基于网络媒体监测语料库的性别语言差异研究

王宇波 著

科学出版社

本书由教育部人文社会科学研究青年基金项目“基于大规模标注博客语料库的性别话语差异实证研究”（12YJC740106）及中国语情与社会发展研究中心资助出版

基于网络媒体监测语料库的性别语言差异研究

王宇波 著

科学出版社

北京

内 容 简 介

本书的研究依托国家语言资源监测与研究中心（网络媒体语言分中心）所建的网络媒体监测语料库，研究对象主要包括字、词、句式、话题优先序列、话语量大小等方面的性别语言差异。作为首次基于目前规模最大的汉语网络媒体监测语料库的汉语本体研究，主要采用语料库语言学的研究手段，在定量和定性相结合的基础上，以统计数据和语言事实为研究依据，通过处理大规模的真实文本得出了具有统计学依据的语言数据，更深入地揭示两性在汉语使用中的一些普遍规律，挖掘自媒体中性别语言呈现的一些特有的语言现象和规律，既符合当前性别语言差异研究的多元化、动态化、微观化和本土化的发展趋势，也对应用语言学、社会语言学等方面的研究有一定的理论价值。

本书的读者对象为高校语言学专业教师、研究生及语言学领域爱好者。

图书在版编目(CIP)数据

基于网络媒体监测语料库的性别语言差异研究/王宇波著.
—北京：科学出版社，2017.10
ISBN 978-7-03-054594-7

Ⅰ. ①基… Ⅱ. ①王… Ⅲ. ①汉语–性别差异–社会语言学–研究 Ⅳ. ①H1

中国版本图书馆 CIP 数据核字(2017)第 236968 号

责任编辑：张 达 / 责任校对：郑金红
责任印制：张欣秀 / 封面设计：正典设计

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

北京京华虎彩印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2017 年 10 月第 一 版 开本：720×1000 B5
2017 年 10 月第一次印刷 印张：12 3/4
字数：210 000

定价：72.00 元

（如有印装质量问题，我社负责调换）

当一个语言学研究的排头兵

我于2004年开始独立指导博士研究生，第一位是香港地区的陈潮忠，他在2007年答辩的学位论文是《香港回归前后普通话使用情况及发展前景考察》。我从2007年开始每年按计划招生，盘算下来，到今年夏季，我指导毕业的博士生共有23名（含越南学生5名、韩国学生2名），只有7名男生（含韩国1名、越南1名），其中就有2008年录取的王宇波，他是当年华中师范大学文学院语言学各专业初试、复试的第一名，也是我很得意和至今看重的学生。

王宇波出身于教师之家，这使得他具有良好的教养和品德。在华中师范大学攻读硕士和博士学位期间，力求上进，勤奋刻苦，严谨踏实，多次获得“优秀研究生干部”“优秀研究生”等称号；思维敏捷，肯下工夫，坐得下来，沉得下去，专业课成绩一直名列前茅，硕士学位论文被评为优秀毕业论文。2009年初，他申请教育部研究生培养创新计划，被选拔进入北京大学中文系联合培养至2010年初。在沈阳教授的指导下，共选修或旁听了23位老师的26门研究生课程。这对于他的专业知识和科研能力无疑是极大的充实和有力的提升。在攻读硕士和博士学位期间，他在核心期刊上发表学术论文多篇，由《中国社会科学文摘》转载摘要。

王宇波还有一个优点就是具有很强的组织沟通能力和团队合作精神，不计得失，任劳任怨。他担任文学院语言学博士研究生专业负责人和班长，多次参与重大国际学术会议的会务工作，协助《汉语学报》编辑部处理日常编务工作。2009年我主持国家社科基金一般项目(编号09BYY018)，他作为合作者和重要助手，勇挑重担，承担了大量基础工作和研究工作，还有效地协调其他几位博士生参与工作，保证了课题按期提交成果，以鉴定等级“良好”结项。王宇波也在项目研究过程中，撰写了博士学位论文《基

于网络媒体监测语料库（汉语）的性别语言差异实证研究》，于2011年夏季通过答辩，获得博士学位。当年我给王宇波的学位论文的“导师评语”是这样写的：

王宇波的学位论文《基于网络媒体监测语料库（汉语）的性别语言差异实证研究》以该媒体监测语料库的各项统计数据为研究基础，以网络语言中性别差异为突破口，综合运用多种研究手段和理论方法，选题具有明显的理论价值和应用价值。论文的突出优点有二。

一、依托大规模真实文本进行汉语本体的数据研究，首开先河，言之成理，持之有据。

论文的研究依托国家语言资源监测与研究中心（网络媒体语言分中心）的语料，是首次基于目前规模最大的汉语网络媒体监测语料库的汉语本体研究。论文主要采用计算机和语料库的一些现代化手段对语言数据进行统计，运用的理论、方法结合了汉语的实际情况，也反映了学术研究的最新动向。同时在定量和定性相结合的基础上，尽量增加了定量研究的成分，试图以数据和事实为研究依据，通过处理大规模的真实文本，得出许多非常珍贵的数据，这些数据既有统计学的依据，又有语言学尤其是应用语言学的意义，而且对于如何运用网络媒体监测语料库统计的数据资源深入开展汉语本体研究具有方法论上的意义，同时拓展了社会语言学从功能角度探索语言使用的视野，其贡献对语言学和社会语言学本身都是具体切实的。

二、注重探求语法事实，讲究运用理论方法，深入比较性别差异，成果具有多重价值。

论文是国内首次以大规模汉语语料为基础的性别语言研究，研究内容具有原创性，研究方法具有跨学科性和实证性。在解释分析上，既有宏观理论性探讨，又有微观的词例、句式的考释；既有文字叙述，又辅以图表。研究成果的价值在于：第一，将有助于更深入地揭示两性在汉语言文字使用上的一些普遍规律，挖掘汉语性别语言自本世纪初以来呈现的一些特有的语言现象和特点；第二，既可以为专家学者研究语言政策、语言规划提供参考，也可以为语言文字工作部门和其他有关部门制定和调整相关的政策提供学术上的支持；第三，也可对汉语教学以及中文信息处理提供一定

的理论基础，从第二语言教学的意义上讲，重视性别语言差异的研究有助于学习者克服跨文化交际的障碍，从而促进汉语语言教学的发展；第四，还将有利于国家制定标准和进行语言资源监测，对中文信息处理有着重要的意义。

总之，这是一篇有理论探求、有方法寻求、有价值追求的博士学位论文。

在博士学位论文的基础上，王宇波结合主持教育部人文社会科学基金青年项目，进一步充实提炼，通过处理大规模的真实文本得出具有统计学依据的语言数据，在揭示男女性在汉语使用中的普遍规律的同时，更深入地挖掘新媒体中汉语性别语言呈现的一些特有的语言现象和规律，既符合当前语言性别差异研究的多元化、动态化、微观化和本土化的发展趋势，也对应用语言学、社会语言学等方面的研究具有一定的理论价值，最终形成了这部专著——《基于网络媒体监测语料库的性别语言差异研究》。

我给王宇波这部专著的序言取了个标题：当一个语言学研究的排头兵。《现代汉语词典》“排头兵”的释义：“站在队伍最前面的兵，比喻带头的人或集体。”（第 6 版 967 页）我认为，在“80 后”这一拨儿研究语言学的青年学者中，对于 2011 年博士研究生毕业进入“985”高校——武汉大学的王宇波来说，意味着阵地更前沿，视野更宽广，步伐更开阔，优势更明显，更容易成为队伍的排头兵。

2011 年王宇波进入武汉大学文学院语言学及应用语言学教研室任教，同时进入武汉大学中国语言文学博士后流动站语言学及应用语言学专业从事博士后研究工作，研究兴趣为自媒体语言监测及相关研究，2013 年博士后出站，入选武汉大学“351 人才计划”，被聘为武汉大学“珞珈青年学者”。2014 年进入武汉大学信息管理学院图书馆、情报与档案管理博士后流动站情报学专业从事二站博士后研究，研究兴趣为中国语言国情信息监测模型及框架构建研究，获国家留学基金委“2014 年高等学校青年骨干教师出国研修项目”资助，赴新加坡南洋理工大学从事自然语言信息处理及相关研究。

王宇波现在主要从事语言应用与信息处理、计量语言学、汉语国际教育等方面的教学与研究工作。他先后主持国家社科基金青年项目（12CYY030）、教育部人文社会科学基金青年项目（12YJC740106）、国家语

委“十三五”科研规划重点项目（ZDI135-11）、中国博士后科学基金第 6 批特别资助（2013T60734）、中国博士后科学基金第 9 批特别资助（2016T90717）、中国博士后科学基金第 51 批面上资助（2012M511647）、中国博士后科学基金第 58 批面上资助（2015M582259）、国家语委委托项目子项目（WT123-43）、武汉大学人文社会科学自主科研项目（2012YB013）等十项；参与国家自然科学基金（71774121）、国家社科基金（09BYY018）和教育部人文社科基金项目（11YJC740084）、（14YJC740079）多项。在《汉语学报》《语文研究》《华中师范大学学报（哲学社会科学版）》等刊物发表论文三十余篇，其中，《新媒体中两性话题选择的优先序列研究》被中国人民大学书报资料中心复印报刊资料《语言文字学》全文转载。曾获中国语文现代化学会第一届国际会议暨第九次学术会议、语言研究与信息处理国际研讨会会议论文一等奖。他担任国家语言文字工作委员会中国语情与社会发展研究中心事业发展办公室副主任，多次参与《中国语言生活绿皮书：中国语言生活状况报告》的编写工作。作为《中国语情》执行主编和编委，为国家语委、教育部语言文字信息管理司和语言文字应用管理司等部门提供决策参考和咨询服务，撰写二十余篇调研报告，参与撰写的《构建国家语言智库体系，服务国家战略和社会需求——关于构建国家语言智库体系的建议》作为“中国语情特稿”呈递给相关政府部门，获教育部领导批示。

窃以为，按照民间二十年一代人的说法，就中国现代语言学研究来说，从事研究工作始于二十世纪四十年代的学者是第一代，六十年代的学者是第二代，我们这些八十年代入门的算是第三代，像王宇波他们这些伴随新世纪走上语言学研究之路的后生是第四代，这也是充满活力、富有潜力、极具实力的新一代。唐代诗人李商隐诗云：“桐花万里丹山路，雏凤清于老凤声。”衷心希望中国语言学事业的新生代站在排头，勇立潮头，奋力前行，再铸辉煌。

是为序。

李向农

二〇一七年夏末秋初于华大家园

序　言

性别语言研究是语言学、心理学、社会学、人类学、交际分析等许多领域所关注的热点课题。20 世纪 60 年代，随着社会语言学的发展壮大和女权运动的兴起，语言与性别的关系问题作为一个独立的语言学问题备受关注。以前一直呈零散状态的性别语言研究开始进入快速发展时期，性别语言研究逐渐成为社会语言学有别于其他语言学研究的一个显著性标志，吸引了众多语言学研究者从不同的角度，采用不同的调查方法和理论框架进行研究。70 年代以后，社会语言学家逐渐认识到，说话人的性别作为一种言语标记，应当与其年龄、地位、角色、身份这些内容具有同等地位，因而应当成为社会语言学研究中的一个独立研究变量。而后，越来越多的研究者以社会构建的眼光来看待性别这个范畴，认为语言是构建性别的重要方式。到今天，语言与性别的关系已经成为一种必不可少的研究变量存在于社会语言学的许多领域之中，研究的内容也远远超出传统语言学所研究的范围，它已经同话语分析、篇章结构、交际策略，以及认知结构等方面的内容结合起来，形成了跨文化、跨学科研究中的新的学术生长点。

国外的性别语言研究集中在语言性别歧视现象和语言性别差异上。语言性别差异方面的研究又是最主要的方面，主要表现在：倾向于探讨两性语言的某一方面的差异，性别语言的跨学科研究，以及对性别语言与其他变量关系的研究。国内的性别语言研究开始较早，20 世纪 20 年代，赵元任先生就通过对吴语的调查来探讨男女在语音和声调上的差异。80 年代之后，越来越多的学者开始关注性别语言问题，相关研究起初零星散见于一些社会语言学著作的相关章节中，而后学界开始注重性别语言差异的理论研究，开始涉及性别语言与文化的关系问题、两性的语音差异及其形成原因、汉字中的性别语言特征等方面。经过几十年的发展，我国的性别语言研究取得了较为丰硕的成果。从研究内容看，可大致分为四个方面：第一，理论

介绍和综述，如杨永林（1991）、赵蓉晖（1999）、肖建安（2000）、李经伟（1998、2001、2002）等。第二，反映在语言上的性别歧视。杜文礼（1993）分别从词汇运用、句法层次、语序、英语的固定表达几个方面讨论了语言性别歧视的表现形式和解决办法；陈建民（1999）对汉语词语意义的不对称模式及词序上的男性为先进行了探讨；钱进（2003）从语义性别原型的角度分析了汉语成语、俗语中的性别观。其他文章也从不同角度描述和分析了语言中的性别歧视，如秦秀白（1996）、潘建（2001）、董晓波（2004）等。第三，性别语言的表现形式。主要从音系学特征、词汇特征、句法特征、语用特征等方面研究性别语言的表现形式。相关研究认为，女性比男性更倾向于使用标准的、权威的发音方式，更常用逆序重音；语调上，女性常用升调来表示礼貌或试探性的口气，而男性语调平淡、严肃；词汇方面，女性比男性更喜欢使用情态词语、特殊的色彩词语、强势语、委婉语等，而男性在言语交谈中常会夹带一些禁忌语、粗话、詈语等；句法上，女性比男性更频繁地使用反义疑问句、强调结构等。如胡明扬（1991）、郭熙（1999）、沈炯（1987）、杨永林（1989、1993）、肖建安（2000）、白解红（2000）等。第四，言语交际上的性别差异。随着研究的深入，研究重点逐渐从两性用语的形式结构差异转到交际策略、话语风格等方面，言语交际上的性别差异研究成为学界的研究热点。曹志耘（1987）、杨永忠（2002）、冯江鸿（2003）、潘小燕（2004）等探讨了两性在日常交往中的话题、话语量、话语方式与交际策略等方面的差异。

虽然国内外相关研究取得了丰硕的成果，但是仍有进一步深入研究和拓展的空间，主要表现在：第一，在研究内容上，主要关注两性在语言使用方面的差异，特别是语言中的性别歧视问题，且过分强调两者存在的差异，忽视了两性在语言和言语两个层面上的共性，从而把两性在语言使用方面的差异简单化和片面化。第二，在研究方法上，国内的语言性别研究还局限在借鉴西方的性别差异研究理论上。研究的层面还停留在介绍西方理论、研究外语为主，兼顾汉外语言性别差异的对比。对汉语词语自身性别差异的规律尚缺乏全面的、深入的、客观的描写和实证研究。第三，在理论解释上，存在解释简单化倾向。造成性别语言差异的原因复杂而多样，至今尚未形成能被普遍接受的理论解释。大部分研究在解释差异产生的原

因时，过度强调引起语言使用中性别差异的某一个因素而忽视其他因素，如支配论和差异论等理论所依据的是范围有限的实证性研究。因此，相关片面化的解释广遭批评（Bing & Bergvall，1996；Henley & Kramarae，1991；Uchida，1992）。第四，从考察范围来看，忽视多重因素，考察单一化。性别语言的某些差异是多种因素共同作用的结果，性别并不是唯一显著的变量。不少研究还只是静态地考察语言与性别之间的关系，而不是综合、全面地观察性别与其他不同因素对语言使用的影响。

迄今，国内语言性别差异研究的专著并不多见，对汉语中的性别语言差异尚缺乏全面而深入的研究。现有的大部分研究所引的观点和语料都来自西方学者的成果，这些研究成果多是在西方国家的社会文化背景下完成的，反映的多是西方主流社会的语言现象，由此得出的结论不一定具有普遍性，未必合乎汉语的语言事实，无法反映我国现阶段的语言生活状况。对此，国内汉语界和外语界的学者们已有共识。

本书的研究依托国家语言资源监测与研究中心（网络媒体语言分中心）所建的网络媒体监测语料库（汉语），研究对象主要包括字、词、句式、话题优先序列、话语量大小等方面的性别语言差异，相关数据涉及“频次、频率、覆盖率、词性分布、词类分布、独用词、共用词、频差、文本数、文本长度”等。作为首次基于目前规模最大的汉语网络媒体监测语料库的汉语本体研究，主要采用语料库语言学的研究手段，在定量和定性相结合的基础上，尽量增加了计量研究的成分，以统计数据和语言事实为研究依据，通过处理大规模的真实文本得出具有统计学依据的语言数据。不仅对于如何运用网络媒体监测语料库统计的数据资源深入开展汉语本体研究具有方法论上的意义，而且有助于更深入地揭示男女性在汉语使用中的一些普遍规律，挖掘新媒体中汉语性别语言呈现的一些特有的语言现象和规律，既符合当前语言性别差异研究的多元化、动态化、微观化和本土化的发展趋势，也对应用语言学、社会语言学等方面的研究有一定的理论价值。

王宇波

2017 年 8 月

目　录

第三篇 话 语 篇

第一篇　字　词　篇