言理解技术及其应用

晋耀红 著

科学出版社

北 京

内 容 简 介

HNC是概念层次网络(hierarchical network of concepts)理论的英文简称,它是一个以概念联想脉络为主线,融语义、语法、语用为一体的自然语言理解的理论体系。本书全面系统地介绍了基于HNC理论的语言理解技术,重点说明了HNC语句理解系统的基本思路、处理策略、相应算法的设计以及试验结果,详细说明了几个利用HNC技术开发的应用技术,包括语境单元萃取技术、语义立场判断技术等,并简单介绍了HNC语言理解技术在文本特征提取、文本相似度计算、文本分类以及信息检索、信息过滤中的应用。本书内容在机器翻译、智能检索、自动文摘以及知识挖掘、人机对话、个性化服务等领域有广阔的应用前景。

本书的读者对象为人工智能、语言信息处理、语言研究等领域的学者、研究生、本科生以及其他对智能信息处理感兴趣的研究者和产品开发人员。

图书在版编目(CIP)数据

HNC(概念层次网络)语言理解技术及其应用/晋耀红著. —北京:科学出版社, 2006

(智能科学技术著作丛书)

ISBN 7-03-016190-4

I. H… II. 晋… III. 自然语言处理 IV. H087②TP391

中国版本图书馆CIP数据核字(2005)第096397号

责任编辑:李 锋 田士勇 张 敏 / 责任校对:宋玲玲

责任印制:安春生 / 封面设计:陈 敬

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>

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2006年4月第 一 版 开本: B5 (720×1000)

2006年4月第一次印刷 印张: 29 3/4

印数: 1-2 500 字数: 544 000

定价: 75.00 元

(如有印装质量问题,我社负责调换〈新欣〉)

《智能科学技术著作丛书》编委会

名誉主编：吴文俊

主 编：涂序彦

副 主 编：钟义信 史忠植 何华灿 蔡自兴 孙增圻 童安齐 谭 民

秘 书 长：韩力群

副秘书长：田士勇

编 委：(按姓氏汉语拼音排序)

蔡庆生(中国科技大学)

蔡自兴(中南大学)

杜军平(北京工商大学)

韩力群(北京工商大学)

何华灿(西北工业大学)

何 清(中国科学院计算技术研究所)

黄河燕(中国科学院计算语言研究所)

黄心汉(华中科技大学)

焦李成(西安电子科技大学)

李祖枢(重庆大学)

刘 宏(北京大学)

刘 清(南昌大学)

秦世引(北京航空航天大学)

邱玉辉(西南师范大学)

阮秋琦(北京交通大学)

史忠植(中国科学院计算技术研究所)

孙增圻(清华大学)

谭 民(中国科学院自动化研究所)

田士勇(科学出版社)

童安齐(科学出版社)

涂序彦(北京科技大学)

王国胤(重庆邮电学院)

王家钦(清华大学)

王万森(首都师范大学)

吴文俊(中国科学院系统科学研究所)

杨义先(北京邮电大学)

尹怡欣(北京科技大学)

于洪珍(中国矿业大学)

张琴珠(华东师范大学)

钟义信(北京邮电大学)

庄越挺(浙江大学)

《智能科学技术著作丛书》序

“智能”是“信息”的精彩结晶，“智能科学技术”是“信息科学技术”的辉煌篇章，“智能化”是“信息化”发展的新动向、新阶段。

“智能科学技术”(intelligence science & technology, 简称 IST)是关于“广义智能”的理论方法和应用技术的综合性科学技术领域,其研究对象包括:

- “自然智能”(natural intelligence, 简称 NI),包括:“人的智能”(human intelligence, 简称 HI)及其他“生物智能”(biological intelligence, 简称 BI)。

- “人工智能”(artificial intelligence, 简称 AI),包括:“机器智能”(machine intelligence, 简称 MI)与“智能机器”(intelligent machine, 简称 IM)

- “集成智能”(integrated intelligence, 简称 II),即:“人的智能”与“机器智能”人机互补的集成智能。

- “协同智能”(cooperative intelligence, 简称 CI),指:“个体智能”相互协调共生的群体协同智能。

- “分布智能”(distributed intelligence, 简称 DI),如:广域信息网,分散大系统的分布式智能。

1956年,“人工智能”学科诞生,50年来,在起伏、曲折的科学征途上不断前进、发展,从狭义人工智能走向广义人工智能,从个体人工智能到群体人工智能,从集中式人工智能到分布式人工智能,在理论方法研究和应用技术开发方面都取得了重大进展。如果说,当年“人工智能”学科的诞生是生物科学技术与信息科学技术、系统科学技术的一次成功的结合,那么,可以认为,现在“智能科学技术”领域的兴起是在信息化、网络化时代又一次新的多学科交融。

1981年,“中国人工智能学会”(Chinese Association for Artificial Intelligence, 简称 CAAI)正式成立,25年来,从艰苦创业到成长壮大,从学习跟踪到自主研发,团结我国广大学者,在“人工智能”的研究开发及应用方面取得了显著的进展,促进了“智能科学技术”的发展。在华夏文化与东方哲学影响下,我国智能科学技术的研究、开发及应用,在学术思想与科学方法上,具有综合性、整体性、协调性的特色,在理论方法研究与应用技术开发方面,取得了具有创新性、开拓性的成果。“智能化”已成为当新技术、新产品的发展方向和显著标志。

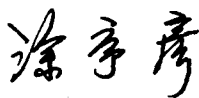
为了适时总结、交流、宣传我国学者在“智能科学技术”领域的研究开发及应用成果,中国人工智能学会与科学出版社合作编辑出版《智能科学技术著作丛书》。需要强调的是,这套丛书将优先出版那些有助于将科学技术转化为生产力

以及对社会和国民经济建设有重大作用和应用前景的著作。

我们相信，有广大智能科学技术工作者的积极参与和大力支持，以及编委们的共同努力，《智能科学技术著作丛书》将为繁荣我国智能科学技术事业、增强自主创新能力、建设创新型国家做出应有的贡献。

祝《智能科学技术著作丛书》出版，特赋贺诗一首：

智能科技领域广
人机集成智能强
群体智能协同好
智能创新更辉煌



中国人工智能学会荣誉理事长

2005年12月18日

序一 研究中文信息处理，需要仰望一下天空

2005年3月14日收到晋耀红博士的一封信，要我为他的专著《HNC（概念层次网络）语言理解技术及其应用》写篇序，随信附来了书稿。我当即答应了。但时至今日，距离2006年只剩下不到10天了，我还没交卷。个中原因是不能用一个“忙”字解释得了的，我也不能以此向晋博士请求谅解。

我是计算机技术的外行。十年来，从中文信息处理界的朋友们那里学到了很多知识，其中从耀红博士的导师黄曾阳先生那里获益尤多。尽管主观上并没有偷巧，但无奈根基（先天）浅，悟性（后天）差，至今我也没能成为他们像样的学生，至多像个球迷，在场边呐喊，为好球欢呼，因球臭而悻悻。现在要给耀红博士的书写序，岂可以草率为之？而我的习惯是，实在不得已要写序，一定要把所序之书至少看一遍才敢下笔。外行、吃力和忙，以致拖到今天才得塞责。

中国学者已经为中文信息处理不懈地奋斗了二十几个春秋，其间的顺利与挫折、进展与停滞、成功与失败，难以言尽；从总体趋势看，还是取得了一个又一个的成果，而且有的不断转化为生产力，产品迭出。中文信息处理技术的巨大进展，即使是终端用户也都感觉到了：平时使用的各种软件、移动通讯工具，越来越方便，也越来越便宜。中文信息处理技术的进步，对我国计算机、多媒体产品和现代通讯工具的普及起了很大作用，在其他学科的研究中也逐渐得到越来越广泛的应用。

但是，已经取得的这些成就，距离人们期望的和计算机可以完成的目标——实现人机对话、机器翻译，使计算机更像电“脑”——还有很长的路程。

今后的路，达到理想彼岸的路，关键在哪里？如何让计算机能够“理解”语言（文本语言和自然语言）？中外专家孜孜以求的就是这个难题之解。语言的本质功能——第一，交际；第二，思维——就决定了对语言的理解最重要的是把握语义；各种语言最主要的差异也在于表达意义的方式方法方面。

这些年我国大多数研究者紧紧跟踪国外计算机语言学的步伐，及时引进，及时转化，并且时时有所发展。但是，不能不看到，国外的计算机语言学是基于英文文本的，这对于缺乏形态的汉语来说，在相当程度上是不适用的。即使在国外国外，与从索绪尔到乔姆斯基的专重语言形式不同，近些年也相继出现了概念依存、逻辑语义等以语义为重心的理论，其间格语法等也考虑到了语义问题。但是，依据这些面向曲折语的可谓层出不穷的新理论，在解决其本国语言（主要是英、德、法、西等语种）时，至今也并没有真正进入处理句子和句群的阶段；沿

着这条路来研究汉语, 似乎也很难看到解决问题的曙光。

汉语的语义在组词成句、组句成篇中的作用尤其巨大。现在中外学者已经逐渐意识到这一点了, 开始有了种种探索。其显要者, 如我在几年前所说的, 有三种风格或者叫三个流派; 这三个流派虽然都把如何突破语义这一瓶颈作为主攻方向, 但各有特长, 从基本理念到解决方案, 从对语义的分析到模型、算法, 都很不相同。我还说过, 看来要较好地解决中文信息自动化处理问题, 哪一家都不能“包打天下”, 必须互相学习, 短长相补(许嘉璐 2005), 很可能当难题接近基本解决时, 那时的核心理论和技术既非此亦非彼, 而是融合了各家(一定不止现在的三家)的新学说。这一预测在我所主持的“863”项目——中文信息处理现代汉语应用基础研究——开展三年过程中得到了初步验证。

我主持的这一课题容纳了上述三个流派, 使大家各尽所长。我们在开展研究时实现了课题组内资源共享; 每年都借阶段性检查和评估的机会进行交流研讨; 这种研讨没有任何泛论, 都是就每个子课题的进展进行评价和质疑, 对所遇到的问题进行颇有会诊味道的讨论、争辩; 某一家的思路变为共有, 另一家的数据库供大家使用, 不少子课题从他人那里获得启示或补充, 于是大家一起进步了。我们之所以能这样做, 是因为每人都知道语言特别是语义的复杂, 只有集多人多派之力才能得到原创性的成果。

在这里我们要特别提到三个流派之一的 HNC(概念层次网络)理论, 近几年的进展尤可称道。其基础理论研究在进一步完善对语境单元萃取和语境生成的探索、概念关联性的表述方式等方面又有了进展, 而且已经进入句群的语义关联的研究; 依据其理论所开发的多种智能型应用软件已经实现产业化, 投入市场后得到广泛好评。

有关 HNC 理论原理的基本框架已经有黄曾阳先生及其弟子的许多论著详细介绍了, 特别是苗传江博士的《HNC(概念层次网络)理论导论》(2005, 清华大学出版社)和晋耀红博士的这部《HNC(概念层次网络)语言理解技术及其应用》, 更从理论用于实践的角度论证了 HNC 理论可靠性和他们对进一步延伸和完善这一理论的思考。在这里我想再从另外一个角度谈谈 HNC 理论的启示性。

说到这里, 我不能不把话题暂时引向离耀红博士的著作稍显远了一点的地方: 我要说说隐藏在中文信息处理领域包括 HNC 在内的种种理论和方法背后的东西。也许这是为了把问题说得清楚些所需要的。

人文社会科学和自然科学一样, 从 18 世纪起, 随着西方工业化国家的殖民“运动”向全世界扩散, “欧洲中心论”形成, 其基本理念和研究问题的方法也逐渐成为“普世真理”。自以牛顿为代表的现代科学和以笛卡儿、伏尔泰和休谟等为标志的启蒙运动兴起(而二者间又有着极其密切的关联, 实际是一个社会的两个方面), 理性主义占据了君临天下的绝对统治地位。

现代科学和启蒙运动的确把人类对客观世界的认识推进到一个新的高度, 极大地提高了人类的生活水平; 同时随着科学的不断进步, 现代科学的一些基本理念, 例如人类自有了现代科技之后就会不断进步, 现代技术是推动人类进步的最大动力, 人类的终极追求是享受世俗的幸福等, 几乎成了欧洲人一成不变的信仰。在自然科学领域, 人们对以下这些结论坚信不疑: 客观世界是可认识和掌握的, 因为一切都是有序的, 不可逆转的; 客观的实在及其演变最终都可以归纳为定义和定理, 一切都应该是标准化的; 人和宇宙都和机器一样, 可以分成若干部分, 只要把各个部分研究清楚后复合起来就可以得到其整体 (即所谓还原论), 等等。这些理念和认识被西方人称之为“理性主义”。

理性主义指导下的科学研究, 始终沿着二元论的道路前进。从自然科学与人文社会科学分家以后, 逐渐分化出数学、物理学、化学、生物学等独立学科; 随着每个学科的发展, 研究变得越来越精细, 又不断地出现分支学科, 有些分支蔚然成为“大国”, 往往其显赫程度远远超过其母体。然而在各分支“大国”越来越精细的同时, 分支之间的综合研究却越来越薄弱了。

但是, 对理性主义的怀疑其实自理性主义形成之日起就开始了。休谟、洛克、卢梭都曾挑战流行于欧洲的上述那些观念和习惯。近代以来, 随着牛顿的运动定律和引力定律等学说被爱因斯坦以及宇宙大爆炸和黑洞理论等新的学术成果突破, 西方很多学者对自启蒙时代以来的“传统”, 包括其基本观念和具体方法产生了怀疑。首先在文学和哲学等领域出现的解构主义、后现代主义等流派从一定意义上说就是由这种怀疑萌发出来的。这也就是为什么 1988 年几十位诺贝尔奖获得者聚集巴黎并发表宣言, 指出“人类要在 21 世纪生存下去, 必须回首 2500 多年前, 去汲取孔子的智慧”。“孔子的智慧”应该不仅指原始儒家的伦理学、政治学, 也包括了中华文化的一元论、整体论、有机论、理性与经验融合论 (此“经验”不能等同于“经验主义”) 的哲学观。

现代学者对启蒙时期主流思想提出怀疑的背景和动机与休谟、卢梭那时已经有了很大不同, 一是有近代许多自然科学成就作为证明, 一是启蒙时期的理念和方法所促成的成就, 负面后果越来越清晰而严重。在社会伦理层面的影响, 诚如美国杰出的历史学家斯特龙伯格所说:

“在旧的秩序下, 人是大自然的一个组成部分, 现在人们或许再也感觉不到与大自然的那种血脉相连的亲近感了。” (罗兰·斯特龙伯格 2005)

“我们今天会倾向于说, 有些事情无疑是在“进步”, 只是要把这个词限定在某些方面: 技术变得效率越来越高, 科学知识积累得越来越多, 交通和通讯越来越快捷。但是, 生活的其他方面依然如故, 甚至变得更糟了。只有鲁莽者才会宣称, 艺术、道德或政治智慧进步了。我们

把所有的领域整合起来的能力也江河日下。有些技术进步造成了其他领域的退步，例如引起精神紧张和道德崩溃。”（罗兰·斯特龙伯格 2005a）

就科学研究而言，“进步”已经到了一往而不知复的地步——认识了蜗角之尖，却恢复不了蜗牛的整体。这无论是在对地球和宇宙的研究中，还是在对人体的研究中都时时表现了出来。“还原论”失灵了。

人文社会科学的研究路线和方法从启蒙时代起就直接由自然科学那里汲取营养、受到启发。在语言学领域，典型的例子就是大名鼎鼎的索绪尔。

索绪尔的语言学即属于理性主义，实际在提倡通过抽象的思辨建构语言的系统，因为在他看来凭着人类的经验是无法发现语言的模式和规则的。及至当代乔姆斯基出现，依然是理性主义、结构主义的老路。

无论是索绪尔还是乔姆斯基，抑或其他理性主义语言学家，先后都为语言学的建立和发展做出过大小不等的贡献，至今全世界的语言学还在享受着他们的成果；但与此同时，理性主义、结构主义的局限也日益显露出来。

结构属于事物的形式层面；以从索绪尔到乔姆斯基为代表的西方语言学所看重和着力的即语言的形式，相对严重地忽视了语言的“内容”：意义。

当语言学只是大学课堂和学者（不止是语言学家，也包括了美学、哲学和社会学家）书斋里讨论对象的时候，只注重形式的局限性还并不明显；而随着信息技术现代化和经济全球化发展势头越来越强劲，不同语言之间的交往越来越频繁、社会对语言信息自动化处理的需求越来越急迫的时候，理性主义、结构主义的局限也就越来越显著了——仅仅靠研究语言的形式，希望先把语言形式分解到最小的单位，然后“复原”为相对完整的语言（依索绪尔的理论，则为“言语”），恐怕是不可能的了。这种思路对于处理当初就是为之设计的语种至今尚且乏力，那么面对语言的“形式”与西方语言有着巨大区别的汉语，就更难以达到预期的效果了。尽管中外专家都在不断地向语义“靠拢”，例如格语法、配价理论等的被引进信息处理领域，但是看来至多也只能做到词（包括短语）处理，做出供今后人们使用的电子工具，至于如何进入句处理、使用那些电子工具又可能走向何方、能够走多远，也还在未知之列。

为什么会如此？这是因为，语言是介乎主观世界与客观世界的现象，极其复杂；语言之成为语言，是因为它表达的是人类思想和感情，思想感情的复杂微妙、使用语言的环境的复杂微妙，以及语言民族性的复杂微妙，决定了语言、尤其是语义的复杂与微妙。这种复杂与微妙主要地体现于语义上，而不是体现在语言的形式上。要使对另一种语言毫无语感的人们和没有感知语言环境能力的计算机掌握和处理语言，只讲或主要讲“形式”，是不行的。

出路何在？我认为，应该回过头来，观照语言的整体；应该以语义为核心；应该研究人是怎样学会语言的，让计算机模拟人掌握和运用语言的过程。例如，

当人之初掌握概念之时, 是一个个概念地学, 然后知其异同, 进行聚合的加工呢, 还是先学会“大概念”(并不等同于“种概念”), 然后再逐一学会与之相关、相近的概念? 概念与概念之间是怎样联系起来的? 要获得这种联系的情况, 固然可以利用统计的方法得出概率最高的配伍(临时借用一下中医的词儿吧), 但是研究者的内省, 辅之以历代人们的语言感知经验可能更为重要。(顺便说一句, 从这个角度看, 中国古代之所以出现“雅学”系列的训诂书, 并不是偶然的。)说到这里, 已经在相当程度上离开了二元论、理性主义和结构主义(但是这里所说和德里达对索绪尔的批判不是一回事), 越来越靠近中国的哲学基本理念了。当然, 这并不意味着拒绝在中国流行了几十年的西方计算语言学及其成果, 如果是那样, 我们就成了傻瓜。事实上, 像 HNC, 就利用了索绪尔的符号体系说, 继承并发展了摩里斯语法、语义、语用三层面说。中西的结合, 首先应该要在文化的底层(哲学)上接触、结合, 然后才有可能实现原创性的突破, 这一论断大概是没有多少疑问的。

内省与感知可能人言人殊, 是否有规律可寻? 我想是可以寻找到规律的。因为虽然客观事物是无限的, 从理论上讲, 词和词义也是无限的, 但是每个人所掌握的词和词义却是有限的; 汉语的句式应该也是无限的, 但其类别是有限的; 词与词、句与句、句群与句群之间的关系(联系)是无限的, 但是这种关系的类别是有限的; 语言环境是无限的, 但人类交流时利用和感知到的语言环境的要素是有限的。从无限的领域里寻找有限, 以有限控无限, 至少在理论上是说得通的。另外, 西方语言学理论给人(至少是给我)以语言的一切都是可以认识和把握的信心。但我始终对此持怀疑态度, 觉得事物的无限和人类认识能力的有限, 这一矛盾恐怕是永恒的。换句话说, 我们对语言研究、中文信息处理的期望, 都不应该超越人和计算机能力的极限。朝着绝对真理不断地前进, 获取力所能及的认识, 这就是我们的任务和目标; 不要指望我们这一代、下一代、再下一代、下 N 代能够达到终点, 这应该成为我们的座右铭。

以上这些想法在我的脑子里已经盘桓了很久, 而且自认为与 HNC 理论有相通之处。但从未安静而踏实地和黄曾阳先生沟通过; 我不是研究普通语言学的, 语言学理论学养不深, 想法也不成熟, 深恐谬种流传, 所以不敢形诸文字, 私下里也只是在给博士生上课时结合着具体语言材料星星点点地作些阐述。而在读晋耀红博士的这部著作时, 颇有“于我心有戚戚焉”的感觉。

几天前, 就在我开始写这篇序的时候, 中国科学院声学研究所、北京师范大学中文信息处理研究所和大正语言知识处理研究院联合举办了第三届 HNC 理论研讨会。黄先生在会上发表了题为“一位形而上老者与一位形而下智者的对话”的讲演。他说到:

“休谟先生当年的深沉感慨依然如故: 科学被奉为至尊, 而忘记了

科学难免有误的基本原则，争辩的反理性特征 300 年来几乎毫无变化，实质上是‘无须争辩的丛林法则’在继续横行无阻。”虽然“伟大的哲学家和科学家都诞生在欧洲，但整个西方已经背上了沉重的包袱，他们不可能再轻装前进了。新的迷信弥漫在西方的上空，从自由和民主，……到逻辑和语法，几乎无所不在。”

“HNC 理论是形而上思维的一项成功试验。其形而上思维的个性特征还没有形成系统的认识，需要大力培育。没有这一培育的创新成果 HNC 技术很难跨越知识方程求解的巨大难关。”

“(形而上老者与形而下智者的)对话认为，HNC 与相关学科的交流首先是形而上层次的交流，但当前这些学科都淹没在生存压力引发的论文浪潮之下，无暇抬起头来仰望一下天空。”

我听了他的讲演才知道，原来他在创建 HNC 理论的同时也一直在苦苦地作着哲学的思辨，而对西方科学与哲学现状的评价、对当前我国语言学界和中文信息处理界尚缺形而上思辨的感慨，竟都与我如此相近（许嘉璐 2000, 2005a），莫非世上确有灵犀一点通之事？这样说，或许我对语言这一相对微观世界的一些想法之所以与 HNC 理论有近似处，也就不是偶然的了。当然，这只是我的感觉，黄先生未必首肯。

说到这里倒引起我的几点担心。一是黄先生的高足们，如耀红、传江博士等，能不能从其师思考的高度和角度理解，进而发展 HNC 理论？一是 HNC 理论显然受到中国古代训诂学的影响，这一点在黄先生及其弟子们的论著中是隐性的，研究训诂学的学者如何能从中嗅到自己的专门学问所助长的花的香气，以启发今后研究的方向和思路？一是业中人士能不能从对西方科学和语言学的反思中悟出中国的语言学和中文信息处理应该另辟蹊径，同时不放弃从西方语言学和计算语言学的成果中吸取有用的东西？

情况的确像黄先生所忧心的那样，许多人“无暇抬起头来仰望一下天空”；但我是乐观的，理性主义对中国学术界的绝对性时代迟早要过去，这不是人们愿意不愿意的问题，而是中国的科学技术要走出一条自己的路，要可持续前进，就必须“仰望一下天空”，必须进行形而上老者与形而下智者的对话。君不见这一迹象已经在不止一处显现了吗？

许嘉璐

2005 年 12 月 31 日夜
改毕于日读一卷书屋

参考文献

罗兰·斯特龙伯格 .2005. 西方现代思想史 . 北京：中央编译出版社：44

罗兰·斯特龙伯格 .2005a. 西方现代思想史 . 北京：中央编译出版社：131

许嘉璐 .2000. 中译本序 . 见：奥格登，理查兹 . 意义之意义 . 北京：北京师范大学出版社

许嘉璐 .2005. 语言文字学论文集 . 北京：商务印书馆：389~396

许嘉璐 .2005a. 呼唤民族哲学的觉醒 . 见：未愆集—许嘉璐论文化 . 贵州：贵州人民出版社：215~217

序 二

在苗传江博士的《HNC（概念层次网络）理论导论》出版的前夕，晋耀红博士完成了《HNC（概念层次网络）语言理解技术及其应用》的撰写，翻阅着这近 400 页的长篇专著，我不禁想起“双剑合璧，所向无敌”的武林佳话。

HNC 理论与技术的双剑合璧之战已经走过了 8 年的艰苦历程，苗著（将简称《理论》）和晋著（将简称《技术》）分别是这 8 年探索历程的理论和技術总结，但是，它的意义又远远超出了通常意义下的总结，因为，它是 HNC 探索进入了催生交互引擎这一新阶段的标志。

双剑合璧的武林佳话不过是神话，HNC 的双剑合璧之战（即交互引擎的研究与开发）曾被人们视为神话，我也同样深感疑虑。因为交互引擎的实现虽然在理论上不过是三组 HNC 方程式的求解，但这是知识方程的求解，而不是微分方程或代数方程的求解。因此，它将面临一系列前所未有的巨大挑战，HNC 能够应对这些挑战么？《理论》和《技术》是对这一重大科学问题的第一份答卷。

把交互引擎的研究与开发变成一个知识方程的求解问题是先验理性思维的一项新尝试，康德先生曾应用这一思维方式开创了形而上学的新局面。HNC 追随康德先生的思路，试图在大脑思维奥秘的局部领域，即自然语言理解领域（这一领域大约只占大脑思维全部奥秘的 1/10），打开一个新局面。开启这一新局面的钥匙只能是康德先生所定义的验前知识的获取与应用，而绝不可能存在什么别的捷径。诚然，计算机强大的计算能力可以形成自己的经验理性，这种经验有可能在抽取自然语言文本蕴涵的某些知识方面获得应用，但绝对无益于计算机自然语言理解能力的提高，如果对此存在幻想，那无异于指望把全世界的盲人动员起来，就可以追捕到容貌已知的逃犯。

HNC 认为，自然语言理解的全部验前知识构成语言概念空间，这一空间的数学结构可以用四组知识方程式来描述。这四组方程规定了交互引擎必须具备的三级理解提升过程。第一级提升是语句的理解，HNC 的专业术语叫句类分析，第二级提升是句群的理解，HNC 的专业术语叫语境单元萃取，第三级提升是篇章的理解，HNC 的专业术语叫语境生成。这三级提升对应着自然语言的三级理解过程，HNC 把每一级理解过程归结为对一组方程式的求解：语句理解对应着句类表示式（HNC2）的求解，句群理解对应着语境单元表示式（HNC3）的求解，篇章理解对应着语境框架表示式（HNC4）的求解。

显然，句类表示式（HNC2）的求解是自然语言理解的第一道难关，《理论》

全面阐释了突破这一难关的验前知识准备（HNC 术语叫句类知识），而《技术》的第一部分则全面阐释了突破这一难关的验前知识运用。这里需要强调指出两点：第一，实际语言的语句理解仅依靠验前知识是不够的，还需要运用大量的验后知识；第二，语句的理解经常不可能在语句的范围内孤立完成，还必须借助于句群甚至篇章的语境知识。《技术》的第二部分对此作了比较系统的分析。

至于语境单元表示式（HNC3）和语境框架表示式（HNC4）的求解，《技术》从应用的角度给出了一个素描，它构成《技术》的第三部分。

如果说《技术》第一部分的阐释遵循了验前知识第一的思路，那么，第二和第三部分的阐释则没有继续遵循这一思路了，这不是作者的过失或疏忽，而是由于句群和篇章理解所依托的验前知识（HNC 术语叫领域句类知识）还处在设计阶段，相应知识库的建设还需要一段时间。

但是，在同一本专著中采取不同思路的写作方式毕竟值得商议，这里引用唐宋八大家之一的曾子固（巩）先生在《赠黎安二生序》里的一段话来表达我的感想，曾先生的原文如下：

“夫世之迂阔，孰有甚于予乎？知信乎古，而不知合乎世；知志乎道，而不知同乎俗，此予所以困于今而不自知也。……然则若予之于生将何言哉。谓予之迂为善，则其患若此。谓为不善，则有以合乎世，必违乎古；有以同乎俗，必离乎道矣。生其无急于解里人之惑，则于是焉必能择而取之。”

就交互引擎的探索来说，我深信，只要坚持验前知识第一的思路，它就一定能够达到胜利的彼岸，我希望 10 年之后，《技术》将以全新的面目再次出版，那个时候，也许我们能够正式宣告，将给网络世界带来理性光辉的交互引擎终于在中华大地诞生了。

但是坚持验前知识第一的思路需要智慧和坚韧，下面引用康德先生的两段论述与作者共勉，也作为本序言的结束语。

“如果经过细致准备之后，经常又重新起头，而接近目标时，又须立刻停止下来；如果常常迫不得已要走回头路而走向另一条新的前进路线；如果各个参加工作的人不能一致采取前进的共同计划，那么我们就可深信，工作是远远没有踏进科学的稳妥途径，而不过是一种在暗中胡乱摸索而已。在这种情况下，如果我们能够找到理性所能安稳进行的途径，我们对于理性就是作出一种贡献，即令这样做时，其结果是会把原来没有经过反思而在我们的初始计划中所采取的许多东西作为无效果而予以放弃。”

“理性有不可否认的自身矛盾，而这种矛盾在理性的独断过程中也是不可避免的。此种矛盾早已败坏了形而上学体系的权威。我们要通过

一种完全与前此所用的绝不相同的方法来最后使人类理性不可或缺的一门科学繁荣兴盛起来——这门科学的枝叶是可以砍掉的，但它的根干是不能毁灭的——如果在这种努力中，我们想不为内部的困难和外部的反对所阻止，我们就得更加坚定。”



2005 年 7 月