

祁文慧 著

三岁幼儿语言的语义研究

语义语法学研究系列

李葆嘉 主编

世界图书出版公司

三岁幼儿语言的语义研究

祁文慧 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

三岁幼儿语言的语义研究/祁文慧著. —北京:世界图书出版公司北京公司, 2011. 12

ISBN 978-7-5100-4188-4

I. ①三… II. ①祁… III. 儿童语言—语义—研究 IV. ①H003

中国版本图书馆CIP数据核字(2011)第258202号

三岁幼儿语言的语义研究

著 者: 祁文慧

责任编辑: 宋筱茜

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: 711mm × 1245mm 1/24

印 张: 18

字 数: 395 千

版 次: 2011年12月第1版 2011年12月第1次印刷

ISBN 978-7-5100-4188-4/H · 1277

定 价: 36.00 元

版权所有 翻印必究

《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词语文字研究,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

前 言

就个体语言的习得和发展而言，是以幼儿话语为基础，逐步发展为成人话语。国内外幼儿语言研究主要为词汇、句法研究，语义研究还刚刚起步。国内幼儿语义的研究体现了两个特点：一是基于个别语言现象创建幼儿语言习得理论；二是基于某一语言习得理论研究个别语言现象。前者的研究比较概括宽泛，缺乏系统分析，所创建的理论缺少实证基础；后者的研究比较琐碎凌乱，没有理论创建和论证，研究结果缺乏指导意义。幼儿语义研究中最大的问题是套用成人词义研究模式和词义分类框架，缺乏幼儿词义和幼儿语言的系统研究，很难真正揭示幼儿词义习得的特点和语义结构发展的规律。

根据“语言系统成长论”（李葆嘉 2003）：个体语言的成长或发展划分为基础阶段（5岁以前）、发展阶段（6~18岁）、扩展阶段（18岁以后）三个阶段。5岁以前称为“小型语言”或“幼儿语言”，18岁以前称“中型语言”或“未成年语言”，18岁以后称“大型语言”或“成人语言”。其中“小型语言”是“中型语言”和“大型语言”的内核。个体母语习得过程实质上就是一个逐步建构个体认知语义网络的过程，这一网络能为语言的理解和生成提供知识背景。因此，幼儿语义网络建构将有助于研究幼儿语言生成和理解的机制，同时也将为建构成人语义网络提供基底。

根据儿童生理和心理发展状况，可将5岁或6岁前儿童分为两个年龄段：出生至3岁为婴幼儿阶段，3岁至5岁或6岁为学龄前阶段。一般认为3岁前是幼儿接触人生的第一个年龄阶段，这个阶段的幼儿已掌握生活中常用的动词，已能保证幼儿与他人基本口语交往的正常进行，这些动词在以后的增加就不十分显著了。（陈帼眉 1989）因此，3岁前幼儿语言研究具有阶段性特点，可以作为儿童语言研究的起点，由此本研究将3岁前幼儿语言确定为研究对象。

许政援等人（吴天敏，许政援 1979，许政援，沈家鲜 1987）将 3 岁前幼儿语言发展阶段划分为：单词句/独词句阶段（1 ~ 1.5 岁）和多词句阶段（1.5 岁 ~ 3 岁）两个大阶段；多词句阶段再进一步分为简单句阶段（1.5 岁 ~ 2 岁）和复合句开始阶段（2 岁 ~ 3 岁）。周兢的汉语儿童语言发展阶段新说（1997）和皮亚杰（Jean Piaget）的认知发展阶段说（1969）认为，根据儿童语言结构能力所达到的水平，0 ~ 2 岁为儿童语言整体结构阶段，认知处于感觉运动阶段；大约从 2 岁开始，儿童逐步获得简单的语言分析结构能力（2 ~ 4 岁），进入认知的前维度阶段（2 ~ 6 岁）。可见，3 岁前幼儿语言和认知的发展紧密相关，而且各年龄段之间认知能力和语言能力存在质的不同，有必要对 1 ~ 2 岁和 2 ~ 3 岁两个年龄段幼儿词汇习得、词汇语义和句法语义的发展分别进行考察。

北京大学第一医院与美国密歇根大学、香港大学等参照母语为英语的儿童语言发育情况量表 MCD 为基本格式，基于 10 名北京市 20 ~ 22 个月幼儿父母的 6 个月报告，修订出了《中文早期语言与沟通发展量表》（普通话版）（Chinese Communicative Development Inventory Mandarin Version）。2001 年，又对 1600 多例 8 ~ 30 月幼儿语言发展情况进行调查，对《中文早期语言与沟通发展量表》（普通话版）进行了再标准化，完成了普通话量表，简称 PCDI（Putonghua Communicative Development Inventory）。2008 年，这一量表在《汉语沟通发展量表》（北京大学医学出版社，谭霞灵〔Twila Tardif〕等著）一书中正式刊出。但是基于幼儿自然语料，在幼儿语言教学、幼儿词汇语法习得研究方面的完整词表尚缺。3 岁前幼儿的词汇习得情况只散见于一些研究论著或者论文之中，且着重于幼儿的实词习得研究。

以往幼儿词汇表构建存在以下问题：1. 词语的获取不是基于幼儿语料，而是参照成人词表采用调查法获得，无法真实反应幼儿词汇掌握情况；2. 提取词语的原始语料无法获取，不能为本研究的词义分析和句模建构提供真实例句；3. 词语获取的年代较早，无法代表当前幼儿语言词汇发展的现状；4. 词类有缺漏，年龄段不全。鉴于幼儿基本词汇的提取是幼儿词汇语义和句法语义研究的基础，本研究汇总了以往研究中的 3 岁幼儿词汇表，作为基本词汇构建的比

对增补之用，以弥补基于语料提取词汇的不足。

由此，本研究所做的主要工作是：首先建立幼儿自然语料库，基于语料库提取基本词汇、分化合并义位获得基本词元；然后以义征分析为基础，建构幼儿基本词元语义场；在语义分类框架内，分析幼儿词义习得发展特点和规律，以词元语义特征为基础建立语义场，抽象句模并建构幼儿语义网络。主要基于以下理论背景：个体语义网络自建构理论、要素离散化—图式结构化理论、语言符号—语境符号置换理论。

语义语法观（李葆嘉 2003）认为，人类语言的共性是语义性。无论是语形型语言还是语义型语言，在语言生成过程中都是首先启动语义编码，语形编码只是语义类别关联化的不完全性和简单化的句法表层投影。自然语言系统是建立在认知符号的实体范畴化（聚合性语义场）和实体关联化（组合性语义场）基础上的语义网络。句法结构并非词语之间的临时组合，而是相关义场之间的语义关联。幼儿语言的习得过程，本质上就是基于认知经验的语义网络自建构过程。幼儿先是基于对象世界感知形成认知图式，然后基于认知图式形成语义结构，最后基于语义结构形成句法结构。认知图式先于语义结构，语义结构固化认知图式。在认知经验发展和符号表达习得的互动过程中，幼儿具有语义网络自建构能力。幼儿认知能力的逐步发展体现为，对认知图式构成要素离散化—图式结构化能力的逐步提高，由最初的整体认知图式向离散化认知图式发展的过程。

幼儿语言具有成长性，幼儿的认知水平发展具有层级性。认知层级的高低突出表现在，对事件图式中的组成要素的离散化和关联化能力的强弱。对事件图式的认知，幼儿最初处于整体状态，表现为事件图式的各组成要素尚未从图式中离析出来。随着认知水平的发展，幼儿才逐步具备了成人的要素离散化能力，并能通过各要素关联化形成结构模型。幼儿从未语义结构化底层，经过不完全语义结构化过渡层。要素离散化表现为词汇数量的增加，图式结构化表现为语义结构的形成和丰化。

幼儿话语从独词句、双词句到多词句的变化，体现了对语境符号极强的依赖性。基于语言符号—语境符号置换理论（李葆嘉 2007：315），无论成人还是幼儿，在语言交际中所传递的信息都是

完整的，信息结构 = 语言符号 + 语境符号。语言符号和语境符号的互补性体现为反比关系，语言符号越充分，对语境符号的依赖性越弱；语言符号越简单，对语境符号的依赖性越强。成人和幼儿对语境符号（非语言符号）和语言符号的依赖程度不同，幼儿话语经历了从语境符号向语言符号置换的过程；随着幼儿所掌握的语言符号趋于丰富，对语境符号的依赖性会逐渐减小。语境符号向语言符号置换具体表现为两个方面：一是事件过程中的要素逐步离析为词汇单位（事件图式颗粒化）；一是认知图式中的词汇单位逐步关联成结构模型（语义结构模型化）。因此，从所形成的语义结构来看，幼儿话语中的独词句、双词句和多词句可被视为单成分结构、双成分结构和多成分结构。在单成分结构中，语境符号所占比例远远大于语言符号，由此构成“语境符号 + 语言符号”复综型语义句模；在双成分结构和多成分结构中，语境符号的比例逐渐减少，语言符号的比例逐渐增加，由此构成主要以语言符号为主的分析型语义句模。

本研究第一章是对国内外幼儿语言研究的综述；第二章系统介绍国内外幼儿语料库的情况，以及本研究中的幼儿自然语料库创建情况；第三章具体介绍基于幼儿自然语料库，幼儿基本词汇表的提取、制作过程和结果；第四章通过比较母语为英语和汉语同龄幼儿词汇习得差异，以及汉语幼儿和成人基本词汇的异同，揭示母语为汉语的幼儿词汇习得特点；第五章首先介绍幼儿词元提取的原则、过程，然后详述幼儿词元语义分类元范畴的建构；第六章至第九章对四大类词元逐一分析义征、建构义场；第十章以典型动作词“拿”为例，比较了2岁幼儿、3岁幼儿以及成人日常“拿”核句语义网络，揭示了幼儿语义网络自建构能力发展的特点。

希望本研究能为幼儿语言研究以及相关领域的研究提供一些参考。第一，探索语言理解和生成机制。幼儿日常交际基本词汇是个体语言词汇增长的基础，与成人日常交际最低限量词汇存在大量重合之处。一般认为，5岁前个体语言已经基本具备了成人语言结构特点，是个体语言成长或发展的基础。基于幼儿自然语料库研究幼儿基本词元语义和句法语义，建构3岁前幼儿的分级语义网络，这将为5岁前儿童语言，甚至成人语言研究提供基础和线索。由此，对真正了解个体语言的理解和生成机制将具有极大作用。第二，丰富

语义语法学理论。本研究从基本词元提取、词元义征分析、词元义场建构、词元语义分类、语义句模抽象、语义网络建构等方面对3岁前幼儿进行系统的语义研究，将为现代汉语词汇语义—句法语义系统提供基础，并验证现代汉语基于句模，为寻找深层句模提供依据。再者，所建构的3岁前幼儿语义网络将与成人语义网络的基底相验证。通过以上研究，不仅能够发现3岁前幼儿词汇语义和句法语义的发展规律，而且将为论证人类语言的共性—语义性提供更加充足的证据，这必将极大地丰富语义语法学理论。第三，完善儿童语言习得理论。基于幼儿自然语料库，以幼儿日常感知和人类语言的共性—语义性为前提，避免了套用成人语言研究的模式。由此发现的幼儿词汇语义和句法语义的发展特点，为“要素离散化—图式结构化理论”以及“语言符号—语境符号置换理论”的提出和论证提供了实证基础。这些理论的提出不仅能为幼儿的母语学习提供理论依据，避免幼儿语言教育的成人化，提高幼儿语言教育的效果，而且幼儿语言的习得规律将为成人第二语言学习提供借鉴。

本研究主要方法如下。

1. 核心模型和受限分析相结合

这一方法的理论基础是李葆嘉（2003）提出的“内核模型”和“受限语言”的思想，以便将语言研究化繁为简。基于受限幼儿语料库，具体做了三个方面工作：一是提取了3岁前幼儿分级词汇；二是对3岁前幼儿基本词元进行义征分析和义场建构，建构幼儿基本词元语义分类框架；三是分别抽象2岁和3岁前幼儿“拿”核句语义句模，并建构语义网络。

2. 计量分析和人工干预相结合

在提取幼儿基本词汇过程中，首先使用软件程序对幼儿自然语料库分词、标注、统计，划分词频等级，提取中频以上词语作为幼儿基本词汇的原始词语；然后再予以人工干预，汰除分布率不足的词语，以及仅出现在故事讲述或者成人语言模仿中的词语；依据词表间的比对和词汇语义的对称性进行词语增补，以增强所提取词表的普适性。

3. 内省法和实证法相结合

为了避免幼儿词汇义征提取的主观化、成人化，研究中坚持人

的经验、词典释义和实证材料相结合，包括比对、提取、验证、优化等环节。幼儿词元义征分析、义场建构，以及“拿”核句句模的抽象，都需要语料和实证的多次检验。

4. 义征分析和语义场建构互动

以义征分析为基础，建构幼儿基本词元语义场；在语义场内抽象语义特征；在语义分类框架内，分析幼儿词义习得发展特点和规律。

5. 场景分析和句模抽象相结合

幼儿在语言交际中传递信息的方式经历了由语境符号向语言符号的置换过程。因此3岁前幼儿语义句模的抽象，需要借助场景的补充提取语境符号，才能明确每个词元在语义结构中所担任的成分，场景分析是义征分析、句模抽象的前提和基础。

6. 范畴抽象和句模建立相结合

依据“动作+语义角色”的语义结构模式，对其中担任相同语义角色的词元进行义征分析、语义场建构。然后抽象语义句法范畴，建立句模，为语义网络的建构打下基础。

7. 聚合义场和组合义场相结合

分别对语义角色以及其他语义标记进行语义分析和抽象，建立聚合义场；继而在“动作+内层角色+外层角色”语义结构中，寻找语义关联，抽象语义范畴，建立组合义场；结合聚合性义场和组合性义场，并抽象情态功能范畴，由此形成纵横交错的幼儿语义网络。

本研究的语料来源主要包括两部分：一是自建“幼儿自然语料库”，包括两个年龄段：1~2岁年龄段和2~3岁年龄段。1~2岁年龄段包括两个个案。一个追踪12个月，采集语音语料37个小时，转写后的成人与幼儿对话混合文本语料（A库）约43.4万字，其中纯幼儿话语文本语料（B库）约9.6万字。另一个追踪9个月，采集语音语料12个小时，转写后的成人与幼儿对话混合文本语料（A库）约9.7万字，其中纯幼儿话语文本语料（B库）约3.5万字。由于后一部分语料的采集和整理完成较晚，所以只作为今后研究完善阶段之用。2~3岁年龄段为一个个案。追踪时间为11个月，共采集到语音语料86.4小时，转写后的成人与幼儿对话混合文本语料

(A 库) 约 62 万字, 其中纯幼儿话语文本语料 (B 库) 18.5 万字。

二是从“国际儿童口语语料库”(Child Language Data Exchange System, 简称为 CHILDES) 中下载的两个语料子库: “Beijing” 和 “Zhou”。“Beijing” 是谭霞灵于 1991 年 8 月至 1992 年 1 月在北京录制, 包含 2 岁幼儿与成人对话的混合语料约 19.8 万字, 纯幼儿话语语料 3.2 万字; 以及 3 岁幼儿和成人对话的混合语料约 14.6 万字, 纯幼儿话语语料约 2.6 万字。“Zhou” 是由周兢 1998 年在南京录制的语料, 包含 2 岁幼儿与成人对话的混合语料约 4.8 万字, 纯幼儿话语语料 2557 字; 3 岁幼儿与成人对话的混合语料约 6.5 万字, 纯幼儿话语语料 6055 字。

在进行汉英幼儿词语比对时, 主要参考了两个英语词表: 一是《国际通用 MCDI 词汇量表》(MacArthur Communicative Development Inventory)。该词表由美国圣地亚哥州立大学心理学家芬森 (L. Fenson) 于 1993 年主持研制, 主要为测量英语儿童词汇的发展建立标准。CDI 分为两级: 8~16 个月为婴儿级, 16~30 个月为学步幼儿级。婴儿级词汇的参照标准是理解、词汇产生、符号和交际手势等, 学步幼儿级的参照标准是词汇的能产以及早期的语法表现。二是《哈特词表》(B. Hart, et al. 1999)。该词表由美国心理学会的贝蒂·哈特等 (Betty Hart, 堪萨斯大学) 通过调查 42 个家庭母子对话, 从每个幼儿最初能表达的 100 个句子中统计出来。词表分年龄段标注了幼儿和母亲对各个词语的使用次数, 并标注了与 CDI 词表共现词语。该调查主要用于研究 3 岁前幼儿的语言发展与家庭经济状况之间的关系。

成人基本词汇的比对, 主要是参照安华林 (2005) 的《现代汉语受限词表》。此表是《汉语词汇的统计与分析》、《现代汉语三千常用词词表》、《普通话三千常用词表》、《生活口语中前 4000 个高频词词表》、《汉语水平词汇与汉字等级大纲》五种词表的交集, 所收录的词汇具有基础性和代表性。

本研究基于幼儿自然语料库、词表比对增补和语义对称性增补, 最后提取 2 岁幼儿基本词汇 460 个; 3 岁幼儿基本词汇 1387 个; 3 岁前幼儿基本词汇 1466 个。基于《3 岁前幼儿基本词汇表》, 参照语义分化和合并原则, 获得 3 岁前幼儿基本词元 1424 个。基于幼儿的

认知特点和语感，参照词典释义，概括提取词元语义特征；采用义征分析与义场建构互动的方法，建构了幼儿基本词元语义系统。系统包括4大类15中类42小类，最深的层级为十层。

通过在语义分类框架内分析幼儿词义习得发展特点和规律，发现汉语幼儿词汇习得不存在明显的名词优势，2岁幼儿动词甚至稍多于名词，体现了幼儿最初习得的动词具有极强的意象性。比对3岁前幼儿词表和成人日常基本词表发现：首先，成人词表中也不存在明显的名词优势；其次，3岁前词表中的名词、动词的总数都已接近成人；但是3岁前幼儿词表中的名词反而多于成人，其中的原因有待研究。

基于自然语料库所提取的基本词汇表以及优化的基本词元表，对幼儿词汇的发展研究具有重要参考价值。采用义征分析—义场建构互动研究所建立的幼儿基本词元语义场，无论对于词汇语义研究还是句法语义研究，都是一项必须完成的基础性工作。“拿”核句的语义网络建构，揭示了幼儿语义网络自建构能力的发展特点，证明了幼儿语义网络的成长性，为语言生成—理解研究提供了样例。本研究的理论价值在于，初步建立了基于语义语法学理论的新型儿童语言学；其应用价值在于，可为幼儿语言教育和语言信息处理提供参考。

幼儿语言研究的关键是语料的获取和研究方法的确定。但在研究中，这两个方面难免有不尽如人意之处。首先，幼儿自然语料的采集方式、过程都存在许多不可控性，本研究所建构的幼儿自然语料库的规模有限，依据此语料库所提取的基本词汇也有待进一步参证。其次，在幼儿词元提取和义征分析中，需要制定更加具体、细化的义位分化原则，对所提取的义征再概括、抽象，以建构一个更加稳定的语义分类框架。如果能多选取几个典型动词建构语义网络，对于揭示幼儿语义网络自建构能力的发展特点，阐释语义网络建构理论、要素离散化—图式结构化理论和语言符号—语境符号置换理论等将更具说服力。如果能对幼儿语言的名核结构句进行研究，将能全面揭示幼儿语言理解和生成机制。