

第二届
汉语中介语语料库建设
与应用国际学术讨论会
论文选集

崔希亮 张宝林 主编



北京语言大学出版社
BEIJING LANGUAGE AND CULTURE
UNIVERSITY PRESS

第二輯

江蘇省分級資料館建設
與應用國際學術研討會

論文選集

劉曉明 曹淑蘭 主編



江蘇省圖書館
Jiangsu Provincial Library
100010 南京

第二届
汉语中介语语料库建设
与应用国际学术讨论会
论文选集

编辑委员会主任：崔希亮

主 编：崔希亮 张宝林

编 委：（按姓氏音序排序）

崔希亮 黄 伟 肖奚强

邢红兵 张宝林 张旺熹 张维佳



北京语言大学出版社
BEIJING LANGUAGE AND CULTURE
UNIVERSITY PRESS

图书在版编目 (CIP) 数据

第二届汉语中介语语料库建设与应用国际学术讨论会
论文选集 / 崔希亮, 张宝林主编. — 北京: 北京语言大学
出版社, 2013.12

ISBN 978-7-5619-3748-8

I. ①第… II. ①崔… ②张… III. ①汉语—中介语
—国际学术会议—文集 IV. ①H1-53

中国版本图书馆 CIP 数据核字 (2013) 第 308632 号

书 名: 第二届汉语中介语语料库建设与应用国际学术讨论会论文选集
DI-ER JIE HANYU ZHONGJIEYU YULIAOKU JIANSHE YU YINGYONG
GUOJI XUESHU TAOLUNHUI LUNWEN XUANJI

责任印制: 姜正周

出版发行: 北京语言大学出版社

社 址: 北京市海淀区学院路 15 号 邮政编码: 100083

网 址: www.blcup.com

电 话: 编辑部: 010-8230 3647/3592/3395

国内发行: 010-8230 3650/3591/3648

海外发行: 010-8230 0309/3365/3080

读者服务部: 010-8230 3653/3908

网上订购: 010-8230 3653 (国内)/3668 (海外) service@blcup.com

印 刷: 北京京华虎彩印刷有限公司

经 销: 全国新华书店

版 次: 2013 年 12 月第 1 版 2014 年 3 月第 1 次印刷

开 本: 710 毫米 × 1000 毫米 1/16 印张: 26.5

字 数: 466 千字

书 号: ISBN 978-7-5619-3748-8 / H · 13374

定 价: 68.00 元

凡有印装质量问题, 本社负责调换。电话: 010-82303590



目 录

大会致辞

第二届汉语中介语语料库建设与应用国际学术讨论会致辞	崔希亮	3
第二届汉语中介语语料库建设与应用国际学术讨论会致辞	陆俭明	5
汉语中介语语料库建设与应用的两个重要特色		
——第二届汉语中介语语料库建设与应用国际学术讨论会致辞		
	冯志伟	9
第二届汉语中介语语料库建设与应用国际学术讨论会致辞	李晓琪	12
第二届汉语中介语语料库建设与应用国际学术讨论会致辞	[韩国] 孟柱亿	15

论文综述

第二届汉语中介语语料库建设与应用国际学术讨论会论文综述	张宝林、黄伟	19
-----------------------------	--------------	----

汉语中介语语料库建设研究

汉语中介语语料库建设面临的任务与对策	张宝林、崔希亮	31
语料库的标注和它的局限性	冯志伟	44
国别化汉语中介语语料库的基础标注和偏误标注	胡晓清、许小星	67
汉语中介语语料库语用标注研究	杨雯	76
汉语中介语综合测度指标及中介语发展语料库的建设	曹贤文、吴从容	88

汉语口语中介语语料库建设刍议	刘运同	99
中介语语料库之建置及应用语料库于华语教学之实证研究		
	(中国台湾)方淑华、王琼淑、陈浩然	107
口语文本语料库检索系统的建立与功能呈现	李航	119
基于语音识别的 HSK 口语语料库研究	刘喆	135
TOCFL 作文语料库的建置与应用	(中国台湾)张莉萍	143
基于偏误反馈的对韩汉语词汇教学信息库的建设	禹德才、胡晓清	155

基于语料库的汉语句式与句法习得研究

基于“HSK 动态作文语料库”的“帮”字句研究	郭曙纶	169
留学生“比”字句偏误类型的分析与习得研究	陆一萍	179
基于“HSK 动态作文语料库”的韩国留学生“比”字句偏误原因分析	李桑	186
基于“HSK 动态作文语料库”的“给”字句习得研究	屈梅娟	201
基于“HSK 动态作文语料库”的外国学生“使”字句回避类偏误研究	郑云霞	212
同义构式“连 XP 都 VP”与“连 XP 也 VP”及其习得研究	朱庆洪	222
韩国学生中介语各句长句子定语复杂度发展研究	肖奚强、黄自然	233
基于越南学生汉语中介语语料库的“上”组趋向补语习得研究	[越南]刘汉武	245
以台湾地区学习者语料库进行“得”字情态补语句之偏误分析	(中国台湾)潘依婷、陈浩然	254
基于“HSK 动态作文语料库”的外国人汉语程度补语习得研究	张颖	265
接触频率对汉语形名、量名搭配习得的影响	张力力	276

基于语料库的汉语字词习得研究

母语背景因素对中高级水平学习者书写汉字的影响

——以日、韩、欧美学生为例 黄伟 289

越南留学生形容词使用偏误分析及教学建议

——以肖像描写写作教学为例 刘荣、李步军 298

基于国别化语料库的“大”的易混淆词的混用表现及成因分析

苏向丽、毛嘉宾 309

基于中介语作文语料库的连词“还是”偏误分析 孙慧莉 321

“互相”语义的意象图式 吴丽、宋春阳 331

句法与概念——以语料库方法对英汉中介语语法中 De 与 Wh- 特征的

对比研究 [英国] 马韬 340

基于语料库的其他相关研究

基于口语语料库的中级阶段留学生汉语韵律研究 [马来西亚] 周宝芯 355

关于自动偏误分析在二语习得中的效果——日本公外基础汉语教学系统

“游”的尝试 [日本] 汤山土美子 368

HSK 动态作文语料库和我们的“假说验证型”研究模式

——以日本学生习得领格标记“的”为例 [日本] 张麟声 383

浅析本科留学生汉语语料库在汉语教学和研究中的应用 张俊萍、李焱 ... 394

教师课堂提问中词汇超纲现象的考察 郑贻瑜 401

附录 411

后记 415

大会致辞



第二届汉语中介语语料库建设 与应用国际学术讨论会致辞

崔希亮

北京语言大学

尊敬的各位嘉宾、各位代表：

大家早上好！

我很高兴作为这次学术研讨会的大会主席第一个发言。这是我们第二届汉语中介语语料库建设与应用国际学术讨论会，首先请允许我代表北京语言大学、代表我的各位同事，向远道而来的各位学者表示热烈欢迎！首届学术研讨会是两年以前，比现在稍早，在南京师范大学由肖奚强教授和我们一起主办的，盛况空前，有许多学者参加。两年以来，我们在汉语中介语语料库的研究方面，取得了很多的成果，尤其是很多新的项目建立起来了，而且现在海内外的很多学者都非常关注我们这个研究。我们的“HSK 动态作文语料库”自从放在网上供大家免费使用之后，使用的人很多，我们非常高兴。因为语料库是一项基础工程，建立之后不能束之高阁，那样毫无价值，免费使用后我们可以满足学者、研究者，包括博士生、硕士生的研究需要，比如提供单机版的语料库。但是我们希望，这也是一个小小的要求，语料库建设不仅仅是某几个人的事情，而是大家的事情，语料库建成之后为大家所用，所以语料库建设当中也需要大家做贡献。因此我们为大家提供了一个完全的访问使用权，希望大家为我们提供语料，然后由我们的工作组对语料进行深度加工，我们会改善我们的搜索工具，使语料库更加好用。我也在使用“HSK 动态作文语料库”，在使用中也发现了一些不好用的地方，因为每个人的需求都是不一样的，语料库不能考虑得那么周到，但是有很多地方我们可以慢慢地深化，比如语法项目的标注还不够细，有很多地方没有标注。

最近我们又申请到了一个教育部社科基金的重大攻关项目——“全球汉语中介语语料库建设和研究”的建设项目，此项目在人文社会科学项目里边，算是

资助力度比较大的项目，当然和理工科是无法比的。我们想借助国家现在正在推行的“2011 协同创新计划”，把我们的语言资源库、语料库和相关的基础研究工作一起做起来，搭建一个平台。我们学校，由我牵头，曹志耘副校长和张旺熹教授，大家正在一起思考“2011 协同创新计划”怎么做。可能有些学者还不很清楚“2011 协同创新计划”是一个什么样的计划，它和“985”“211”不是同样性质的计划，国家现在的目的是希望能够在科学、技术以及人文社会科学方面推动建设一些协同创新的科研中心。所谓协同创新的科研中心，它的目标一个是解决国家急需的课题，第二个是培育世界一流的研究团队。语料库的建设应该是国家急需的，但是语料库只是其中之一，并不是我们工作的全部。我们汉语国际教育以及中国文化走出去，是国家发展战略方面非常重要的工作，所以北京语言大学准备在这个方面做些工作。目前我们得到了一些兄弟院校和相关部门的支持，他们都希望和我们合作。现在有很多单位都希望和我们一起来做，但是有一个前提，就是要互补。我想我们的语料库建设在两年以前就已经开始了协同创新，应该说开始得比较早。在这个基础上，很多事情我们能进一步地做得更好。所以今天在这个研讨会的开幕式上，我代表我的同事表一个态：我们愿意继续承担国际汉语或者说全球汉语中介语语料库的研制工作，尤其是在深度开发方面。我们有很多人，如语言学及应用语言学、对外汉语方向的博士生和硕士生，都愿意参与到这样的工作里边来，还有我们信科院的荀恩东教授（今天也在座）一直很支持我们的这个工作，我们很多检索工具都是由他们开发的，还有平常网站的管理、更新、维护都是由他们在做。我们的汉语水平考试中心现在由邢红兵教授主持工作，还有二十几位同仁也非常支持这些工作。我在这里代表我们课题组、代表学校，向对语料库建设做出贡献的海内外专家学者表示衷心的感谢。

今天参加这次研讨会的人数并不很多，共有 80 位，但是来自八个国家和地区。目前是举办学术研讨会非常频繁的时间，尤其是八九月份，是北京气候最好的时候，我们正好又马上就要庆祝北京语言大学建校 50 周年。在这样一个特殊时刻，我也代表学校邀请在座的各位专家学者在九月九号上午来参加我们的庆典活动，九月九号下午是大学校长论坛，主题就是“汉语的第二语言教育”或者说“汉语作为外语的教育问题”，会有来自 50 个大学的校长发表高见，所以借此欢迎大家参加，或者旁听，或者发表意见。

谢谢大家！祝这次研讨会圆满成功！



第二届汉语中介语语料库建设 与应用国际学术讨论会致辞

陆俭明

北京大学中文系

校长、各位专家学者、各位老师、各位同学：

首先我要热烈祝贺第二届汉语中介语语料库建设与应用国际学术讨论会在北京语言大学隆重举行。

汉语教学发展到今天，迫切需要建设完备的汉语中介语语料库，而且要充分利用这语料库来进一步改进汉语教材，革新汉语教学方法，提高汉语教学的质量。大家知道，汉语中介语语料库的建设与应用，有助于我们客观了解汉语第二语言学习者的汉语习得表现和发展过程，有助于我们深入揭示学习者汉语习得的内部规律和机制，而这是我们进一步改进汉语教材、革新汉语教学方法、提高汉语教学质量所依赖的基础。因此汉语中介语语料库的建设与应用，对推进汉语教学，对推动汉语教学的本体研究，对让汉语健康、稳步地走向世界，对汉语的本体研究，都具有重要的作用和意义。

我们知道，中国人学习外语也好，外国人学习汉语也好，都不可能一步到位，都不可能将所学的语言一下子就学会并掌握了，都会经历一个在学习过程中不断克服语音、用词、成句或书写等诸方面种种不合规范的毛病甚至错误从而不断进步，逐步学会目的语这样的过程。因此我们在汉语教学的过程中，除了正面教给学习者该学、该掌握的东西外，还得有针对性地帮助他们克服在学习过程中不断出现的这样那样的毛病与问题。那么怎么才能做到有针对性地并且有实效地帮助学习者克服在学习过程中不断出现的这样那样的毛病与问题呢？最早，譬如上个世纪 70 年代之前，在“以教为中心”（teaching perspective）理论的指导下，一般致力于对比研究。对比研究，即对比研究学习者的母语和教师所教的目的语二者之间语音、词汇、语法等诸方面的异同。人们以期通过这样的对比研究来获得指导学生、解决学习者语言问题的钥匙。这个办法有一定作用，但实际效果并非如预期那么好。原因是受当时学术水平的局限，人们的思维普遍受到风靡学界

的结构主义语言学理论的束缚,因而只是局限于两种语言系统的静态对比研究,以为只要将学习者的母语与目的语这两种语言的相同点、不同点归纳出来,就能预测学习者的学习难点,就能解决学习者在学习过程中所出现的种种问题,而没有深入思考学习者在学习过程中为什么会出现母语的负迁移和目的语的负迁移,更没有考虑来自同一母语、由同一位老师教、大家都在一起学习的学习者为什么在母语的负迁移和目的语的负迁移上会出现不同的情况,即为什么各个学习者的学习难点并不一样;再说,由于语言现象的复杂性,研究者无论怎样认真细致,两种语言的静态对比研究也不可能真正获得十分理想的、能为语言教学所需要的两种语言的异同规则,往往是教师刚将某个语言规则教给学生,学生却一用就错。事实上,语言教学中学习者出现的种种语言上的问题,在相当大的程度上与学习者的心理因素有关,而以往那种对比研究和偏误分析的最大问题就在于没有考虑到学习者这一主体的心理因素以及学习过程中可能有的种种变数。于是,在上个世纪 60—70 年代国外学者提出了中介语理论,其中最具代表性的是 Larry Selinker 1972 年发表的《中介语》(Interlanguage)一文。

按照中介语的理论观点,我们在教学过程中,教学目光不仅要投向学习者的母语系统和目的语系统,更要注意投向每个学习者所特有的、而且在学习过程中不断变动的语言系统,在这个特殊的语言系统中既有偏误的一面,更有正确的一面,这个特殊的语言系统就是中介语系统。

对于这个中介语系统,从上个世纪 70 年代至今,国内外学界有不少讨论,存在着不同的看法与意见。不过有几点认识还是共同的:

第一,都承认任何外语/第二语言学习者都存在中介语系统,这是一个既不同于母语、也不同于目的语的独立的语言系统。

第二,这个语言系统是可观察的,又具有一定的抽象性。换句话说,中介语既可以视为学习者的言语行为或言语表达,也可以视为潜藏在学习者言语行为之后的某种心理过程。

第三,这个中介语语言系统是学习者在特定的时点所构建的,而且会随着学习的进展逐渐向目的语正确形式靠拢,因此具有动态性。

第四,这种中介语系统,实际上都有一个潜在的心理结构(latent psychological structure)和潜在的语言结构(latent language structure)作为基础。

第五,影响中介语的潜在的心理结构主要由以下几方面心理过程构成:语言的迁移过程,包括训练造成的迁移过程;目的语的泛化过程,包括目的语语言规则和语言材料的泛化过程;学习者的学习策略;学习者的交际策略;语言学习过

程中出现的石化现象 (fossilization, 也有人翻译为“僵化现象”, 也有人翻译成“化石化现象”), 即停滞现象 (stagnation) 等。

中介语理论的提出使外语 / 第二语言教学发生了重大的理论导向的变革, 那就是从传统的“以教为中心”(teaching perspective) 向“以学为中心”(learning perspective) 的转变, 而且改变了人们对于学习者偏误只具负面性的看法。

在我国内, 最早引进中介语理论的是鲁健骥教授 (1984), 随后越来越多的学者思考、研究中介语问题。可是, 有了中介语理论, 还得考虑这样的问题: 如何有效运用中介语理论来为我们的汉语教学服务。汉语中介语语料库的建设正是出于这样的考虑, 逐步酝酿、建设与实施的。具体说:

第一, 建设汉语中介语语料库可以改变只在单一环境下只用某种单一方法搜集中介语语料的状况, 可以更好、更全面地搜集、反映汉语学习者所出现的并在不断变化的汉语中介语的真实面貌。

第二, 不仅可以更好地对汉语中介语进行系统性研究, 而且可以更好地进行重复性研究。

第三, 可以更好地进行汉语中介语的定量分析和定性分析, 而且使中介语研究由主观定性分析转向数据驱动下的定性定量结合型分析, 从而更好推进中介语研究。

在我国内, 第一个建设汉语中介语语料库的是北京语言大学的前身北京语言学院, 那就是 1995 年由储诚志、陈小荷主持建立起来的汉语中介语语料库系统。20 年过去了, 如今研究中介语的学者越来越多, 已形成了一支不小的研究队伍; 在汉语中介语语料库建设方面也很有成绩, 继北语之后, 目前已有暨南大学、南京师范大学、中山大学、鲁东大学、上海交通大学、厦门大学等多所高校建设了不同规模、不同标注程度、不同内容的汉语中介语语料库, 听说还有多所高校也在建设汉语中介语语料库; 而在这一时期, 英语教学界也在多所高校建立了中国人学习英语的中介语语料库。相应的, 应用中介语语料库开展语言教学研究、语言习得研究和汉语本体研究的研究成果也日益增多。2010 年北京语言大学和南京师范大学联合发起举办了“首届汉语中介语语料库建设与应用国际学术讨论会”。总之, 从上个世纪 80 年代至今, 有关中介语的理论引介和汉语或英语基于中介语语料库的实证研究都呈现不断增长的趋势。

虽已有不小的成绩, 但从目前的情况看, 无论是汉语中介语语料库的建设与研究, 也无论是汉语中介语语料库的应用研究, 还是中介语本身的研究, 都远不能适应和满足汉语教学发展的需要。真要发挥中介语理论和中介语语料库在汉语教学中的效应, 一方面需要加大财力上的投入, 因为建设一个高质量的语料库是

很花钱的，更需要进一步以科研引航，开展多方面的理论与实证研究，包括软件开发。譬如，单就汉语中介语语料库建设来说，我想就需要在以下几方面做深入研究与思考：

第一，汉语中介语语料库的核心是能真实反映汉语学习者学习进展状况的日趋完备的汉语中介语语料。如何更全面、细致、有效、科学地搜集中介语语料，以确保资源的优质？

第二，如何对语料进行更为科学、合理的标注？该从哪些方面去标注？什么是最理想、最醒目的标注？应该标注到什么程度？如何确保标注的一致性？如何能逐步实施机器自动标注？

第三，如何实现汉语中介语语料库的多模态，即如何建设具备较好的资源交互性、集成性，具有很强的资源库所必需的各种界面，特别是音频、视频等的可控性，使语料库具备健全的功能体系，从而全面提升语料库的效能？

第四，除了综合性的总库外，如何建设不同视角、不同层面、不同方位的各个分支库，即如何建设能分别呈现语音、文字、用词、语法、篇章等不同领域，能分别呈现不同国别、不同母语语种、不同性别等这样的汉语中介语分支语料库？

第五，如何建设专业搜索引擎，实施便捷的检索？如何使语料库能滚动式扩展，具有很强的再生能力？如何增强语料库的调集功能？如何使语料库的操作、控制尽可能简单？

第六，如何逐步实施统一规划，总体设计，多方参与，跨地区、跨领域协作，以克服各自为政、低水平重复开发与建设的弊端，从而实施资源共建和共享？

就中介语理论本身而言，也有许多问题有待进一步研究与探索，以使之逐渐深化，比如：中介语的起点与终点问题、中介语的描写方法问题、中介语的系统性与变异性问题、中介语石化现象的本质问题，等等。

本届研讨会，从所列的9个方面的议题看，其目的就是希望通过大家的各抒己见、畅所欲言，探讨在中介语理论的深化、中介语语料库的建设、中介语语料库的应用等方面怎样真正做到科研引航，并能获得更好的共识。

我对中介语没有研究，但对有关中介语理论和中介语语料库方面的论著比较关心，乐意学习。张宝林教授一定要我在开幕式上讲话，不好推辞。以上算是我的一点学习心得，跟大家交流，请大家批评指正。

最后，预祝研讨会圆满成功！谢谢大家！



汉语中介语语料库建设与应用的两个重要特色

——第二届汉语中介语语料库建设与应用国际学术讨论会致辞

冯志伟

杭州师范大学外国语学院

2010年在南京师范大学召开首届汉语中介语语料库建设与应用国际学术讨论会之后，我国的汉语中介语语料库研究取得了很大的进展，汉语中介语语料库的研究已经在国家社会科学基金项目得到重视，肖奚强教授、胡晓清教授、张宝林教授提出的关于汉语中介语语料库研究的课题先后在国家社会科学基金获得立项。这是值得我们庆幸的！

在2010年的首届汉语中介语语料库建设与应用国际学术讨论会上，我应邀做了“从语料库中挖掘知识”的主题发言。会议之后，我一直关注汉语中介语语料库的研究。我相信，第二届汉语中介语语料库建设与应用国际学术讨论会一定会取得圆满的成功！

我觉得，汉语中介语语料库建设与应用有两个重要的特色。

第一，汉语中介语语料库建设与应用是一种跨语言的研究。

汉语中介语语料库要研究不同语言背景的学生学习汉语的特点，这就要求研究者不仅要懂得汉语，还要懂得其他语言，有了其他语言作为研究的参照系，有助于加深我们对汉语的认识。

据中国互联网络信息中心（CNNIC）统计，截至2008年6月底，我国的互联网网民人数已经达到2.53亿，超过了美国的网民人数，成为了世界上互联网用户最多的国家。

接着，中国互联网络信息中心连续报道了我国互联网的发展情况。下面是中国互联网络信息中心报道的一些数据：

——2008年6月，中国的网民已经达到2.56亿。截至2008年12月31日，我国网民数达到2.98亿人，互联网普及率达22.6%。宽带网民规模达到2.7亿人，占网民总体的90.6%。

——截至2010年5月，我国网民的数量已经达到4.04亿之多，使用手机

上网的网民达到 2.33 亿人。我国 99.1% 的乡镇和 92% 的行政村接通了互联网, 95.6% 的乡镇接通了宽带, 3G 网络已基本覆盖全国。2009 年, 我国电子商务交易总额突破 4 万亿元。互联网已经成为我国经济发展的火车头。

——到了 2011 年, 我国网民的数量已经达到 4.44 亿之多, 互联网渗透率为 37.2%, 每 3 个人中就有 1 个网民。

这些数据说明, 我国已经成为了首屈一指的互联网大国。

目前, 在互联网上除了使用英语之外, 越来越多地使用汉语、西班牙语、德语、法语、日语、韩国语等英语之外的语言。从 2000 年到 2010 年, 互联网上使用英语的人数仅仅增加了 301.4%, 而在此期间, 互联网上使用俄语的人数增加了 1825.8%, 使用葡萄牙语的人数增加了 990.1%, 使用中文的人数增加了 1476.7%, 使用法语的人数增加了 398.2%。

据统计, 在 2011 年, 日语的网络渗透率已经达到 78.4%, 德语的网络渗透率已经达到 79.5%。尽管汉语 2011 年的互联网渗透率只有 37.2%, 但是也高于世界上十大网络语言的平均网络渗透率 36.4%, 仍然是一种非常重要的网络语言。

由于互联网上使用英语之外的其他语言的人数增加得越来越多, 英语在互联网上独霸天下的局面已经彻底被打破, 互联网确实已经变成了“多语言的网络世界”(multilingual web)。因此, 网络上的不同语言之间的翻译和信息处理自然也就越来越迫切了。

“多语言”这个特性使得互联网变得丰富多彩, 同时也造成了不同语言之间交流和沟通的困难, 互联网上的语言障碍问题显得越来越突出, 越来越严重。因此, 网络上的不同自然语言之间的计算机自动处理也就变得越来越迫切了。网络上多语言的机器翻译(machine translation)、信息抽取(information extraction)、文本数据挖掘(text data mining)正在迅猛地发展。语言辨别(language identification)、跨语言信息检索(cross-language information retrieval)、双语言术语对齐(bilingual terminology alignment)和语言理解助手(language comprehension aids)等多语言的在线处理技术(multilingual online processing)已经成为了新一代互联网和语义网络(semantic web)技术的重要支柱。

在这个多语言的网络世界, 我们应当重视多语言的跨语言研究。汉语中介语语料库建设与应用不仅要研究汉语, 还要研究外语。建立“国别化”(country-specific)的中介语语料库, 具有明显的跨语言的特色, 正好适应了多语言的网络世界对语言学的要求。

这是汉语中介语语料库建设与应用的一个特色。