

四岁幼儿语言的语义研究

王 彤 著

世界图书出版公司

北京·广州·上海·西安

图书在版编目(CIP)数据

四岁幼儿语言的语义研究/王彤著. —北京:世界图书出版公司北京公司,2012.12

ISBN 978-7-5100-5657-4

I. ①四… II. ①王… III. ①儿童语言—语义—研究 IV. ①H003

中国版本图书馆CIP数据核字(2013)第004976号

四岁幼儿语言的语义研究

著 者:王 彤

责任编辑:夏 丹 郭晓慧 徐 娜

出 版:世界图书出版公司北京公司

出 版 人:张跃明

发 行:世界图书出版公司北京公司

(地址:北京朝内大街137号 邮编:100010 电话:64077922)

销 售:各地新华书店和外文书店

印 刷:三河市国英印务有限公司

开 本:711 mm × 1245 mm 1/24

印 张:15

字 数:331 千

版 次:2013年2月第1版 2013年2月第1次印刷

ISBN 978-7-5100-5657-4

定 价:38.00 元

版权所有 翻印必究

南京师范大学
国家“211工程”三期重点学科建设项目
“语言科技创新及工作平台建设”

《语言科技文库》

总主编 李葆嘉

计算语言学研究系列	陈小荷主编
语义语法学研究系列	李葆嘉主编
汉语方言学研究系列	刘俐李主编
古代汉语学研究系列	黄 征主编
语言教学与研究系列	肖奚强主编
语言新专题研究系列	倪传斌主编
跨学科名作译著系列	李葆嘉主编

《语言科技文库》总序

李葆嘉

当代语言学已经进入了一个科学与技术的互补时代，信息处理水平成为衡量国家现代化水平的重要标志之一。知识世界的载体是语符系统，信息处理的根本对象是语言信息处理。与计算机的出现使得语言符号有可能成为数据处理对象相似，神经科学实验仪器设备的应用，使得在大脑神经层面探讨语言机制成为可能。这些无疑都引导语言研究走向科技化，“语言科技新思维”（李葆嘉 2001）应运而生。所谓“语言科学”包括理论语言学、描写语言学、历史语言学、应用语言学等分支学科，所谓“语言技术”指语言研究的现代技术手段，包括语言信息处理、语音实验分析，以及语言的神经、心理和行为实验分析的技术手段等。就语言信息处理而言，又可以分为语料库研制技术、知识库研制技术、知识挖掘和抽取技术、句法信息处理技术、词汇信息处理技术、语音信息处理技术、语义信息处理技术、语用信息处理技术等。

2001年5月，南京师范大学文学院创办了史无前例的“语言科学及技术系”，率先迈出了从传统文科教育范型向现代科技教育范型转变的步伐。“十五”期间，南京师大“211工程”重点学科建设项目“语言信息处理与分领域语言研究的现代化”（陈小荷教授主持），以基础平台建设、资源建设和理论探索等为主，迈出了语言科技研究的一大步。

“十一五”期间，南京师大文学院、外国语学院和国际文化教育学院联袂申报“211工程”三期重点学科建设项目。该项目以“语言科技”为引导，以“多学科交叉、跨院系整合、开放型营运”为理念，建设具有前瞻性、原创性、成长性的语言科技高级工作平台。以典型课题的工作原理为核心，进行资源开发和系统研制，拓展语

音科技、二语习得的神经机制研究、言语能力受损儿童的语言能力研究等新方向。同时造就新一代学术领军人物和培养一批高层次复合型人才，以期形成一支高水平的交叉学科团队。该项目设计，体现了工作平台建设、理论创新、应用研究、人才培养、团队建设的学科发展一体化思路。其旨趣在于，加速语言研究从传统文科范型向现代科技范型的转变，以引领 21 世纪语言科技的新潮流。

作为新兴交叉学科项目，通过教育部组织的专家匿名评审，“语言科技创新及工作平台建设”（2008 ~ 2011）获批，总投入 1 000 万元。总体而言，这一“语言科技创新”团队，分支学科齐全，专业知识互补。涵盖了理论语言学、计算语言学、语义科技、语音科技、实验方言学、历史语言学、神经语言学、二语习得研究、话语行为语言学等领域。这一期间，项目组成员获批的国家级基金项目达 20 多项。该项目理念之前瞻、实力之雄厚、工程之浩大、经费之保障，为学界瞩目。

2008 年秋，本项目以南京师范大学语言科技研究所为实施单位正式启动。主要有三大任务：建设一个领先性的语言信息科技实验室、建立一个独创性的语言科技工作平台、撰著一套有特色的语言科技文库。

从实验室方案设计到设备招标采购，再到实验室用房改造，经过 8 个月的努力，2009 年 12 月，语言信息科技实验室建成，为语言研究从传统范型向科技范型的转变提供了基本保障。该实验室划分为实验工作区、科研工作区和管理服务区。实验工作区建有语音实验与计算室、神经认知实验与计算室、课堂话语实录室三个专门实验室。科研工作区建有语义科技工作室、语音科技工作室、方言实验工作室、知识工程工作室 I（先秦词汇）、知识工程工作室 II（中古词汇）、知识工程工作室 III（敦煌俗语言文字）、语言习得神经机制工作室、语言习得中介机制工作室，以及参研工作室。管理区服务包括办公室、管理室、编辑室和交流室。出席“语言科技高层论坛暨语言信息科技实验室落成仪式”（2009 年 12 月 14 日）的专家认为，该实验室体现了语言学跨学科研究的当代性和先进性，具有整体性、科技型、开放型三个特点，处于全国领先地位，是“语言科技新思维”的又一体现。同时认为，该实验室的科研工作涵

盖了四个二级学科、四个博士学位点，有稳定明确的研究方向，有合理的设计规划和很好的科研基础；整体设计合理，功能齐备。以教育部重点实验室建设标准衡量，很多方面超过了指标。

语言科技工作平台是基于工作原理（课题定位—理论方法—技术路线—关键技术—评估方式）而建设的高级平台。一方面，从语言信息、语言知识和语言机制三个层面，围绕典型课题进行设备配置、资源建设和软件开发；一方面，将典型课题研究与工作平台建设融为一体，依据典型课题建设的子平台应具有解决同类课题的功能。

建设语言科技工作平台的目标是要实现语言研究手段的技术化和模型化，总体设计包括三个二级平台和八个子系统。

一、语言信息工作平台 1. 语义科技工作系统（李葆嘉教授主持）：基于词汇语义—句法语义的一体化研究思路，开发“人—机交互语义标注工具”，研制“深度语义标注信息库”；研制“幼儿（2~6）日常话语跟踪语料库”，完成幼儿语义系统和话语行为分析研究。2. 语音科技工作系统（顾文涛教授主持）：研制“多语言、多语境、多语用的语音语料库”，基于声学信号分析、感知实验和数学建模，完善语音韵律理论与相关技术应用。3. 方言实验工作系统（刘俐李教授主持）：完成“网络版汉语方言有声语料库”，拟定系统的可操作性语音、词汇、语法实验模型和研究方法，进一步完善新兴交叉学科“实验方言学”。

二、语言知识工作平台 1. 先秦词汇统计与知识检索系统（陈小荷教授主持）：研制“先秦文献语料库”、“专名知识库”、“汉语词汇档案库”等，开发先秦文献自动分词算法、古籍版本异文自动发现算法、同指专名检索软件工具等，完成“先秦汉语词汇统计与知识检索”。2. 中古词汇统计与知识检索系统（董志翘教授主持）：研制“中古文献语料库”、“专名知识库”、“中古汉语词汇档案库”等，开发中古文献自动分词和标注工具等，完成“中古汉语词汇统计与知识检索”。3. 敦煌俗语言文字统计与检索系统（黄征教授主持）：研制“敦煌文献资料库”、“敦煌文献俗词语档案库”，开发相应工具，完成“敦煌文献资料与知识检索”。

三、语言机制工作平台 1. 二语习得的神经机制研究系统（倪

传斌教授主持): 研制“英语受蚀词汇库”等, 基于行为学、脑成像和脑电三维度模型, 进行中国人英语习得与磨蚀的神经机制研究, 完成“基于神经机制的英语个性化学习分析系统”。2. 二语习得的中介机制研究系统(肖奚强教授主持): 研制“留学生汉语口语中介语语料库”, 基于中介语理论、对比分析理论、偏误分析理论以及二语习得影响因素等, 完成“留学生汉语习得的中介机制研究”。

这一工作平台, 既是科技研究平台, 也是人才培养平台, 即一个现代化的科学研究和人才培养工作体系。

作为本项目的文本成果, 《语言科技文库》包括计算语言学研究、语义语法学研究、汉语方言学研究、古代汉语学研究、语言教学与研究、语言新专题研究六个系列。其总体特征为: 领域的开拓性、理论的原创性、选题的新颖性、方法的交叉性、考据的精审性、成果的应用性。在研究过程中, 除了数据采集分析、资源建设和软件开发, 更重要的还是要有新思路、新理论和新材料。陈小荷提出的先秦文献信息处理新方法, 从先秦典籍注疏文献中挖掘出用于自动分词和词义消歧的知识, 再注入已开发的古汉语分词和词性标注工具中去, 所取得的先秦古籍版本异文自动发现、先秦词汇知识自动挖掘等成果均具开拓性。李葆嘉提出的语义语法学理论和话语行为理论, 基于研制专用语料库或语义信息库和技术手段, 开拓了语义网络建构、深度语义分析和话语行为研究等新的领域。刘俐李建构的实验方言学理论和方法, 为方言学向现代科技方法的转型研究提供了新路, 并取得了一系列新成果。黄征多年来从事敦煌文献及其俗语文字研究, 古代汉语学研究系列中的敦煌文献校录整理, 以及敦煌写本字词考释、以古佚和疑伪经为中心的敦煌佛典词语和俗字研究、两汉声母系统研究等新见迭出。肖奚强基于汉语中介语语料库的二语习得研究, 在对外汉语教学研究界已经产生了影响。钱玉莲的汉语介词与相应英语形式比较研究等专著各有亮色。倪传斌依据语言测试和认知实验等数据, 从行为学、生理学和语言学三个层面分析影响中国英语学习者外语磨蚀的相关因素。刘宇红基于隐喻的理论探讨, 对各类隐喻形式的结构、特性和解读规律进行了多视角的深入探讨。

《语言科技文库》所收论著, 由作者在2008年12月申报选题,

2011年始逐步完稿。系列主编审读了书稿，主要就其学术价值、章节安排、内容关联、行文表述、图表绘制等方面，提出审阅意见。此后，作者们对书稿又进行了修改和润色。《语言科技文库》的作者，大多数是具有博士学位的年轻教师。对于我们这些20世纪80年代走进语言学研究领域的而言，出版论著可能已不足为道。然而，对于年轻学者而言，其论著的出版既是几年来研究的结晶，也是对其继续探索的促进。换言之，“211工程”重点学科建设的目的之一，就是为年轻教师搭建一个可持续发展的科研和教学平台。学科带头人的主要任务之一就是提携后进。

尽管从根本上来说，科学或学术研究是一种个人的探索行为，然而复杂问题的研究，无疑需要群体协作。“学科建设”或团队合作模式，是20世纪90年代后期出现的一个新概念。这种模式涉及总体规划、多方协调，是需要付出精力和心血的。2008年，通过投票方式推举我担任该项目总负责时，就意识到自己成了一个“劳动班委”。2009年，前往安徽大学拜访黄德宽教授时，曾谈到“学科负责人的任务就是规划设计，争取项目经费和提供科研设备设施”，得到黄教授的赞许。2010年，申报江苏省高校哲学社会科学重点研究基地时，评审专家柳士镇教授提问的“作为一个交叉学科项目，各学科之间的协调是怎么考虑的，有什么做法”，可谓一语中的。作为后学，深知交叉研究之艰、学科整合之难。相关学科之间的整合协调需要借助行政机制，但凭借行政方式并非就能完成。当时的回答是，目前做到的是建成了一个可以合作研究的场所，至于学科之间的进一步沟通合作应有较长过程。有一点很明确，只有通过交叉项目，相应学科才能渗透，合作者才能逐步磨合。我们只是在一步步探索。

十一五期间的“211工程”建设项目即将完成，但是学科建设的任务并没有结束。2010年，“语言信息科技研究中心”被评审为江苏省高等学校哲学社会科学重点研究基地，为“语言科技”这一交叉领域注入了新的建设活力。重点研究基地建设，除了“跨院系整合、多学科交叉、开放型运行”理念，需要凸显“合作性攻关”。围绕交叉性项目，实施计算语言学、语音科技、神经语言学、语义科技等力量的联合攻关计划。只有通过全面开放以及和与国内外同

行的合作交流，才有望建成具有影响的语言科技研究、人才培养和学术交流基地。

十年前，我（2001）曾写道：“语言科技”的内涵是以理论研究为指导，以描写研究为基础，以应用研究为枢纽，促使语言研究向计算机应用、认知科学和现代教育技术领域等延伸，沟通文理工相关学科以实现语言研究过程及其成果的技术化。“语言科技”的外延为语言工程科技、语言教育科技和语言研究科技。其中，“语言研究科技”是将语言研究活动与资源建设、软件开发相结合，其目标是实现语言学自身的科技化。还应包含语言实验、数据处理这些实验语音学、神经语言学研究的科技手段。

虽然语言学家不可能也不必要都转向语言计算或实验研究，尽管描写、考据和内省始终是最基本的方法，但是具有一定的语言科技意识却非常必要。语言学家只有了解有哪些可供利用的资源、软件或仪器，才能提高其研究深度、精度和效率。语言学家也只有了解到信息处理的语言研究需求，才有可能为之提供可资应用或参考的基础成果。“语言科技”是21世纪语言学研究的潮流。

此为出版缘起。是为总序。

2011年8月谨识于南都

李 序

一

2001年，我提出层级性元语言系统的研究理论和方法，将元语言系统分为：词汇元语言、释义元语言、析义元语言和认知元语言四个层面。其中提出：

词汇元语言系统是处于词汇功能层面用来进行日常交际的限量词汇集，也可以称之为“习得性元语言”，是任何个体语言能力继续发展的基础。其主要功能是儿童语言发展能力的检测标准和语言教学中的择词依据。

无论是从群体语言的发生发展，还是从个体语言的习得发展，词汇量总是从少到多。之所以将最低限量词汇看作一种元语言，是因为最低限量词汇不仅反映日常熟悉的基本事物、基本动作、基本性质以及基本关系，可以满足日常语言活动的需要，而且也是习得或理解其他词汇的基础，或者分化为其他衍生词语的基本词元。

相对于庞大的整个词汇系统而言，最低限量词汇天然具有“元”性质。

词汇元语言系统的研究，可以基于成人日常会话语料加以抽取。如唐萍的《建湖方言词汇研究》（南京师范大学汉语言文字学专业，2004年硕士论文），定点跟踪（2003年3月~6月）采集建湖方言日常会话语料，转写成50余万汉字的语料库，通过词频分析，抽取中频以上的词，建立了《成人日常话语基本词表》（2028个）。或者通过基本词表和高频词表的交集加以抽取。如安华林的《现代汉语释义元语言研究》（南京师范大学语言学及应用语言学专业，2004

年博士论文), 选取五种具有代表性的词表抽取共现词, 建立了《成人通用受限词表》(1400余)。然而, 最重要的途径应是基于幼儿语料加以研究, 而当时找不到一定规模的幼儿语料库。

2005年, 开始思考研制幼儿日常会话自然语料库。直到遇到王彤, 以她的女儿为采集对象, 12月才正式着手。到2010年, 所建成的《幼儿(2~6岁段)日常会话自然语料库》(李葆嘉主持, 400万汉字), 包括6个分库、14个子库。

1 两岁段日常会话自然语料库

1.1 DD 两岁段日常会话语料库 (薛松馨负责)

话语采集: 幼儿母亲 (2007.10.24 ~ 2008.10.06), 录音时长: 2221分钟。文本转写 (按工作量排序): 薛松馨、王婷婷、钱韵然、孙杨、施书宇、徐丽、李文涛。合计44万汉字。

1.2 FF 两岁段日常会话语料库 (薛松馨负责)

话语采集: 幼儿母亲 (2008.08.07 ~ 2009.08.06), 录音时长: 698分钟。文本转写: 薛松馨、马子恩、王瑶、李小敏、黄晓晓、徐丽、钱韵然、施书宇、孙杨、王婷婷、宋纯、李文涛。合计9.7万汉字。

1.3 Beijing 两岁段会话语料库

语料来源: CHIDES 中的共享语料 (香港中文大学谭霞灵, 1991.08 ~ 1991.10), 录音时长: 未知。文本转写: 嵇红英、祁文慧。合计19.8万汉字。

1.4 Nanjing 两岁段会话语料库

语料来源: CHIDES 中的共享语料 (华东师范大学周兢, 1998.08 ~ 1998.10), 录音时长: 540分钟。文本转写: 嵇红英、祁文慧。合计4.8万汉字。

2 三岁段日常会话自然语料库

2.1 XX 三岁段日常会话语料库 (嵇红英、祁文慧负责)

话语采集: 王彤 (2005.12.20 ~ 2006.11.13), 录音时长: 5184分钟。文本转写: 嵇红英、祁文慧。合计62万汉字。

2.2 MM 三岁段日常会话语料库 (薛松馨负责)

话语采集: 幼儿母亲 (2007.09.15 ~ 2008.08.26), 录音时长2142分钟。文本转写: 薛松馨、钱韵然、徐丽、孙杨、施书宇、李

文涛、王婷婷。合计 22 万汉字。

2.3 Beijing 三岁段会话语料库

语料来源：CHIDES 中的共享语料（香港中文大学谭霞灵，1991.11 ~ 1992.01），录音时长：未知。文本转写：嵇红英、祁文慧。合计 14.6 万汉字。

2.4 Nanjing 三岁段会话语料库

语料来源：CHIDES 中的共享语料（华东师范大学周兢，1998.08 ~ 1998.10），录音时长：480 分钟。文本转写：嵇红英、祁文慧。合计 6.5 万汉字。

3 XX 四岁段日常会话自然语料库（王彤负责）

话语采集：王彤（2006.11.14 ~ 2007.11.13），录音时长：8 243 分钟。文本转写：王彤、陈瑞琴、李玲、薛松馨、张彩琪、姚志勇、徐夫梅、黄小莉、朱振华、王悦、颜明。合计 89 万汉字。

4 五岁段日常会话自然语料库

4.1 XX 五岁段日常会话语料库（王彤负责）

话语采集：王彤（2007.11.14 ~ 2008.11.13），录音时长：6 018 分钟。文本转写：王婷婷、李文涛、施书宇、陈秀利、钱韵然、孙杨、李尧、徐丽、薛松馨。合计 62 万汉字。

4.2 QT 五岁段日常会话语料库（嵇红英负责）

话语采集：嵇红英（2006.04.01 ~ 2006.06.30），录音时长：1 200 分钟。文本转写：嵇红英。合计 18 万汉字。

4.3 YY 五岁段日常会话语料库（嵇红英负责）

话语采集：嵇红英（2006.07.01 ~ 2006.08.30），录音时长：180 分钟。文本转写：嵇红英。合计 2 万汉字。

5 XX 六岁段日常会话自然语料库（李尧负责）

话语采集：王彤（2008.11.14 ~ 2009.11.13），录音时长：3 936 分钟。文本转写：黄晓晓、马子恩、李小敏、王瑶、宋纯、薛松馨、李文涛、徐丽、王婷婷、钱韵然、施书宇、孙杨、陈秀利。合计 42 万汉字。

6 幼儿话语行为效能信息库（李尧负责）

选取语料 175 万汉字（2 岁段 43 万、3 岁段 59 万、4 岁段 20 万、5 岁段 28 万、6 岁段 25 万）。标注人员：王瑶、黄晓晓、宋纯、

马子恩、王瑜婷、段潇雪、马琦敏、匡海波。

先后参加《幼儿日常会话自然语料库》文本转写的 04、05、06、07、08、09 级硕士生，以及 05、07、08 级博士生，近 30 人。

在研制幼儿语料库的过程中，牵涉到幼儿年龄的计算问题。在以往的幼儿语言研究中，通常将“满周岁”定为“一岁”，以之类推年龄段。由此造出一个“零岁”，用来指称“第 1 个月第一天到第 12 个月最后一天之间的幼儿年龄”，而将“第 13 个月到第 24 个月之间的幼儿年龄”称之为“一岁”。一方面，一个 12 个月之内的婴儿，居然没有岁数，显然违背日常知识；一方面，一个进入第 13 个月的幼儿，怎么可能还是一岁呢？实际上已经进入两岁段了。之所以如此，盖因为持有“零岁”概念的人，搞不清“一岁”（处于一岁段之中）与“一周岁”（已满或超过一岁段）之间的区别。计算岁数的“出生年龄”或日常年龄，其依据是个人的出生时间和所活时长，自出生起应为一岁段。作为日常计岁的“一岁段”，有以下几种基本含义：①一岁段，即处于出生的第一天到满第 12 个月的最后一天的这一年龄段之中；②一周岁，即满 12 个月的这一天（生日）；③已满或超过一周岁，在第 13 个月到 24 个月之间。如：

(1) A：这孩子一岁吧？ B：11 个月了。

(2) 他今天过一岁的生日。

(3) 他已经一岁（多）了。

在日常年龄计算上，当以中国传统的“虚岁”为准。所谓“足岁”，即“年满多少岁”，通常用于一些专门场合，作为区分界限的年龄。如：

(4) 年满七周岁可以上小学。

(5) 学校必须组织当年 12 月 31 日以前年满十八周岁的男生进行兵役登记。

区分年龄的界限不应作为计算日常年龄的标准。因此，《幼儿（2~6 岁段）日常会话自然语料库》的年龄段，分别是：2 岁段（幼儿的 13~24 个月）；3 岁段（幼儿的 25~36 个月）；4 岁段（幼儿的 37~48 个月）；5 岁段（幼儿的 49~60 个月）；6 岁段（幼儿的 61~72 个月）。

此处，我要强调的是，嵇红英是研制最初几个语料库中，转写花费时间和精力最多的（转写总数估计在100万汉字），也是我所指导的研究生中最早从事幼儿语言研究的（《儿童语料库建设与幼儿词汇语义分类》，南京师范大学语言学及应用语言学专业，2007年硕士论文）。研制自然语料库非常辛苦，其制作权务必尊重。在用这些语料库做学术研究时，使用者务必一一交代研制者姓名（包括负责、录音，以及文本转写、校对人员），而不能一字不提，有隐匿之嫌。

二

从2001年以来，我所提出的与幼儿语言研究的相关理论主要有：

1. 语义语法学理论（李葆嘉2001a；李葆嘉2001b；李葆嘉2003a；李葆嘉2005b；李葆嘉2007b；孙道功、李葆嘉2009；李葆嘉2009）。

2. 层级性元语言理论（李葆嘉2001a；李葆嘉、安华林2001；李葆嘉2002；李葆嘉2003a；李葆嘉2003b；李葆嘉2005a；李葆嘉2005b；孙道功、李葆嘉2008；孙道功、李葆嘉2010）。

3. 个体语言成长论和语义网络自建构理论（李葆嘉2001a；李葆嘉2003a；李葆嘉2003b；李葆嘉2005c；李葆嘉2007b）。

4. 基于义征缠绕机制的语义网络理论（李葆嘉2007b第十四章，章婷2007，孙道功2008）。

5. 意象的颗粒—图式化与表达的词汇—句法化互动理论（李葆嘉2008b，祁文慧2008）。

6. 语境符号与语言符号的置换理论（李葆嘉2008a第四章）。

7. 话语表达的生成机制（话语表达在认知主体和认知对象之间，人们凭借认知策略形成基于认知经验的事件图式，再凭借建模策略形成基于事件图式的语义模式，然后凭借语用策略进行语义句法范畴排序，形成基于语义模式的话语表达）（李葆嘉2008b）。

8. 语义结构的稳态性与话语表达序列的可选性（李葆嘉

2008a 第四章；李葆嘉 2008b)。

9. 基于语境 + 语用的语义句法范畴排序理论 (李葆嘉 2008a 第四章；李葆嘉 2008b)。

10. 话语行为效能范畴系统 (李葆嘉 2006；李葆嘉 2007b；李素琴 2009；李尧 2011)。

这一系列见解，有些是在幼儿语言研究之前提出的，有些则是幼儿语言研究过程中冒出来的。同时，学生就研究中提出的疑难问题，促使我去进一步思考，提出相应的理论和研究对策。应当说，许多见解体现了研究过程中的师生互动。

基于以上理论背景和幼儿日常会话自然语料库，研究生完成的幼儿语言研究的学位论文有：

1. 嵇红英《儿童语料库建设与幼儿词汇语义分类》(2007, 硕士论文)；

2. 祁文慧《三岁幼儿语言的语义研究》(2008, 博士论文)；

3. 薛松馨《两岁幼儿的基本词汇和语义结构研究》(2011, 硕士论文)；

4. 施书宇《幼儿名核结构的语义研究》(2011, 硕士论文)；

5. 李 尧《幼儿话语行为效能研究》(2011, 博士论文)；

6. 王 彤《四岁幼儿语言的语义研究》(2012, 博士论文)。

这些学位论文的研究成果，包括：

1. 两岁、三岁、四岁幼儿语言的基本词汇

2. 两岁、三岁、四岁幼儿语言的词汇元语言系统

3. 两岁、三岁、四岁幼儿语言的析义元语言系统

4. 幼儿名核结构的语义范畴以及结构类型 (析义元句法)

5. 幼儿拿核句、给核句的语义网络建构

6. 幼儿话语行为效能范畴系统

接下来，还要继续完成的是：

1. 《五岁幼儿语言的语义研究》

2. 《六岁幼儿语言的语义研究》

3. 《幼儿动核结构的语义研究》

4. 《幼儿日常会话的情境与主题分析》

5. 《幼儿日常会话的语言计量研究》

除此之外，目前正在启动的是“幼儿话语的话说结构分析”，基于已完成的汉语话说结构功能范畴系统及其匹配框架（邱雪玫《汉语话题语－说明语结构研究》，2011 博士论文），拟研制首个大规模的“幼儿话说结构信息库”，开发“话说结构功能范畴标注工具”等。当然还有一个专题需要拓展，那就是“幼儿话语行为的情境及其非语言符号研究”。

三

2007 年秋，王彤开始攻读语言学及应用语言学专业博士学位，她原来的专业是日语教学与研究。她曾写过一篇《这笔财富将陪伴我终身》（载《随园情缘》，南京师范大学出版社，2012），研究幼儿语言道路上的艰辛与欢乐已跃然纸上。由于在职攻读博士学位，由于语料库研制的任务繁重，由于语义研究的艰难，王彤的博士论文前后耗时 5 年。

在学术界有一种看法，研制语料库不能作为学位论文选题的研究主体，理由就是语料库是基础性工作，还不是正式的研究。问题在于，如果不作为学位论文选题的研究主体，那么谁来完成如此耗时费力的语料库呢？而如果没有这样的语料库，又如何就相关领域的问题展开下一步的深入研究呢？显然，语言资源建设、语言工具开发，这些都是语言研究科技化时代的成果形式。如果拘守传统文科研究的成果形式，显然不适应语言研究科技化的需求。

就活的语言研究而言，在攻读学位的几年时间内，能够采集一定规模的日常会话并建库（如 50 万字），与主要就文献资料撰写一篇几万字的学位论文相比，无论付出的精力与时间，还是对学术研究的贡献，都不可同日而语。因此，我以为，语言学专业学位论文选题的研究主体可以是研制一项语言资源。从语料库设计，到语料的采集、转写、加工，再到配上检索系统，得出一些基本数据，不但可以作为毕业论文申请学位，而且是一项造福学界的科研成果。

另外，也不知出于何种理由，一些学者总是强调群案研究，而

对个案研究似乎不以为然。好像群案研究的结论才有普遍性，而个案研究的结论没有普遍性。这牵涉到关于学术研究的几个基本问题。

第一、个性与共性的关系。根据哲学的基本原理，共性寓于个性之中，普遍性存在于个别性之中。所谓“一叶知秋”、“一滴水可以看见大海”。不同的幼儿当然存在一定的个体特征，但只要是一个正常的幼儿，其语言成长轨迹中必然包含着普遍性特征。

第二、横向与纵向的关系。群案研究取样个体多，时间长度短，通过横向比较得出一些结论。但其弊端在于，许多重要细节无法深入，容易流于表面化的数据。个案研究着眼于时间跟踪，取样个体单一，通过纵向比较得出一些成长轨迹。但其弊端在于，有可能放大个体差异，尚须增加不同个体的数据。尽管理想的方法应是纵横研究法，但是在横向研究盛行的情况下，纵向研究或个案研究显得更有价值。反之，有了个案纵向研究，再以之为基础的横向研究才更可靠。

第三、混杂与单一的关系。一些学者，依据书面语语料库（有一些充其量只能称之为“语料堆”）的传统要求，认为语料越多越好（美其名曰：语料的海量性），甚至越杂越好（美其名曰：语料的平衡性）。从理论上讲，似乎越多的语料可能有越多的发现。从操作上讲，任何人也不可能搜集到全部语料，再多的语料仍是有限语料。由此我一直强调，基于受限语料和受限方法，进行受限研究而得出受限结果。实际上，只要语料达到一定的量（如30万、50万汉字），其研究结果就富有系统性的价值。即使增加再多的语料，也很少能发现有新的系统性价值的东西。因此，我一直主张建立单一性的专门语料库。

总而言之，在自然状态下，跟踪采集成人日常语料就非常困难，跟踪采集幼儿日常语料则难上加难。就某一个案的语料采集和加工而言，已相当耗时费力，何谈在几年内同时就几个个案的语料进行采集和加工？其实，之所以研究个案跟踪的人少，原因就在于个案研究比一般的群案研究要耗时费力多了，跟踪性研究比取样性研究要耗时费力多了。

语言的深入研究不仅需要新的理论和方法，更需要研究理念的彻底更新。