




中国民族语言 工程研究新进展

**Advances on the Minority Language
Processing of China**

 中国社会科学院民族学与人类学研究所

江 荻 孔江平 • 主编



社会科学文献出版社
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

总 序

中国社会科学院民族学与人类学研究所是一个多学科、综合性的研究机构。从学科的设置和专业方向来看，包括了马克思主义研究、历史学、语言学、民族学、社会文化人类学、经济学、宗教学、文献学、政治学、法学、国际关系、影视人类学、民俗学、古文字学等，还包括蒙古学、藏学、突厥学等专门的学问。这些学科和专业方向的多样化构成了研究所的多学科、综合性特点，而这些学科的研究对象则是人类社会民族现象及其发展规律，着重于对不同历史阶段和不同含义的民族共同体（people、ethnos、nationality、nation、ethnic group）及其互动关系的研究，显示了研究所诸多学科的共同指向。研究所以民族学和人类学冠名的目的是为多学科建构一个共同的学术平台，在研究对象统一性的基础上实现多学科的互补与整合，在多学科的视野中建立综合性研究优势，增强中国民族学和人类学的学科性发展。

人类社会的民族现象及其所伴生的民族问题，是人类社会最普遍、最复杂、最长久、也是最重要的话题之一。中国是世界上古代文明延续不断的东方国度，也是统一的多民族国家。在数千年的发展进程中，多民族的互动关系不仅是历朝历代最突出的社会现象之一，而且也是统一的多民族国家形成和不断发展的重要动能。因此，中国几千年来民族现象和各民族的互动关系，为我们解读人类社会的民族现象及其规律性运动提供了一个相当完

整的古代模式。《礼记·王制》中说：

“凡居民材，必因天地寒暖燥湿，广谷大川异制，民生其间者异俗；刚柔、轻重、迟速异齐，五味异和，器械异制，衣服异宜。修其教不易其俗，齐其政不易其宜。中国戎夷，五方之民，皆有性也，不可推移。东方曰夷，被发文身，有不火食者矣；南方曰蛮，雕题交趾，有不火食者矣；西方曰戎，被发衣皮，有不粒食者矣；北方曰狄，衣羽毛穴居，有不粒食者矣；中国、夷、蛮、戎、狄，皆有安居、和味、宜服、利用、背器；五方之民，言语不通，嗜欲不同；达其志、通其欲，东方曰寄，南方曰象，西方曰狄鞮，北方曰译。”

这就是中国先秦文献中所记载的“五方之民”说，可谓中国最早具有民族志意义的记录。它所提示的内涵，对我们今天认识和理解民族现象也是启迪颇多的。例如，构成民族特征的文化差异的自然基础是生态环境，即在“天地寒暖燥湿”、“广谷大川”等不同生态环境中生存的人类群体“皆随地以资其生”所表现的“异俗”，“五方之民”在民居、饮食、服饰、工具、器物等方面的“异制”，不同的语言、不同的价值观念及其相互沟通的中介（翻译）等。其中也包括了处理“五方之民”互动关系的古代政治智慧，即“修其教不易其俗，齐其政不易其宜”。可以说，中国是一个有民族学传统且民族学资源十分丰富的国家。

古往今来、时过境迁，今天的中国已经自立于世界民族之林，正在为实现中华民族的伟大复兴推进中国特色社会主义现代化进程。这一进程正在展示现代民族过程的发展前景，它同样会对现代人类社会的民族现象及其发展前景提供一种范式，也就是中国解决

民族问题的成功例证。当然，我国正处于社会主义初级阶段的发展进程中，在解决民族问题方面我们不仅面对着中国 56 个民族共同发展繁荣的历史重任，而且也面对着全球化时代多民族的大千世界。无论是内政治理，还是融入国际社会，广义的民族问题仍旧是我们需要高度重视的课题。当代中国民族问题的基本特征和普遍反应是经济文化的发展问题，这是由当代中国社会所处的发展阶段及其基本矛盾所决定的。同时，我们也面对着一些棘手的问题，如“台独”问题、达赖集团问题、“东突”势力和国际恐怖主义问题，以及世界范围和周边国家民族问题的交互影响。这两个方面的问题为我们提出了责无旁贷的研究任务。履行这一职责需要我们付出多方面的艰辛努力，其中学科建设是最重要的保障。

科学化是学科建设题中之义，任何一门学科只能在科学化的过程中实现发展。中国的学术传统源远流长，也形成了诸多学科性的研究领域。近代以来，随着西学东渐，中国的学术事业在不断地吸收西方科学规范的过程中逐步形成了现代学科的分化，其中民族学、人类学也取得了很大程度的发展。自 20 世纪 70 年代末中国改革开放以来，中国的哲学社会科学事业在与世界学术领域交流互动的过程中取得了新的发展和显著的成就，哲学社会科学在认识世界、传承文明、创新理论、资政育人、服务社会等方面的不可替代作用，得到了党和国家的充分肯定。但是，能否充分地发挥哲学社会科学各学科的这种作用，涉及诸多因素，而学科建设所包含的指导思想、基本概念和范畴、学科理论、研究方法和学术规范等方面的内容是具有重要意义的。体现这些基本要素的研究成果，不仅对推进学科建设至关重要，而且也是繁荣发展哲学社会科学事业不可或缺的内在条件。中国社会科学院重点学科建设工程的启动，是进一步繁荣发展哲学社会科学事业的重要举措，我所推出中国社会科学院重点学科建设工程丛书·民族学

人类学系列是贯彻落实这一重要举措做出的一种尝试。

如上所述，我所是一个多学科、综合性的研究机构，经过学科调整和研究室重组，所内的大部分学科都纳入了重点学科建设工程，如民族理论、民族历史、民族语言、语音学和计算语言学、民族学（社会文化人类学）、世界民族和诸多专业方向。因此，这套丛书的出版及其所关涉的研究内容也体现了多学科的特点。这套丛书根据基础研究和应用研究并重的学科建设要求，或以学科或以专题反映我所科研人员新近的研究成果。根据中国社会科学院重点学科建设工程的协议要求，在今后几年中我所列入工程范围的学科和专业方向将完成一系列具有重要理论价值和现实意义的研究课题，而这套丛书则主要反映这一过程中的阶段性学术成果。

2003年，我国获得了2008年国际人类学民族学世界大会的举办权，这对中国的民族学和人类学以及广义的民族研究事业来说是一次重大的发展机遇，也是与来自世界各国的民族学家、人类学家进行广泛对话和空前交流的机会。同时，这也意味着挑战。我们不仅需要展现中国各民族的现代发展成就，而且需要在民族学、人类学研究方面推出一批批引人注目的高水平研究成果。因此，加强民族学、人类学的学科建设，整合传统民族研究的学科性资源，做好充分的学术准备是今后几年我国民族学、人类学界的重要任务。从这个意义上说，这套丛书的陆续出版在一定程度上也将体现我所在迎接这次世界大会进行的学术准备。

在此，我们非常感谢社会科学文献出版社对这套丛书的出版给予大力支持和真诚帮助，也期待着广大读者给予关注和指正。

中国社会科学院民族学与人类学研究所所长

郝时远

2004年6月

Foreword of the Series

The Institute of Ethnology and Anthropology of the Chinese Academy of Social Sciences is a multi-disciplinary and comprehensive research institution. In terms of disciplinary arrangement, the institute covers Marxist studies, history, linguistics, ethnology, socio-cultural anthropology, economics, religion, historical records, politics, law, international relations, video anthropology, folklore, and ancient scripts, as well as some special learning like Mongol studies, Tibetology and Turk studies. The disciplinary diversification forms the multi-disciplinary, comprehensive feature of the institute. All these disciplines have one thing in common, that is, they all study *minzu* (a general word in Chinese for people, ethnos, nationality, nation and ethnic group) phenomenon and their evolutionary law in the human society, with the emphasis on *minzu* communities in different historical stages and with different meanings, as well as on the interaction between various *minzu*. The institute is named with the term of ethnology and anthropology for the purpose to build a common academic platform for all disciplines it involves, to realize the mutual complementarity and integration of all the disciplines, to form the advantage of the comprehensive studies, and to foster the development of ethnology and anthropology in China.

The *minzu* phenomenon and the problems resulting from the phenomenon have been among the most widespread, most complicated, most prolonged and most important subjects in the human society. China is an Oriental country with an ancient civilization that never discontinued. Also, it is a unitary country with ethnic plurality. In the course of several thousand years, the ethnic interaction was not only a social highlight in each historical dynasty, but also a motivator for the formation and continuous development of a unitary country with ethnic plurality. So, the *minzu* phenomenon and ethnic interaction in China' long history of several thousand years provide us with a full ancient model for understanding the *minzu* phenomenon and their law in the human society.

In China's Pre - Qin historical literature, there is a term of "*wu fang zhi min*", literally, five - direction peoples. It comes from the following paragraph:

The material used for shelter must vary with different climate, cold or warm, dry or moist, and with different topography, such as wide valley or large river. And people living in different environment have different customs. They may have different character, behaving way, dieting habit, instruments and clothes. It is proper to civilize the people without changing their customs and to improve their administrative system without changing those suitable to them. *Wu fang zhi min* (people inhabiting five directions), either in middle plain or in frontier, all have their own character, which can not be transformed. People in the east, known as Yi, grow long hair hanging down over the neck and have tattoos, and some of them have their food without cooking. People in the south, known as

Man, tattoo their foreheads and cross their feet when sleeping, and some of them have their food without cooking. People in the west, known as *Rong*, grow long hair hanging down over the neck and wear pelt, and some of them do not have grain as their food. People in the north, known as *Di*, wear feather and live in caves, and some of them do not have grain as their food. Both people in middle plain and the *Yi*, *Man*, *Rong* and *Di* have their own shelter, diet, dress, instruments and carriers. The people in five directions can not understand each other and may have different desires. The way to make each other's ideas and desires understood is called *ji* in the east, *xiang* in the south, *didi* in the west and *yi* in the north. (cited from *Liji*, an ancient Chinese book.)

This may be regarded as the earliest record with ethnographical sense in China.

The citation suggests a lot for us to understand today's *minzu* phenomenon. For one thing, the cultural difference that usually constitutes the ethnic feature has its natural foundation in ecological environment. Human groups living in different ecological environment, like different climate (cold or warm, dry or moist) and topography (wide valley or large river), all depend upon their local resources and thus may have different customs. As mentioned above, the people in five directions varied in shelter, diet, dress and instrument, and people speaking different languages with different values can be communicated only through the medium of translation. Also, here is displayed the ancient political wisdom to deal with the relations of the people in five directions, namely, to civilize the people without changing their customs and to improve their administrative system

without changing those suitable to them. So it shows that China is country with ethnological tradition and rich ethnographical resources.

Now, old time has passed and the situation has changed. Today, as an independent member of the international community, China is promoting the modernization with Chinese characteristics in order to realize the great rejuvenation of the Chinese nation. This development has revealed the prospect of the modern *minzu* process. And at the same time, it will also provide a pattern, that is, the successful example in which China deals with the *minzu* problem, for the *minzu* phenomenon and their evolution of the human society in the modern time. Of course, China still remains at the initial stage of socialism. So far as the *minzu* problem is concerned, we are now facing not only the historical task of common development and prosperity for the 56 nationalities in China, but also the ethnically plural, complicated world in the time of globalization.

The ethnic problem in broad sense still remains to be a subject to which we should pay much attention, either in the management of internal affairs or in the merging to the international community. In contemporary China, the fundamental feature of or the widespread response to the ethnic problem is how to promote the economic and cultural development. This is determined by the current developmental stage as well as the fundamental contradiction of the contemporary Chinese society. At the same time, we are confronted with certain difficult problems, such as Taiwan's attempt for "independence", the problem of the Dalai clique, the issue of "East Turkistan", the international terrorism, as well as the influence of ethnic problems both in our neighboring countries and all over the world in general.

The problems in the two larger respects put forth our duty – bound tasks for research. To perform this duty, we should make our effort in many aspects, among which, disciplinary construction serves as the most important guarantee.

Disciplinary construction calls for scientific spirit, only with which can a discipline realize its development. China has its academic traditions of long standing, and a number of academic domains developed in the history. Since the influence of the Western learning went eastward, Western norm of science has been introduced and the disciplinary division in modern sense gradually came into being in China. And it is just in this process that ethnology and anthropology acquired development to large extent. Since the late 1970s when China began to take reforms and open to the outside world, new development and remarkable achievement have been made in China's philosophy and social sciences through the exchange with foreign academia. The Party and the State highly appreciate the irreplaceable role of philosophy and social sciences in understanding the world, passing on civilization, innovating the theory, consulting for government and educating the young, and serving the society.

However, the full play of the role of philosophy and social sciences involves many factors. And in this respect, disciplinary construction is of importance, such as in the guiding thought, fundamental concepts and categories, disciplinary theories, research methods and academic norm. The research achievements that represent these fundamental factors will not only be of vital importance in promotion of disciplinary construction, but also make up the indispensable inherent conditions for prospering and fostering philosophy

and social sciences. The launch of the construction project for prior disciplines at the Chinese Academy of Social Sciences is a significant move for further prospering and fostering philosophy and social sciences. And the Series of the Construction Project for Prior Disciplines at the Chinese Academy of Social Sciences our institute has put out is just an attempt to carry out the significant move.

As mentioned above, our institute is a multi - disciplinary and comprehensive research institution. Since the discipline adjustment and research department restructuring, most disciplines in the institute, such as ethnic theory, ethnic history, ethnic linguistics, phonetics and computational linguistics, ethnology (socio - cultural anthropology) and world ethnic - national studies, have been brought into the construction project for prior disciplines. So, the series and the content involved reflect the feature of multi - disciplines, too. Placing emphasis both on basic and applied studies, the series reflects recent research achievements either in the unit of a discipline or in a special topic. In accordance with the requirement from the agreement on the construction project for prior disciplines at the Chinese Academy of Social Sciences, our institute will complete in the next few years a series of research projects both with important theoretical value and actual significance. So, the series mainly reflects the academic products at the current phase.

In 2003, China succeeded in bidding for the host for the 2008 Conference of the International Union of Anthropological and Ethnological Sciences (IUAES). This will be a significant developmental opportunity, not only to China's ethnology and anthropology, but to

ethno – national studies in general as well. Also, it will be an opportunity of widespread dialogue and unprecedented exchange with ethnologists and anthropologists from various countries in the world. At the same time, however, it means a challenge. We need to exhibit the developmental achievements of the nationalities in China, and moreover, we need to exhibit plenty of striking research achievements with a high level. Therefore, it will be the important task of China's ethnology and anthropology in the next few years to strengthen the disciplinary construction, integrate the disciplinary resources of traditional ethno – national studies, and make full academic preparation. In this sense, the publication of the series in succession can be regarded to some extent as the academic preparation made by our institute for the coming congress of IUAES.

Finally, we appreciate very much the vigorous support and sincere assistance of the Social Sciences Literature Press to the publication of the series. And we also expect the attention and criticism from the readers.

HAO SHIYUAN

Director

Institute of Ethnology and Anthropology

Chinese Academy of Social Sciences

June 2004

前 言

回顾起来，语音学与计算语言学重点实验室迄今已走过了整整 20 个年头。初创当时，机房已经有了一些先进装备，除了一台身躯庞大的国外小型计算机和一大堆打孔卡片之外，还有一套配备台式计算机的语音分析设备（语图仪）。正是在这些设备的支持下，研究室的科研人员同心协力、默默耕耘，历经从小到大、成果积累的漫长阶段，铸就了两个学科的成长，形成了今天中国语言学界颇具影响力的民族学与人类学研究所实验语音学学科和计算语言学学科。

在学术发展的道路上，实验室的两个学科都经历了从描写到理论分析乃至社会应用的过程。

实验语音学初期的目标是关于中国民族语言的语音基本要素分析，解决民族语言学界对各种民族语言语音的定性描述，包括音高或声调，声母和韵母，如鲍怀翘教授等发表《佤语浊送气声学特征分析》（1990），《蒙古语察哈尔话元音松紧的声学分析》（1992），《藏语元音声学分析》（1995）；孔江平教授发表《道孚藏语双擦音声母的声学分析》（1991），《紫云苗语五平调系统的声学及感知研究》（1992），《苗语浊送气的声学研究》（1993）；江荻教授的

《论藏语声调高低分化的噪音机制》(1998)等。这些研究不仅积极支持了民族语言学界深入开展描写和语言历史的研究,而且还为本学科的深度发展铺垫了基础。进一步的发展则开辟了两个新的研究领域,一是注重建立各个少数民族语言的声学参数数据库,二是开展噪音基础理论研究。经过多年积累,研究室已建立了藏语拉萨话的语音声学参数数据库、哈萨克语语音声学参数数据库、蒙古语声学参数数据库,这一系列中国民族语言语音声学参数数据库的建立和研究极大地推动了我国民族语言语音学研究水平,为民族语言语音研究开创了一个崭新的局面。噪音研究方面,孔江平教授承担了自然科学基金和院重点基金多项噪音声学课题,录制了多种民族语言的语音数据,发表多篇研究论文,例如《汉语普通话噪音特征相关分析》,《哈尼语发声类型声学研究及音质概念的讨论》,《汉语普通话噪音抖动及波动的声学研究》,《汉语双音节调位的矢量量化(VQ)》研究等等。近年,实验室还开展了动态腭位的研究,如鲍怀翘、郑玉玲《普通话动态腭位图数据统计分析初探》,以及普通堵塞音、塞擦音成阻持阻时程的统计分析等等。

计算语言学方面,实验室从基本语料开始着手,首先构筑学科建设的基础,所谓兵马未动粮草先行。第一步是研制出面向藏语文本处理的大型电子机载词典和一定数量的标注文本语料,这项研究有力地推动了学科的发展。通过对静态和动态藏语语料的统计研究,摸清了藏语文本的基本句法和信息特征,发表了《藏语计算机统计用语料抽样文本的遴选》(1999)、《书面藏语的熵值及相关问题》(1998),以及《藏文的音韵结构及频度现象》(1998)等多篇论文,为下一

步工作铺好了前进的道路。其次,开展了藏语文本语料库的特征分析,提出了藏语文本词根归一化方法,设计了第一个藏语分词基本规则以及依据句类自动分句的原则。期间还发表了《语篇索引技术在藏文文本中的应用》(1999)、《藏语拉萨话现在时的标记及功能》(1999)、《藏语计算机自动分词的基本规则》(2000)等论文。近年,计算语言学科开始构筑较全面的藏语信息处理的基本平台和工具,包括藏语的排序、检索,句法属性标注集,组块分词方法以及最迫切的面向机器处理的句法本体系统研究。这个时期所发表的论文体现出学科开始走向成熟的趋势,研究内容逐步深化。如《现代藏语的句法组块与形式标记》(2003)、《现代藏语谓语的识别与信息提取》(2003)、《书面藏语排序的数学模型及算法》(2004)等等。这些研究无论从国外研究状况还是从国内研究进展来看,都处在藏语文研究的前沿,也标志着当代藏语计算语言学和语料库语言学的发展趋势。而最近我们开始构建的面向文本处理以及机器翻译的藏汉双语语法信息词典则表现出本学科未来的进程和发展方向。

除了以上两个主要学科之外,实验室研究人员利用信息技术手段积极参与民族语言的应用研究。如郑玉玲副教授主持开展了“藏缅甸语料库与比较研究的计量描写”课题,发表《计算机进行藏缅甸语相关分析尝试》等多篇论文。而在民族语言类型学研究方面,江荻教授协助语言类型学课题组采用语料库语言学的方法进行文本处理,完成了机器抽取词语单位、自动生成文本词表、以及对文本进行词法、句法、语义属性标注,有力地配合了该项目的顺利进行。另一项颇具影响的应用是我所与香港科技大学合作研制出的“汉藏语同

源词研究计算机检索系统”(2002),该系统共收集国内外最重要的语言或方言130余个。其中包括汉语方言、中上古汉语以及南岛语构拟以及藏缅语、苗瑶语、壮侗语、南亚语、南岛语等多种代表性语言或方言。除此之外,系统中还装入了汉语与民族语言对照词典十余部。至于检索方法更是多种多样,汉语检索、英语检索、民族语言检索,还可以采用语义类别检索、组合检索等等。其他应用方面,实验室近年还初步开发了“语言地理信息系统”(2003)、“藏文文本扫描识别系统”(2004)以及“中国民族语言研究论文文摘数据库”(2000)和“《元朝秘史》全文检索系统”等多个项目。

按照重点实验室的五年计划和十年规划,实验室还将开展一系列可持续发展的研究项目。如中国语言的动态腭位分析和韵律分析,民族语言的声学参数数据库,建立部分民族语言的语音图谱档案。计算语言学方面,将在已有研究基础上迈进更深层次的研究项目,包括面向机器处理的现代藏语句法规则研究,藏语机器自动处理研究的平台和工具研制,现代藏—汉—英多语机器自动翻译系统,以及更基础性的机载藏语语法信息电子词典和现代藏语本体知识体系研究。而在语言应用研究方面,实验室希望通过中国民族语言资源联盟的建设项目,联合国内研究机构和高校对我国民族语言资源进行全面汇总,建立技术开发和支持平台,该项目的实施将使我国形成最具完整性、系统性、规范性和权威性的中国民族语言国际性资源基地。并可与国际“语言资源联盟”以及国内“汉语资源联盟”全面链接。

在20年的创业发展过程中,实验室相继培养了一批批出色的科研人才。曹雨生教授积极为年轻人创造研究机遇,争

取项目，开展学术交流，使实验室呈现朝气蓬勃的发展局面。现任国际语音学会理事的鲍怀翘教授退休后仍一直活跃在语音学界，有力地带动了实验室的语音研究工作。沈米遐工程师严谨的工作作风和宽厚的为人理念可以说是实验室团队前进的粘合剂。陈嘉猷副教授是实验室资源和设备的总管，多年来不曾出过毫厘差迟，而这股韧劲在研究上也表现出特有的认真，保证了实验数据的准确和可靠。郑玉玲副教授可以说是科研上的多面手，她开发研制的藏缅语数据库曾荣获国家科技优秀成果奖，而今她进一步在语音韵律和腭位分析方面努力，始终保持学术研究的进取精神。孔江平教授长期致力于研究我国语言的嗓音声学分析项目，他用高速摄影技术采集的动态声带振动影像令学界叹为观止，而他所构建的中国语音嗓音分析方法将深刻影响未来中国语音研究的方向。江荻教授特别执著于藏语的计算处理，从最基础项目做起，逐步构建出藏语自然语言处理的基本平台、操作框架和理论。

这本论文集汇集了实验室最新的一些研究成果，一定程度上反映了我们的科研动向和研究范围。其中还有我院研究生院研究生或高校交流研究生的创作。论文涉及的范围正如实验室的科研方向，包括了计算、语音和应用三个方面，因此分三个部分编排。作为首次出版的实验室论文集，我们特别编辑了近年来我们发表的主要论文目录作为附录，以方便今后的学术交流。

在这本论文集出版之际，我们要感谢中国社会科学院科研局的支持和资助，感谢研究所领导的帮助，郝时远所长自始至终关心和督促论文的汇集和编辑，亲自帮助联系出版事宜，撰写序言。黄行副所长不仅参与课题研究，还帮助协调