

中国满通古斯语言语料数据库建设及研究^{*}

□赵阿平

【摘要】满通古斯语族语言文化研究是对文化传承具有重要价值且急需抢救的濒危学科,亦是对中国特色哲学社会科学整体构建具有重要作用而有待加强的特色学科。随着高科技迅猛发展与经济全球化,民族传统文化多样化遇到了空前的生存危机。目前,满通古斯语族语言由于自然生态环境与社会环境的变化都已处于濒危状态,尤其是满语、赫哲语更为严重,已处于极度濒危走向消亡的状态。浩瀚的满文档案文献史料极为珍贵,亦亟待开发利用。满通古斯语族语言文化遗产抢救与语料数据库建设,以及全面科学系统的综合研究显得更加紧迫。

【关键词】满通古斯语言语料数据库;满通古斯语族;濒危语言

【中图分类号】H0-09 【文献标识码】A 【文章编号】1673-8179(2019)05-0164-06

濒危语言资料的调查同一般的语言调查相比有其独特性,这是由濒危语言的特点所决定的。所以,对濒危语言资料的记录会同以往的语料收集不尽相同,其记录的目的是为了全面展示一种语言的真实面貌,长久地保存自然的原始语料,为后代留存一份文化遗产的活化石。^[1]中国境内的满通古斯语族语言目前都已处于濒危状态,尤其是满语、赫哲语已属极度濒危语言。为保护濒危语言,为保护人类非物质文化遗产提供可靠的语料数据,满通古斯语言语料库建设极为必要且已具备条件。此项目研究将使珍贵的民族语言文化遗产得以有效保存,为持续深入研究满通古斯语言人类学及相关学科奠定基础。

(一) 价值意义

1. 学术价值

满通古斯语言口语是研究民族语言文化的第一手珍贵资料,也是满通古斯语言人类学及相关学科研究的活化石,具有极其重要的学术价值。满通古斯语言文化田野调查研究经历了长期持久的进程,成就斐然。^[2]以往国内外对满通古斯语言研究,虽然取得诸多成就,获得重大发展及突破性成果,但仍存在一些问题。主要是未能整合国内外专业人才梯队协同攻关,致使研究成果缺乏系统性与全面性,最为缺憾的是未能建设满通古斯语族语言语料库,并进行相关研究。本课题首次开展中国满通古斯语言语料全面系统调查研究,完成濒危满语、锡伯语、赫哲语、鄂伦春语、鄂温克语语料数字化处理,长期有效保存珍贵语言文化遗产,为深入研究满通古斯语言奠定基础。在课题研究过程中,培养一批青年专业人才,并通过本课题研究,可以凝聚海内外满通古斯语言文化及相关领域的研究力量,有力推进中国满通古斯语言学科建设发展与协同创新,从而确立中国在此领域的领先优势地位。

2. 应用价值

语料库建设作为现代语言学研究领域的一个重要问题,是通过利用计算机信息技术深化语言应用与研究的新方向,近年来日益为国内外学术界所重视。语料库是经过科学收集和整理大量在实际情况下使用的语言信息而建立的专供研究使用的资料库。语料库作为一个或者多个应用目标而专门收集的,是有一定结构的、有代表的、可被计算机程序检索的、具有一定规模的语料集合。^[3]本质上讲,语料库实际上是通过自然语言运用的随机抽样,以一定大小的语言样本来代表某一研究中所确定的语言运用的总体。语料库建设

* 收稿日期 2019-06-21

基金项目:2018年度国家社会科学基金重大项目“中国满通古斯语言语料数据库建设及研究”(项目编号:18ZD300)。

©1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

进行自然语言文本的采集、存储、检索、统计、语法标注、句法语义分析,以及具有上述功能的语料库可在语言教学、语言定量分析、词汇研究、词语搭配研究、词典编制、语法研究、语言文化研究、作品风格分析、自然语言理解、机器翻译等进行研究。

目前在国际范围已建立了诸多系统化的语料库,也相继建立了濒危语言语料库。这对于保存濒危的满语、赫哲语、鄂伦春语、锡伯语、鄂温克语与大量的满文档案文献翻译工作都具有重要的学术价值与开发应用意义。建立满通古斯语言语料库,长期保存珍贵的民族语言文化遗产,为保护濒危语言、保护人类非物质文化遗产提供可靠的语料数据,为持续深入研究满通古斯语言人类学及相关学科奠定基础。在满通古斯语言语料库建设中,进行词汇结构分析、语法分析及词汇语义分析,可以为编纂大型满通古斯语言辞书、多语对照词典及分类工具书奠定基础,提供参考。在此基础上,使满汉双语平行语料库建设提升层次,对于满文档案的初步翻译问题可以加速解决,扭转中国现存大量满文档案相对翻译整理人才匮乏的紧迫局面,促进满文文献整理研究,使之开发利用获得实效。本课题最终成果是反映该领域当代研究水平,经得起历史检验的具有高水平学术价值、理论与应用创新意义的标志性成果。

3. 社会意义

语言的演化与生物的演化一样,都经历了漫长的过程,在以演化生物学和生物统计的技术研究语言的演化和形态分类,具有相通的可行性,大型的统计计算还需要很长时间以及充足的语料库。在此期间,不同的语言和生物一样,逐渐出现演化的差别、“技能树”的分异、地缘的隔离与差异,而同时又出现了交融,最后形成了目前这种复杂的树状—网状结构。这些语言的演化、新语言的诞生、语言的发展的过程,是自有人类语言以来就已经出现的。正如生物多样性一样,语言多样性的最核心的意义在于“基因库”的保存,人类的每一个认知的可能性,都是这个库里的一个“基因”,只有合起来,才能构成完整的人类能力,而语言多样性,就是反映这一可能的有力证据。^[4]

人类通过语言与思维创造了文化,人类的语言结构反映了一个族群(民族)的性格和文化特点。语言承载着文化,传承创造了人类文明丰富多彩的结晶,语言与人类文化的关系是相互依存的。因此,保护濒危语言的意义,最根本在于其承载着一种文化。人类语言学是通过语言来反映不同文化的共性与特性,特定的自然环境、社会环境与一个民族的存在和发展有着密切的关系。当今世界科学技术迅猛发展,对于全球一体化的进程产生强烈影响。在这样多元一体化历史发展的总趋势面前,一个民族或一个国家的特色文化显得更加珍贵,更有价值。尤其是至今仍保存着人类早期社会诸多古老文明和特殊形态结构的民族,更显示出其特有的价值与作用。在当今社会许多科学命题难以解释,通过对某些民族发展足迹的探寻可以得到启示与答案。满通古斯语族语言文化正是呈现如上诸多特征,因而作为典型案例更具有社会研究的重要价值和意义。^[5]满通古斯语言研究对于保护濒危语言、探讨文化多样性、克服文化冲击、提高跨语言交际能力以及推广官方语言、加强双语教育等都具有实际的应用价值。特别是对研究中华民族多元一体化进程、快速推进中国现代社会经济文化发展、协调民族关系以及外交政策、促进人类和谐发展都具有重要的参考价值与现实意义。

(二) 主要内容

满通古斯语族语言文化研究是对文化传承具有重要价值且急需抢救的濒危学科,亦是对中国特色哲学社会科学整体构建具有重要作用而有待加强的特色学科。随着高科技迅猛发展与经济全球化,民族传统文化多样化遇到了空前的生存危机。目前,满通古斯语族语言由于自然生态环境与社会环境的变化都已处于濒危状态,尤其是满语、赫哲语更为严重,已处于极度濒危走向消亡的状态,浩瀚的满文档案文献史料极为珍贵,亦亟待开发利用。满通古斯语族语言文化遗产抢救与语料数据库建设,以及全面科学系统的综合研究显得更加紧迫。

1. 满通古斯语言语料调查采录研究

满通古斯语言语料库建设及研究,主要是运用结构语言学、人类文化语言学、社会语言学、计算机语言学

等多学科理论方法,以国家权威语料库建设标准,定位课题研究工作,对濒危满语、锡伯语、赫哲语、鄂伦春语、鄂温克语的基础语料进行田野调查采集,并对以往调查采集的资料进行整体梳理录制,分析研究。

课题突出“语言特色、地域特点、技术特色”,即突出满语、锡伯语、赫哲语、鄂伦春语、鄂温克语的语料调查研究及对比研究,填补前人研究空白。课题调研地域涉及满语、锡伯语、赫哲语、鄂伦春语、鄂温克语现存地黑龙江省富裕县三家子村、黑河市四季屯、大五家子、新疆察布查尔锡伯族自治县、同江街津口、饶河、抚远、鄂伦春十八站、呼玛县、黑河新生鄂伦春乡、呼伦贝尔鄂温克旗等地。调查对象以熟练掌握满语、锡伯语、赫哲语、鄂伦春语、鄂温克语的七十岁以上老人为抢救调查重点;其次为五十岁以上语言会话较好者;再次为五十岁以下各年龄段语言会话者。

调查研究内容主要为满语、锡伯语、赫哲语、鄂伦春语、鄂温克语的音系、词汇、句子、语法、话语、口头文化、口述史等,按照国家语言保护标准采集音视频信息,对书面语进行拉丁转写注音,对口语进行国际音标注音,对话语开展语法分析。注重将理论研究与应用研究相结合,探讨研究满语、锡伯语、赫哲语、鄂伦春语、鄂温克语濒危形成、演变、发展的过程、规律及特点。

2. 总体研究与子课题构成

本课题总体研究分由五个子课题展开,子课题一,濒危满语、锡伯语语料调查收集及研究;子课题二,濒危鄂伦春语语料调查收集及研究;子课题三,濒危鄂温克语语料调查收集及研究;子课题四,濒危赫哲语语料调查收集及研究;子课题五,中国满通古斯语言语料数据库建设及研究。

本课题围绕中国满通古斯语言语料数据库整体建设及研究,在充分搜集、整理和吸收国内外相关语料、文献资料及已有资料成果的基础上,注意攻克课题各个环节的重点和难点。以上五个子课题是总课题的组成部分,相互密切联系组成有机整体。虽然各子课题的内容和侧重点有所区别,承担各自的义务与权力,但是必须服从总课题的统一要求。例如:子课题一、二、三、四,各语言以往调查保存资料有所不同可以根据具体情况适当整理后,先交由子课题五进行数字化处理收录。而各语言现状田野调查基本内容结构和体裁体例必须是一致的,田野调查的词汇2 000~3 000个、句子300句、民间传说、神话故事、家族史传承等语言片段数十段,制定统一调查大纲,分别进行调查,最后形成有机联系的整体课题。我们允许各子课题之间,对于特殊问题的处理可以有所不同,可以根据具体情况适当安排。课题组由国内外有关高校、科研机构相关领域知名专家学者和学术骨干组成,并联合美国、日本、韩国、中国香港等国家和地区相关高校、科研机构,协同实施课题研究任务。在研究方法上,拟采取结构语言学、社会语言学、人类文化语言学、历史比较语言学等多学科交叉研究方法,以解决研究中的关键点与盲点。特别注重运用历史比较语言学、人类文化语言学方法,考察满通古斯各族语言,尽量体现各语族不同的特点及相通的共性。

(三) 重点难点问题

1. 重点问题

首次开展中国满通古斯语言语料全面系统调查研究,完成濒危满语、锡伯语、赫哲语、鄂伦春语、鄂温克语语料的数字化处理,建设濒危满通古斯语言语料数据库,永久保存珍贵语言文化遗产,为满通古斯语言本体深入研究提供坚实的原始资料基础。撰写《濒危满语、锡伯语调查研究》《濒危鄂伦春语调查研究》《濒危鄂温克语调查研究》《濒危赫哲语调查研究》《中国满通古斯语言语料库建设及研究》5部专著,并进行词汇结构分析及语法分析,增添学术研究创新成果。同时对满语、锡伯语、赫哲语、鄂伦春语、鄂温克语濒危形成、演变、发展的过程、规律及特点进行分析探讨,为相关民族濒危语言提供参考。

2. 难点问题

满通古斯语言口语是研究民族语言文化的第一手珍贵资料,也是满通古斯语言人类学及相关学科研究的活化石,具有极其重要的学术价值。满通古斯语言文化田野调查研究经历了长期持久的进程,成就斐然。以往国内外对满通古斯语言研究,虽然取得丰硕成果及重大发展,但仍存在诸多问题。其中最主要的难点问题是未能整合国内外专业人才梯队协同攻关,致使研究成果缺乏系统性与全面性,最为缺憾的是未能建设满

通古斯语族语言语料库,并进行系统深入的综合研究。

中国境内的满通古斯语族分布地域广泛,主要分布在黑龙江省、新疆维吾尔自治区、内蒙古自治区等地区。在长期发展中,满通古斯语族与中原文化不断相互接触交往,相互影响乃至交融,在语言文化各方面既有各自的特性,又有诸多的相通性。满通古斯语族族群相互间关系多重,错综复杂,具有诸多的扩散性与神秘性,其历史渊源及发展复杂,语言文化习俗丰富且相互影响,该族群至今仍保存着人类早期文明的诸多形态和特征。因此,本课题对其进行全面系统调查研究,并完成濒危满通古斯语言语料数据库建设,具有艰巨的开创性与挑战性。

(四) 主要观点与研究方法

1. 问题选择

迄今为止国内外尚无对中国满通古斯语言语料全面系统调查研究,并同时完成建设濒危满通古斯语言语料数据库的整合调查研究成果,因此本课题选择的研究对象,即是问题选择方面的一个重大突破与创新。在永久保存珍贵的濒危满通古斯语言文化遗产的同时,撰写《濒危满语、锡伯语调查研究》《濒危鄂伦春语调查研究》《濒危鄂温克语调查研究》《濒危赫哲语调查研究》《中国满通古斯语言语料库建设及研究》5部专著,进行词汇结构分析、语法分析及语义分析,为学术研究增添创新成果。并对满语、锡伯语、赫哲语、鄂伦春语、鄂温克语濒危形成、演变、发展的过程、规律及特点进行分析探讨,为相关民族濒危语言提供参考。本课题组经过较长时间的准备、酝酿和论证,组合了精通掌握满通古斯诸语及英语、日语、俄语等语言并具有丰富成果与经验的专家学者,已经具备了解决本课题诸多难题的能力和条件。本课题的最终成果将是一套体例完备、内容丰富、观点正确的标志性成果,将集资料性、学术性、理论性、应用性于一体,反映中国当代学术水平,引领本学科领域向高度与深度发展,并为相关学科领域提供借鉴参考。

2. 学术观点

对于中国满通古斯语言语料全面系统调查研究,并同时完成建设濒危满通古斯语言语料数据库的整合调查研究成果,从而发现并掌握满通古斯语言长期动态变化的共性与特性,以及变化规律,具有一定的突破与进展。满通古斯诸族语言文化既有共通性,又有各自的特色。赫哲人以渔猎为主要生计方式,语言文化具有鲜明的渔猎习俗特色。鄂伦春人以狩猎为主要生计方式,语言文化带有狩猎习俗特色。驯鹿鄂温克人的主要生计方式亦决定了语言文化的特色。锡伯人经历了由东北向西北的迁徙,主要生计方式为畜牧与农耕,语言文化体现了相应特色。满族及其先人的生计方式则涉及了渔猎、狩猎、采集、畜牧、农耕多种,语言文化与之相应具有兼容并蓄的特色,并且在历史发展中多次崛起,曾经建立了地方政权的渤海国、半壁江山的大金国,乃至建立了一统天下历经 268 年的清王朝。其吸取了历代封建王朝的统治经验,继承了丰厚的遗产,并融入自身的民族文化特征。

3. 研究方法

严格遵循语言学的传统研究方法,同时采用多种新的跨学科研究方法和分析工具,增加研究成果的技术含量,使观点与结论更加接近科学。

(1) 多视角交叉研究

从濒危语言学、描写语言学、结构语言学、人类文化语言学、语料库语言学、实验语言学、计算机语言学等多视角切入,相互交叉展开研究。对濒危的满语、锡伯语、赫哲语、鄂伦春语、鄂温克语进行历时的系统调查,共时的异同比较,语言本体的结构分析,语言历史文化的深层透视,语料库数据化处理与客观分析的建设研究等等,相互交叉配合研究,获取立体丰富且客观科学的有效成果。

(2) 多学科综合研究

运用结构语言学、人类文化语言学、历史语言学、比较语言学等跨学科综合研究方法,以解决研究中的关键点与难点。特别注重运用比较语言学、人类文化语言学方法,考察濒危满语、锡伯语、赫哲语、鄂伦春语、鄂温克语,尽量体现其不同的特点及相通的共性。学科综合研究法,是将语言学研究方法与多学科研究方法相

结合 相互引进其他学科的指导方法。即立足语言学 ,与其他学科多向交流 ,吸取其他学科的研究成果和分析方法。科际交流 尤其表现在方法的借鉴与习得上。这个方法的特点是多种方法的相互渗透和融合 ,如 ,融会语言学上的共时描写法、历时分析法 ,社会学上的社会调查法、阶层分析法 ,以及其他学科的统计法、历史研究法等等。因此 ,本课题研究得以拥有更广泛的交流空间 ,实现突破性研究的预期结果。

(3) 多单位协同创新研究

组织大规模、跨学科、应用价值与社会价值较高的科研课题 ,方能将满通古斯语言文献与文学、社会学、历史学、民族学、人类学相结合 ,在注重基础理论研究的同时 ,加强应用领域的研究。本课题组织协调了相关高校、科研机构及专家学者 ,形成专业化、强强联合的优势团队。项目合作单位有中央民族大学少数民族语言文学系、香港大学饶宗颐学术馆、东北大学中国满学研究院、呼伦贝尔学院蒙古学学院、民族历史文化研究院等 ,汇集能力强、成果多、思路新的专业人才团队 ,形成经验交流、资源共享、优势互补、协同创新攻关合力 ,从而保障实现计划 ,获取突破性的成果。

(4) 使用现代化技术手段

人类社会高科技的飞速发展 ,进入了以电子技术为核心的数字化和网络化时代 ,这为科学研究提供了难以估量的便利条件。文献资料的查询与保存利用、田野调查现代化仪器设备的利用等各种现代化技术手段 ,都为我们当今调查研究提供了新的突破空间 ,超越了以往的不可能。本课题研究将传统的查询参考文献资料与当今的现代化技术手段相结合 ,踏实深入调查研究 ,获取真实可靠且科学有效的成果。

4. 文献资料

满通古斯语言是研究满通古斯语言文化的第一手珍贵资料 ,也是阿尔泰学及相关学科研究的活化石 ,对于学科建设发展具有极其重要的学术价值。本课题研究广泛精选了满通古斯语言综合研究文献^[6-9] (包括文献【6】至【9】) ,满语调查与研究文献资料^[10] (包括文献【10】) ,锡伯语、赫哲语、鄂伦春语、鄂温克语专门调查与研究文献资料^[11-14] (包括文献【11】至【14】) ,满通古斯语言相关文献研究文献资料^[15] (包括文献【15】) ,工具性文献资料^[16-17] (包括文献【16】和【17】) 等作为参考 ,突破了以往较为专门单纯性的文献资料选取参考 ,便于本课题对于满通古斯语言个体深入、整体系统全面的研究。语料库建设是现代语言学研究的一个重要领域和利用计算机信息技术深化语言应用与研究的新方向 ,对于保存濒危满语、锡伯语、赫哲语、鄂伦春语、鄂温克语与研究及翻译利用浩瀚的满文档案具有重要的科学价值与开发应用意义。

5. 话语体系等方面的探索

在课题研究过程中 ,培养一批青年专业人才 ,并通过本课题研究 ,凝聚海内外满通古斯语言文化及相关领域的研究力量 ,有力推进中国满通古斯语言学科建设发展与协同创新 ,从而确立中国在此领域的领先优势地位。☉

[责任编辑 秦红增] [专业编辑 康忠德] [责任校对 夏福立]

[参 考 文 献]

- [1]徐世旋.濒危语言研究[M].北京:中央民族大学出版社,2001.
- [2]赵阿平,郭孟秀,何学娟.濒危语言——满语、赫哲语共时研究[M].北京:社会科学文献出版社,2013.
- [3]范俊军.中国濒危语言有声语档数据规则[J].西北民族大学学报(哲学社会科学版),2016(3).
- [4]王士元.演化论与中国语言学[J].南开语言学刊,2008(2).
- [5]赵阿平.满通古斯语族语言文化研究新发展[N].中国社会科学报,2014(8).
- [6]朝克.满通古斯语族语言词源研究[M].北京:中国社会科学出版社,2014.
- [7]朝克.满通古斯语族语言词汇比较[M].北京:中国社会科学出版社,2014.
- [8]朝克.满通古斯语族语言研究史论[M].北京:中国社会科学出版社,2014.
- [9]赵阿平.满通古斯语言文化研究文库[M].北京:民族出版社,2004~2008.
- [10]山本谦吾.满语口语基础语汇集[M].东京:东京外国语大学亚非语言文化研究所,1969.

- [11]李树兰.锡伯语口语研究[M].北京:民族出版社,1984.
[12]胡增益.鄂伦春语研究[M].北京:民族出版社,2001.
[13]朝克.鄂温克语三大方言基本词汇对照集[M].北海道:日本小樽商科大学,1995.
[14]津曲敏郎.中国黑龙江省赫哲语调查[J].环北极文化比较研究,1992(1)..
[15]戴庆厦.中国濒危语言个案研究[M].北京:民族出版社,2004..
[16]冯志伟.中国语料库研究的历史与现状[J].汉语语言与计算学报,2002(1)..
[17]黄昌宁.语料库语言学[M].北京:商务印书馆,2002.

[作者简介] 赵阿平,满族,黑龙江大学满族语言文化研究中心教授、博士生导师。黑龙江哈尔滨,邮编:150080。周庆生,中国社会科学院民族学与人类学研究所研究员、中国语言学会语言政策与规划专业委员会副会长、中国少数民族双语教学研究会副会长。北京,邮编:100080。

Construction and Research of the Manchu-Tungusic Language Corpus Database in China

ZHAO A-Ping

(*Heilongjiang University, Harbin 150080, China*)

Abstract: The study of the language and culture of the Manch-Tungusic language group is an endangered principle which is of great value to the cultural inheritance and is in urgent need of rescue. It is also a unique principle which needs to be strengthened and plays an important role in the overall construction of the philosophy and social sciences with Chinese characteristics. With the rapid development of high technology and economic globalization, the diversity of traditional ethnic culture has encountered an unprecedented subsistence crisis. At present, due to the changes of natural ecological environment and social environment, the Manchu-Tungusic languages are severely endangered, especially the Manchu and Hezhen languages, which are on the edge of extinction. The vast historical materials of Manchu archives are extremely precious and in need of exploitation and utilization. It is more urgent to rescue the Manchu-Tungusic language and cultural heritage, to build their corpus database, and to study them comprehensively and scientifically.

Key Words: Manchu-Tungusic language corpus database; Manch-Tungusic language group; endangered language