

山东省“十五”期间重点学科强化建设项目

鲁东大学汉语言文字学科

“语言文字理论与应用研究”文库

面向信息处理的现代 汉语语法研究

亢世勇 著

上海辞书出版社

山东省“十五”期间重点学科强化建设项目
鲁东大学汉语言文字学科

“语言文字理论与应用研究”文库编委会

学术顾问：王 宁 黄伯荣

主 编：张志毅 张绍麒

副 主 编：陈淑梅 亢世勇 吕永进

编 委：（按姓氏拼音排序）

陈淑梅 丁信善 亢世勇 李 敏 吕永进

王宝刚 肖德法 徐德宽 张绍麒 张志毅

执行主编：亢世勇

责任编辑：宦荣卿

封面设计：杨钟玮

目 录

总序	张志毅(1)
鲁东大学汉语言文字学重点学科群介绍	(5)
绪论 改变观念,抓住机遇,大力发展中文信息处理研究	(1)
第一章 语义构词研究	(13)
第一节 《汉字义类信息库》的研究与实现	(13)
第二节 基于数据库的现代汉语字、词义类分布统计 比较研究	(34)
第三节 基于数据库的汉语语义构词法初探	(46)
第二章 词类问题研究	(62)
第一节 基于数据库的现代汉语词类语法特征统计 研究	(62)
第二节 现代汉语名谓宾动词、谓宾动词分类统 计研究	(74)
第三节 现代汉语名宾动词分类统计研究	(94)
第四节 “动结宾”动词语法特点的分析	(154)
第五节 双宾动词语法特点分析	(166)
第六节 现代汉语兼类词及其区分策略	(176)

第三章 句法结构研究	(199)
第一节 词语组合所受限制的类型及其关系	(199)
第二节 组词成句所受的语用制约	(213)
第三节 动词对其所带宾语的选择	(225)
第四节 试析“写你的”——兼论句法、语义、语用三个 平面的结合	(235)
第四章 句子语义结构研究	(246)
第一节 基于标注语料库的现代汉语句子语义结构 研究	(246)
第二节 现代汉语句子语义结构系统	(262)
第三节 现代汉语句子语义结构频率的统计研究	(313)
后记	(347)

总 序

张志毅

甲申猴年,渝州叶黄,迎来了中国英汉语比较研究盛会和该会诞生十年庆典。应邀,我从东海之滨乘车经秦岭蜀道,到达了天府之国。在开幕式上,我(还有沈家煊先生和美国的根茨勒教授)作了主题学术演讲。那是空前盛会,英语界资深教授、博导、名家、权威咸集。更令人喜出望外的是,我们学科点的年轻博士解海江的《汉英颜色词对比研究》的理论之高和语料之丰,在那鸿篇巨制之林,也属上乘之作。真不愧为是他的博士论文《汉语编码度研究》一章的升华。

会后,乘船顺流而下,大三峡、小三峡、小小三峡,那世间的奇峰异岭,两岸对峙,葱茏辉映,夹出一条蜿蜒高耸的绿色长廊。它比美国的大峡谷有生机盎然之美,比意大利的威尼斯有天造地设之美,天姿地色,绝世无匹。然而,一回到海滨,这美感居然被更美的东西冲淡了,那便是我们学科的琳琅满目的、美不胜收的一批新成果,比海江的《汉英颜色词对比研究》更胜一筹,或不分轩轻。如张绍麒教授的《汉外语言对比研究报告》、亢世勇教授的《面向信息处理的现代汉语语法研究》、陈淑梅博士的《东汉碑隶构形系统研究》、王宝刚博士的《〈方言笺疏〉因声求义研究》、徐德宽博士的《现代汉语

双宾构造研究》、吕永进教授的《汉字规范形音义》等,即将由上海辞书出版社出版。还有一些重要成果,将陆续由权威出版社推出。

这批语言学与应用语言学成果及其后续研究,是很有意义的。小而言之,对我们这个省级重点学科起巩固、强化、提高、发展作用;大而言之,对中国语言学也是添砖加瓦。我们都有责任推动整个语言学的发展。因为,“这门学科的重要性完全可以同17世纪伽利略改变了我们关于物质世界的整个观念的新科学媲美”(卡西勒)。语言学的“结构主义在社会科学中所起的革新作用——如原子物理学家在自然科学中所起的作用”(列维·斯特罗斯)。“历来有一种传统:把语言学看作是这样一种活动,它能导致发现新的认识论,发现具有启发意义的革命纲领”(莫里斯·格罗斯)。(以上引文皆见伍铁平《剑桥语言百科全书·中译本序》)

我们的研究,还有待于“脱去匠气,生发师魂”。匠气,就是工匠式的缺乏创新的习气;师魂,就是大师式的重大原创的思想。木匠的产品是桌椅板凳,画匠的作品是毫无新意的图案。大师的产品是蒸汽机、电动机、日心说、相对论等等。

新学说、新理论的诞生,需要社会或民族的五个基石:一是民主自由的环境,二是多学派的哲学思想,三是科学的理论思维,四是先进的多种文化,五是创造性的教育。需要个人的十二个条件:一是多种哲学思想,二是逻辑和数理逻辑训练,三是高等数学思维、理性思维、创新思维,四是精通一二门外语,五是古文阅读能力,六是良好的母语素养及表达能力,七是占领本学科的四级理论制高点(一二三级学科和

专题的),八是广阔的视野和临界学科的知识,九是掌握多学派学术思想史,超越一个学派的狭隘视域,分清学说的新旧高低,十是优先注重演绎法、比较法、定量定性法,十一是高超的观察能力、信息处理能力,十二是超人的勤奋。对于个人来说,社会民族的五个基石是非可控因素,而个人的十二个条件中至少有十个是可控因素。我们的创新,只能在可控因素中最大限度地调动积极因素,多脱去一些匠气,多生发点师魂,多产生一点大师级或近于大师级的新学说、新理论。这样一来,我们学科和我国的语言学研究才有望同国际同行对话,同国际接轨。

愿与同事、同行、同道共勉。是为序。

甲申猴年立冬后七日

2004年11月17日

鲁东大学汉语言文字学 重点学科群介绍

“语言文字理论与应用研究”是山东省强化建设重点学科、鲁东大学(原烟台师范学院)汉语言文字学科“十五”期间进行重点学科建设的阶段性成果的总结。

鲁东大学汉语言文字学学科群是学校重点发展的优势学科群之一,目前汉语言文字学为山东省强化建设的 B 类重点学科、语言学与应用语言学为校级重点学科、中文信息处理实验室为校级重点实验室。

学科群现有专任教师 27 人,其中教授 7 人,副教授 10 人;博士 8 人,在读博士 4 人。有省级拔尖人才 1 人、享受国务院特殊津贴的专家 1 人、全国优秀教师 1 人,省级中青年学术骨干培养对象 1 人、校级中青年学术骨干培养对象 6 人;是一支高职称、高学历、年轻化的充满朝气的教学研究队伍。近年来在《中国语文》《中国外语》《语言文字应用》《辞书研究》《古汉语研究》《世界汉语教学》《修辞学习》《汉语语言与计算学报》《中文计算语言学期刊》等学术期刊上发表论文 130 多篇,出版著作 20 多部;承担国家社科规划项目、山东省社科规划项目、教育部语言文字应用“十五”规划项目、国家 973、863 项目子课题、国家自然科学基金项目子课题等项目

30 多项,获得各种科研经费 200 多万元。获得各种优秀科研成果奖 30 多项,其中省社会科学优秀研究成果奖 10 项、山东省高校人文社科优秀研究成果奖 8 项;获得优秀教学成果奖 11 项,其中国家级 1 项、省级 4 项,现代汉语课被评为山东省精品课程。1998 年和 2004 年两次被评为国家语言文字工作先进单位。本学科群现有汉语言文学、汉语言两个本科专业,汉语言文字学硕士研究生专业。汉语言文字学专业 1999 年开始招收硕士研究生,到目前共招收培养硕士研究生 60 多人,已经毕业 20 人,有 8 名分别考取南京大学、山东大学、南京师范大学、厦门大学、北京师范大学、广东外语外贸大学、上海师范大学等高校的博士研究生,其他被商务印书馆、北京外语教学与研究出版社等单位录用。除汉语言文学学院外,学科群还有烟台师范学院国际交流学院对外汉语教学部、语言文学教育研究所、普通话测培中心和中文信息处理研究所的有力支持。

本学科群形成了鲜明的特色:(1)各个学科统筹建设,互为依托,相互促进,共同发展。(2)以中文信息处理和语言教学与研究的现代化为龙头带动学科群的发展,狠抓基础理论研究,理论研究为应用服务。

学科群形成了四个稳定的研究方向:词汇语义学研究、中文信息处理研究、语言教学研究、汉语与汉字规范化研究。

(一)词汇语义学研究。本方向的特色在于吸收借鉴国内外先进理论,深入研究汉语词汇,建立汉语词汇语义学、流俗词源学理论体系,并向辞书编纂、计算机汉语词汇语义信息处理等应用方向发展。学术带头人张志毅教授,山东省专业技术拔尖人才,享受国务院特殊津贴的专家,出版专著《词与词典》、《词汇语义学》(商务印书馆),在《中国语文》等学术

期刊上发表论文 70 多篇,其中论文《现代语文学辞书的整体观》、《柏拉图以来词义说的新审视》均获山东省社会科学优秀成果二等奖,《词汇语义学》获山东省社会科学优秀成果三等奖,并被教育部确定为汉语言文字学专业研究生教学用书;还出版了《简明同义词词典》《反义词大词典》,《同义词大词典》即将由商务印书馆出版。学术带头人张绍麒教授的专著《语词符号结构探索》、《汉语流俗词源研究》分别获山东省社会科学优秀成果三等奖、山东省高校人文社会科学优秀研究成果一等奖。论文《异形词是词位的无值变体》(《语言文字应用》2003 年 1 期),在词汇规范方面有一定的影响。解海江博士在《古汉语研究》等刊物上发表《面部语义场历史演变》等语义场研究论文,其博士学位论文《汉语编码度研究》在对比语言学领域里,建立跨语言词义对比研究的编码度理论模式,获得好评。该方向现主持省级科研项目 5 项、横向科研项目多项。

(二) 中文信息处理研究。本方向的主要特色在于紧扣计算机语言信息处理的需要,开发大规模的语料库、电子词典,开展面向中文信息处理的汉语汉字研究、计算机辅助词典编纂的理论研究与实践、中文信息的提取研究等,是本学科群的龙头。学术带头人亢世勇教授为现代汉语省级精品课负责人。主持国家社科规划项目 1 项、山东省社科规划项目 1 项,和北京大学计算语言学研究所、清华大学智能技术与系统国家重点实验室合作完成国家 973 项目、国家 863 项目、国家自然科学基金项目多项,承担上海辞书出版社等单位委托项目多项。在《辞书研究》《汉语语言与计算学报》等国内外学术期刊上发表《〈现代汉语新词语信息电子词典〉的开发与应用》《〈汉字义类信息库〉的研究与实现》等学术论文

40多篇,出版著作(含词典)6部,其中《新词语大词典》(上海辞书出版社,2003)为国内规模最大的新词语词典,也是利用计算机语料库技术辅助新词语词典编纂的一项成果,代表了信息时代辞书编纂发展的方向。还开发了标注各种语言学信息的大规模的现代汉语语料库和现代汉语新词语信息电子词典等重要语言资源。获得省社会科学优秀研究成果奖1项、省高校人文社科优秀研究成果奖2项、市社会科学优秀研究成果奖2项。应邀多次赴新加坡、马来西亚、香港、台北等地参加学术会议。学术带头人张绍麒教授主持的山东省社科规划重点项目“胶东方言电子语音语料库的建设及应用”很有特色,受到同行的关注。王东海副教授(在读博士)的论文《汉语同义语素的组配规则与参数》(《中国语文》2002年1期)获山东省社会科学优秀成果二等奖。徐德宽博士的“现代汉语双宾构造汉英机器翻译模型研究”以及国家汉办项目《汉语实词搭配词典》之子课题“《汉语实词搭配词典》用料库建设及研究”引起了同行的关注;还承担了《中华纸业》杂志社委托的“造纸类学术期刊论文摘要的汉英机器翻译模型研究”等横向课题。多年来供职于新加坡的姬东鸿博士也签约加盟,成为本方向的学术带头人之一,其丰硕的研究成果及影响将壮大本方向的力量,推动本方向的快速发展。

(三) 语言教学研究。本方向主要特色在于抓住新时期高等师范中文专业语言教学、对外汉语教学等所面临的挑战,积极进行语言专业课程体系、教学内容的改革探索。学术带头人张绍麒教授,全国优秀教师、山东省十佳优秀教师、山东省中青年学术骨干培养对象,近年来主要致力于语言教学研究。所主持的教学改革项目“现代汉语课规范化建设”获山东省优秀教学成果一等奖、国家普通高校优秀教学成果

二等奖；主持建立了适应培养厚基础、宽口径、高素质人才的“四·四”辐射型汉语教学体系，主编教学成果论文集《面向21世纪中文专业教学内容和课程体系改革探索》，获山东省优秀教学成果二等奖。主持山东省教学改革试点专业——汉语言专业的改革和建设取得了丰硕的成果。其论文《现代汉语考试系统》《汉语词语文化标记及其在跨文化交际中的策略》在汉语教学改革及对外汉语教学中有一定的影响。李敏教授在面向对外汉语教学的现代汉语语法研究方面进行了多年的研究，在《语言教学与研究》《世界汉语教学》等期刊上发表了多篇学术论文、出版了《实用语言文字教程》等著作，获市社科优秀成果奖3项，校优秀教学成果奖2项。徐德宽博士承担了国家汉办“十五”科研规划项目“汉语方位词语的认知语义基础及意象隐喻系统”之子课题“汉语方位词语的隐喻研究”。

（四）汉语与汉字规范化研究。本方向的特色在于面向实际应用，深入进行汉语汉字规范化研究。学术带头人陈淑梅教授（博士）有《汉字应用通则》（合作）、《汉字》和《汉字学概论》等著作，在《语文建设》《古汉语研究》等学术期刊发表论文30余篇，博士论文《东汉碑隶构形系统研究》即将由上海教育出版社出版。其中《汉字应用通则》（合作）获省社会科学优秀研究成果一等奖。主持教育部语言文字应用“十五”规划项目“语言规划研究”1项，省古籍整理项目1项。吕永进教授主持完成了山东省教育厅科研项目“汉字异体学”和“现代汉字偏旁部首变异系统研究”，在《语言文字应用》《语文建设》等学术期刊上发表《现代汉字左声右形结构析得》等汉字规范化论文20多篇，还参加了教育部“十五”规划重大项目“《规范汉字表》的研制”。徐德宽博士在《中国学术期刊》等刊物上发表《学术期刊中汉族作者四字姓名的汉语

拼音拼写问题》等论文。王宝刚博士、宋洪民博士参与了国家语标委语言学名词规范化的工作。

《语言文字理论与应用研究文库》是重点学科强化建设成就的一个阶段性总结,是烟台师范学院语言文字学科群部分成员近年来在各个方向开展学术研究的一个汇报,随着学科群成员教学、科研的不断深入和成熟,《文库》还要继续出版他们的研究成果。尽管其中不成熟之作在所难免,但我们愿意把它们奉献出来,接受学界的检验,并就教于方家。

如果说鲁东大学语言文字学科群的学科建设取得了一点成绩的话,完全得益于我们有幸处于国家注重进行高等教育改革、重视发展学科建设的大环境之中。我们的学科建设自始至终得到了山东省教育厅、烟台师范院校党委的正确领导和全力支持,使学科群在实验室设备和图书资料的购置、人才引进和培养等方面都能够走在前列。

根据国家在新形势下对高等学校发展的要求,学科群将在进一步加强汉语言文字学基础理论研究的同时,重点发展直接服务于我国现代化建设的应用语言学研究。具体目标:(1)加大中文信息处理实验室的建设力度,争取成为省级重点实验室,进一步增强其在学科群中的龙头作用。(2)语言学与应用语言学成为硕士研究生专业。(3)汉语言文字学成为博士研究生专业。

总之,该学科群具有稳定的、理论与应用密切结合的研究方向,有一支专业结构、学位结构、职称结构和年龄结构均较合理的教学、科研队伍,显示了强劲的发展势头。现在,学科群全体成员正齐心协力、奋发向上,为实现学科群发展的既定目标而努力奋斗。

绪论 改变观念,抓住机遇,大力发展中文信息处理研究

近年来,在校院两级领导的大力支持下,我院中文信息处理取得比较大的进步。我们将进一步改变观念,抓住机遇,大力发展中文信息处理事业。

1 紧紧抓住中文信息处理发展的大好机遇,努力扩大影响

中文信息处理就是利用计算机处理汉语汉字信息。中文信息处理包括基础研究和应用研究两个方面。基础研究包括汉字信息处理研究、词处理研究、语句和篇章处理研究、中文信息处理的语言学理论研究;应用研究包括自然语言理解、机器翻译、情报检索、计算机辅助教学和研究等。而就目前中文信息处理的发展来看,语言学问题已经成为汉语信息处理的“瓶颈”,积极开展面向中文信息处理的汉语研究是当务之急。同时运用计算机进行辅助研究和教学又有非常广阔的市场。这两个方面,无论从哪一方面来讲,中文

专业都有得天独厚的条件。积极开展面向中文信息处理的汉语汉字研究、利用计算机进行辅助教学、辅助研究也是中文专业的义不容辞的责任和义务。21世纪是信息产业的世纪。信息产业的基础是信息技术,中文信息处理是信息技术领域的一个重要方面。从20世纪80年代起,中文信息处理引起了国内外信息处理界的广泛关注,取得了可喜的进步。国家为了解决中文信息处理所遇到的困难,促使中文信息处理的迅速发展,特别重视中文信息处理人才的培养。1998年颁布的高等师范院校中文专业教学大纲中,将“中文信息处理”和“计算语言学”确定为中文专业的基础课。这给中文信息处理的发展带来了新的机遇。有鉴于此,我们确定了以中文信息处理为龙头的教学改革思路。从中文系的专业设置到课程改革建设,都体现出浓厚的中文信息处理的特色。中文专业主要分为两块:一块是语言,一块是文学。语言这一块重点发展中文信息处理的基础理论研究和应用研究,特别要开展面向中文信息处理的汉语的教学和研究,同时加强各课程的改造和新增课程的建设。文学这一块主要是加强计算机多媒体辅助教学,以提高教学质量。两方面共同的任务是加强各种辅助教学软件的建设,包括题库、练习库、讲义、多媒体课件、教学研究资料信息库等,开发出各种学习软件包。一方面满足我们自己教学的需要,一方面投放市场,以期获得更好的社会效益和经济效益。争取把中文系办成一个产、学、研相结合的基地。为了具体落实这些规划,并且随着中文系发展的实际需要,在原有的中文信息处理实验室和汉语言文字学重点学科的基础上,申请并获准成立了中文信息处理重点实验室。重点实验室下分了“语言信息处理”“计

“计算机辅助教学”“计算机辅助研究”等三个方向,即将成立中文信息处理研究所。经过短短几年的努力,我们的中文信息处理得到了快速的发展,取得了很大的进步,发表论文 40 余篇,出版著作 8 部,承担各类科研项目 20 余项,2001 年在完成两个国家 973 项目子课题的基础上,又获得了国家哲学社会科学规划项目“《现代汉语新词语信息电子词典》的开发与应用”(项目编号: 01CYY002),而且得到“东方与中文语言信息处理学会”的资助应邀参加了在新加坡举行的“2001 中文电脑国际研讨会”,所提交的论文《〈汉字义类信息库〉的研究与实现》得到广泛的好评,被大会评为“最佳论文”,发表于《汉语语言与计算学报》2002 年第二期,2004 年获得山东省社会科学优秀研究成果奖。2002 年得到资助应邀赴台北参加台湾中央研究院主办的“第三届中文词汇语义学学术会议”,发表论文《基于数据库的汉语字、词语义分布比较研究》《〈现代汉语新词语信息电子词典〉的开发与应用》,后者又发表于《中文计算语言学期刊》2002 年 2 期。2004 年应邀得到资助赴新加坡、马来西亚参加“第五届中文词汇语义学学术会议”,发表论文《基于数据库的汉语语义构词法初探》。我们所做的工作和所取得的成绩在中文信息处理界产生了一定的影响。

2 “借炉炼钢”,苦练内功——中文信息处理发展的途径

我们烟台师范学院是小学校,许多方面都无法和一些著名大学、研究机构相比,尤其是中文信息处理研究刚刚起步,独立申请并承担高水平、有影响的科研项目,可能性几乎是

零。在这样一种情况下,我们不等不靠,积极与国内外著名高校和科研机构联系,走横向联合发展之路。1999年我们和北京大学计算语言学研究所等单位合作,完成了国际合作项目“大规模的现代汉语标注语料库”,和北京大学计算语言学研究所合作,完成了国家社科基金项目“古诗计算机辅助研究系统及其应用”,和德国 KONSTANZ 工业技术大学合作完成了“全信息标注语料库的开发与应用”,和香港城市大学合作了“现代汉语句法语义动词研究”等项目。此后,我们逐渐开始独立承担任务,先后承担了上海辞书出版社立项资助的“现代汉语新词语研究”,清华大学智能技术与系统国家重点实验室主持的国家 973 项目“面向大规模真实文本的汉语计算的理论、方法、工具”的子课题“《汉字义类信息库》的研究与实现”,北京大学计算语言学研究所主持的国家 973 项目“面向新闻领域的汉英机器翻译系统”的子课题“汉英对齐语料库对比研究”。通过合作或独立承担完成这些项目,开阔了我们的视野,提高了我们的能力,壮大了我们的队伍,积累了一定的资源,扩大了我们的影响。有了这样的基础,我们趁热打铁,在继续保持横向联合的同时,积极主动地走独立发展的道路,我们先后申请并获得批准了山东省教育厅科研项目“基于语料库的现代汉语新词语研究”、山东省社科规划项目“胶辽方言语音语料库的研究与实现”、国家社科规划项目“《现代汉语新词语信息电子词典》的开发与应用”,并相应地发表、出版了一批比较有影响的成果,进一步明确了我们的发展目标及发展方向。目前正在承担的项目有山东省社科规划项目“面向信息处理的汉语语义构词规则研究”、国家 973 项目子课题“基于标注语料库的现代汉语句子语义结