



第三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

Disanjie Hanyu Zhongjiejyu Yuliaoku Jianshe yu Yingyong Guoji Xueshu Taolunhui Lunwen Xuanji

林新年 肖奚强 张宝林 主编

世界图书出版公司

第三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

林新年 肖奚强 张宝林 主编



世界图书出版公司

北京·广州·上海·西安

图书在版编目 (C I P) 数据

第三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 / 林新年, 肖奚强, 张宝林主编. —北京: 世界图书出版公司北京公司, 2016.9

ISBN 978-7-5100-8090-6

I. ①第… II. ①林… ②肖… ③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 (2016) 第226852号

主 编: 林新年 肖奚强 张宝林
责任编辑: 武传霞

出版发行: 世界图书出版公司北京公司

地 址: 北京市东城区朝内大街 137 号

邮 编: 100010

电 话: 010-64038355 (发行) 64033780 (客服) 64033507 (总编室)

网 址: <http://www.wpcbj.com.cn>

邮 箱: wpcbjst@vip.163.com

销 售: 新华书店

印 刷: 北京建宏印刷有限公司

开 本: 787 mm × 1092 mm 1/16

印 张: 29.5

字 数: 514 千

版 次: 2016 年 10 月第 1 版 2016 年 10 月第 1 次印刷

定 价: 72.00 元

版权所有 翻印必究

(如发现印装质量问题, 请与本公司联系调换)

编辑委员会

主任：崔希亮

主编：林新年 肖奚强 张宝林

编委：（以姓氏拼音为序）

陈 晟 崔希亮 黄 伟

林君峰 林新年 肖奚强

颜 明 张宝林 周文华

目 录

大会致辞及总结

- 第三届汉语中介语语料库建设与应用国际学术讨论会大会致辞
..... 崔希亮 (003)
- 汉语中介语语料库建设与应用中的计量研究
——第三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 (005)
- 汉语中介语语料库建设与应用大有可为
——第三届汉语中介语语料库建设与应用国际学术讨论会总结报告
..... 肖奚强 (011)

论文综述

- 第三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
..... 林新年 (017)

汉语中介语语料库建设研究

- 谈汉语中介语语料库的建设标准……………张宝林 崔希亮(029)
- 论汉语中介语语料库建设的多样性和层次性……………周文华(046)
- 汉语中介语语料库建设和二语习得研究:回顾与展望
……………曹贤文(059)
- “非汉字文化圈国家学生错别字数据库”网络应用平台
建设研究……………戴媛媛(072)
- 对外汉字教学语料库建设的构想与尝试……………郭 菁(079)
- “TNR 汉语学习者偏误语料库”的开发与实践 ……于 康(089)
- 论汉语中介语语料库标注的全面性及类别问题
……………肖奚强 周文华(104)
- 汉语中介语语料库偏误标注的层次及类型 ……焉德才(117)
- TOCFL 学习者语料库的偏误标记 ……张莉萍(131)
- 汉语中介语语料库的零位标注……………陈海峰(160)
- 汉语中介语语料库分词规范的研制与思考……………李子月(171)
- 先秦汉语分词规范和词性标记集研究及实验……………梁社会(184)
- 常用口语转写系统的比较……………刘运同(199)
- 汉语中介语口语语料转写规则初探……………张 蕾(205)
- 试论中介语研究中偏误来源的分类……………颜 明 肖奚强(216)

基于中介语语料库的语言习得及应用研究

口语语料库为本的发音偏误分析：以日籍汉语学习者的韵母为例

.....方淑华 陈庆华 杨惠媚 陈浩然 (231)

汉字笔画的信息量与外国留学生语篇环境下汉字字形的再认

.....张小峰 (244)

基于语料库的汉语二语关系从句产出研究.....李金满 (252)

基于自然语料的越南学生汉语“了”的偏误分析

.....何黎金英 (Hà Lê Kim Anh) (266)

越南留学生汉语体标记“了₁”的习得研究 ...彭 臻 周小兵 (291)

基于中介语语料库的“以后”“之后”“而后”“往后”

多维度辨析.....林琳琳 (303)

基于语料调查的“连……都(也)……”出现的语义背景分析

.....张江丽 (312)

学习者语料库在混淆词语研究中的效用评估.....李 华 (322)

越南学生汉语词汇偏误调查统计分析

.....丁氏红秋 (Dinh Thi Hong Thu) (335)

HSK 动态作文语料库中“挑选”与“选择”的使用考察

.....郭曙纶 (349)

日本学生汉语量词“个”与“种”混用现象分析

——基于 HSK 动态作文语料库的专项调查.....林新年 陈 晟 (363)

韩国学生汉语含“见”词语习得顺序研究

..... 沈规利 宋春阳 李 琳 (373)

基于汉语中介语语料库学习者泛义动词“弄”“搞”

使用情况的考察..... 乔 侠 (387)

汉语同义动词“拥有”与“具有”研究..... 陈梦恬 (402)

基于汉语中介语语料库的外向型词语辨析词典的搭配设计原则

..... 孟 凯 (414)

论汉语书面语篇篇章结构特征分析的支撑点与维度

——基于 HSK 动态作文语料库的研究 娄开阳 (432)

基于中介语语料库的日本留学生论证性语篇信息结构习得研究

..... 张迎宝 (443)

大会致辞及总结

第三届汉语中介语语料库建设与应用 国际学术讨论会大会致辞

崔希亮

(北京语言大学)

各位代表：

在炎热的酷暑季节，在热情似火的双福之地——福建福州，我们迎来了“第三届汉语中介语语料库建设与应用国际学术讨论会”。我谨代表大会主办方，向来自不同国家和地区的各位代表表示热烈的欢迎和亲切的问候！

我们这个会议已经召开了三届。首届会议由我校与南京师范大学联合主办，于2010年7月在南京召开。第二届会议作为我校50周年校庆系列庆祝活动之一由我校主办、南京师范大学协办，于2012年8月在北京召开。今天召开的第三届会议，除我校和南京师范大学以外，福建师范大学也加入了主办者行列。主办者队伍的壮大，表明汉语中介语语料库的学术影响在不断扩大，得到了越来越多的院校的重视，我们为此感到非常高兴，并期待有更多院校加入到主办者行列中来。

会议的宗旨是一贯的，即：交流海内外汉语中介语语料库建设与应用研究的最新成果，探讨相关问题，促进学界的交流与合作，进一步推动汉语中介语语料库的建设与基于语料库的汉语教学研究的发展，从而为汉语国际教育服务。前两届会议很好地实现了会议的预定目标，汉语中介语语料库的建设与相关研究得到了跨越式的发展。

我昨天粗略地看了一下会议手册，发现和前两届会议相比，本届会议的论文选题和研究倾向有一些重要变化。

首先是语料库的建设研究受到越来越多的关注，论文数量大幅增加。首届会议这方面的论文只有8篇，第二届会议增至18篇，而本届会议是27篇。这和近年来投入建设的语料库和建成的语料库日渐增多的发展趋势是完全一致的。

其次是语料库建设方面的研究越来越细致,越来越深入。在27篇论文中,宏观探讨语料库建设的7篇,讨论语料标注的10篇,研究汉语中介语语料库分词规范的4篇,研究口语语料转写规范的3篇,研究汉字库、口语库、多模态语料库建设的各1篇。汉语中介语语料库的建设标准、标注规范、专用分词规范及专用词表、口语语料和多模态语料的转写规则,这些方面的研究是我们在教育部重大攻关项目“全球汉语中介语语料库建设和研究”中提出的语料库建设的本体研究的重要内容。本届会议有这么多论文做这些方面的探讨,无疑将使相关研究更加深入。我们为此感到由衷的高兴。请我们课题组包括子课题组的成员在论文中标明课题名称和批准号,以作为我们课题组研究成果的组成部分。

毫无疑问,建设语料库的目的是应用。基于语料库的汉语教学研究、偏误分析与习得研究一直是我们这个会议的重要内容,这次会议仍然有很多此类文章。这些文章的研究水平也在不断提升,为汉语教学提供了很多见解独到、数据翔实、极富参考价值的研究成果。

我们这个会议历经5年,连开3届,已经成为汉语国际教育领域众多学术会议中的一个知名品牌,而且特色鲜明,是一个专门性的会议。历届会议参会人数虽然不是很多,但远见卓识迭出,学术成果丰硕,极大地推动了汉语中介语语料库的建设与应用研究,在汉语的教学与相关研究中发挥了很好的作用,在一定程度上促进了教学研究与习得研究模式的转变和研究水平的提高。这和各位参会代表的积极努力密切相关,和学界的大力支持密切相关。我们希望各位与会代表一如既往地继续支持我们这个会议;希望各主办方精诚合作,团结奋斗,把我们这个会议继续办下去,并且越办越好,让我们这个品牌越来越响亮,在汉语教学和研究中发挥越来越大的作用。

预祝大会圆满成功!

汉语中介语语料库建设与应用 中的计量研究

——第三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）

“首届汉语中介语语料库建设与应用国际学术讨论会”于2010年在南京由北京语言大学与南京师范大学联合主办，“第二届汉语中介语语料库建设与应用国际学术讨论会”于2012年在北京由北京语言大学主办、南京师范大学协办。今天，我们在福州召开“第三届汉语中介语语料库建设与应用国际学术讨论会”，这次讨论会由北京语言大学、南京师范大学、福建师范大学联合主办。

本次会议将秉承前两届会议的宗旨，交流海内外汉语中介语语料库建设与应用研究的新成果，探讨相关问题，促进学界的交流与合作，进一步推动汉语中介语语料库的建设与基于语料库的汉语教学研究的发展，继续推动汉语中介语语料库的建设与相关研究。

我是一个普通的语言学研究者，一直关注汉语中介语语料库建设与应用问题，曾经参加了第一届和第二届的会议，今天有机会来到福建师范大学参加第三届汉语中介语语料库建设与应用国际学术讨论会，我对于这次会议的召开，表示热烈的祝贺。

目前，汉语中介语语料库建设已经取得了很大的成绩。早在1995年，北京语言大学的前身北京语言学院就建立了第一个汉语中介语语料库。后来，北京语言大学又建立了“HSK动态作文语料库”，放在网上供大家免费使用，受到普遍的欢迎。最近，崔希亮教授又主持了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”，在全球范围内来研究汉语中介语语料库的问题。南京师范大学、鲁东大学、暨南

大学、中山大学、上海交通大学、厦门大学、台湾师范大学也先后建立了不同规模的汉语中介语语料库。香港中文大学、新疆医科大学建立了汉语口语习得语料库。

在这些中介语语料库的基础上,研究者们进行了很多卓有成效的研究。特别值得注意的是,其中有一些研究使用了计量的方法。

例如,南京师范大学肖奚强、黄自然采用统计分析的方法,探讨了中介语中句子定语复杂度的发展情况。他们的研究发现,各级中介语定语的复杂度,随句长的增加呈现出逐步上升的趋势,初级、中级、高级的中介语,定语的复杂度呈逐级上升。他们还发现,在中介语中,多处定语的比例高于目标语,而多层定语的比例低于目标语。因此,他们推断:多处定语与多层定语之间可能存在着此消彼长的关系。也就是说,在中介语中,多处定语的出现数量与多层定语的出现数量之间是相互影响的,这种相互影响,也就是语言中各种特征之间的协同关系(synergetic relation)。

北京语言大学黄伟用统计方法研究了母语背景因素对中高级汉语学习者书写汉字的影响,发现母语背景因素对于独体字的书写没有显著的差异,而对于合体字的书写存在显著的差异。影响独体字书写的主要因素是字形特征(笔画),而影响合体字书写的主要因素除了字形特征(部件、结构类型)之外,还有母语背景,日本学生、韩国学生和欧美学生的汉字书写错误,呈现出递增的趋势。黄伟的研究很有价值,笔画对于独体字书写的影响,部件、结构类型对于合体字书写的影响,也是一种协同关系。其实,独体字书写对于合体字书写也是有影响的,合体字书写对于独体字书写也是有影响的,这也是一种协同关系,可惜黄伟在他的研究中,还没有涉及这样的协同关系。

协同语言学(synergetic linguistics)是系统科学和语言科学相结合的交叉学科。其理论核心是“语言是一个自组织和自调节的系统,是一个特殊的动态系统”^①。在协同语言学的理论框架下,语言既被看作是一种心理-社会现象,同时又是一种生物-认知现象。和其他自组织系统一样,构成语言系统的大量子系统之间存在一系列相互竞争、相互协作、相互依存的过程。这些运作过程融合了语言系统要求(system requirements),来自生物、心理、物理、社会系统以及其他方面的外部

① Köhler, R. Synergetic linguistics. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.) *Quantitative Linguistics. An International Handbook*. Berlin: Walter de Gruyter. 760-774, 2005.

力量一起产生作用,构成了语言体系的运作机制,这些性质不同的语言系统要求只有通过相应的语言功能之间的相互影响、相互作用才能得到满足,“而语言功能间的这些相互影响关系的具体形式也就是语言运作的规律”^①。因此,语言的运作过程并非杂乱无章而是有规律可循的。

协同语言学主张采用演绎的方式针对语言运作机制,特别是语言交际活动中呈现的各种语言现象、语言结构、结构的属性以及它们之间的相互协作关系提出具有普遍性的理论假设,然后运用概率论、随机过程、微分方程、函数论等统计的、非离散的数学定量方法,对其进行精确的度量、观察、建模和解释。在此基础上通过验证理论假设,揭示语言系统的自适应机制以及语言演化的动因,即语言系统从杂乱无章变成井然有序以及有序结构形成之后语言系统发展变化的规律,并倡导将这些规律以数学定律的形式呈现出来,从而形成表述严密的语言理论体系。^②

协同语言学是计量语言学(quantitative linguistics)的一个分支学科。

1964年,英国统计学家赫丹(Gustav Herdan)首次使用了“计量语言学”这个术语。他出版了《计量语言学》(*Quantitative Linguistics*, 1964)、《语言作为选择和机会的理论》(*Advanced Theory of Language as Choice and Chance*, 1966)等著作,产生了广泛的影响。

当代计量语言学的主要代表人物都是来自德国和东欧地区的,其中最著名的是德国波鸿大学(Bochum University)的阿尔特曼(Gabriel Altmann)教授和德国特里尔大学(Trier University)的柯勒(Reinhard Köhler)教授。目前有两本计量语言学的国际学术刊物“*Journal of Quantitative Linguistics*”和“*Glottometrics*”,前者为国际计量语言学协会(International Quantitative Linguistics Association, IQLA, <http://www.iqla.org/>)的会刊,是SSCI的国际检索刊物。IQLA的网站上有一些非常有用的关于计量语言学的网上资源。国际著名的语言学出版社Mouton de Gruyter正在出版一个计量语言学的图书系列,已出60多种。

计量语言学以真实的语言交际活动中呈现的各种语言现象、语言结构、结构属性以及它们之间的关系作为研究对象,使用概率论、随机过程、微分方程、函数论等统计的、非离散的数学定量方法,对其进行精

① Altmann, G. Zur Funktionalanalyse in der Linguistik. In J. Esser & A. Hübner (eds.) *Forms and Functions*. Tübingen: Gunter Narr Verlag. 25-32, 1981.

② Naumann, S., P. Grzybek, R. Vulcanovi & G. Altmann. Preface. In Naumann, S., P. Grzybek, R. Vulcanovi & G. Altmann (eds.) *Synergetic Linguistics: Text and language as dynamic systems*. Vienna: Praesens Verlag, 2012.

确的度量、观察、模拟、建模和解释,以探索语言现象中隐藏的数学规律,刻画语言的数学面貌,揭示语言系统的自适应机制以及语言演化的动因,从而发现自然语言中存在的各种定律,如分布定律、函数定律、演化定律等。因此,计量语言学是以真实语料为基础、用计量的方法来研究语言的结构和发展规律的一门语言学科。^①

协同语言学是计量语言学的奠基人之一柯勒教授在20世纪80年代提出的,目前已经取得众多的成果。在我们研究中介语语料库的时候,如果能够应用协同语言学的理论和方法来研究中介语中各个特征的协同关系,将有助于揭示隐藏在中介语的语言表层现象之下的内部规律,从而提高我们对于中介语规律性的认识。

计量语言学和协同语言学都使用实证的方法来研究语言,这种实证的方法来源于哲学中的经验主义(empiricism)。

许嘉璐先生主张我们语言学工作者不要只埋头于具体的研究,而应当抬起头来仰望一下天空。^②这个“天空”就是哲学。他大声疾呼,“必须进行形而上老者与形而下智者的对话”。

这是振聋发聩的呼吁!这是高瞻远瞩的主张!我非常赞同。

我们这些埋头于具体研究的人,都是些“形而下智者”,我们确实应当仰望一下哲学这片无比广阔的天空,与“形而上老者”的哲学研究者进行对话,用哲学来指导我们的研究工作。

哲学中的经验主义以培根(Francis Bacon, 1561—1626)、霍布斯(Thomas Hobbes, 1588—1679)、洛克(John Locke, 1632—1704)、休谟(David Hume, 1711—1776)为代表。他们都是英国哲学家,因此,经验主义也被称为“英国经验主义”。

培根批评理性派哲学家。他说,理性派哲学家只是从经验中抓到一些既没有适当审定,也没有经过仔细考察和衡量的普遍例证,而把其余的事情都交给了玄想和个人的机智活动。培根提出“三表法”,制定了经验归纳法,建立了归纳逻辑体系,对于经验自然科学起了理论指导作用。

霍布斯认为归纳法不仅包含分析,而且也包含综合。分析得出的普遍原因只有通过综合才能成为研究对象的特殊原因。

洛克把理性演绎隶属于经验归纳之下,对演绎法做了经验主义的理解。他认为,一切知识和推论的直接对象是一些个别、特殊的事物。

① 冯志伟,用计量方法研究语言,《外语教学与研究》,2012年,第2期。

② 许嘉璐,研究中文信息处理,需要仰望一下天空,《语言文字应用》,2006年,第2期。

我们获取知识的正确途径只能是从个别、特殊进展到一般。他指出：“我们的知识是由特殊方面开始，逐渐才扩展到概括方面的。只是在后来，人心就采取了另一条相反的途径。它要尽力把它的知识形成概括的命题。”^①

休谟运用实验推理的方法来剖析人性，试图建立一个精神哲学体系。他指出：“一切关于事实的推理，似乎都建立在因果关系上面，只要依照这种关系来推理，我们便能超出我们的记忆和感觉的见证以外。”他认为，“原因和结果的发现，是不能通过理性，只能通过经验的”。^②经验是我们关于因果关系的一切推论和结论的基础。

在这些经验主义哲学家的影响之下，现代自然科学的代表人物牛顿（Isaac Newton, 1642—1727），建立了经典力学的基本定律，即牛顿三定律和万有引力定律，使经典力学的科学体系臻于完善。他的哲学思想也带有明显的经验主义倾向。他认为自然哲学只能从经验事实出发去解释世界事物，因而经验归纳法是最好的论证方法。他说，“虽然用归纳法来从实验和观察中进行论证不能算是普遍的结论，但它是事物本性所许可的最好的论证方法，并随着归纳的愈为普遍，这种论证看来也愈有力”。他把经验归纳作为科学研究的一般方法论原理。他认为，“实验科学只能从现象出发，并且只能用归纳来从这些现象中推演出一般的命题”。^③正是由于牛顿遵循经验主义的归纳法，才在物理学上取得了划时代的伟大成就。

法国启蒙运动的代表人物伏尔泰（Voltaire, 1694—1778）也有明显的经验主义倾向。他以洛克的经验主义为武器去反对教会至上的权威，否定神的启示和奇迹，否认灵魂不死。他赞美经验主义哲学家洛克，“也许从来没有一个人比洛克头脑更明智，更有条理，在逻辑上更为严谨”。^④他积极地把英国经验主义推行到法国，推动了法国的启蒙运动。

计量语言学和协同语言学都使用实证的方法来研究语言，这种实证的方法来源于哲学中的经验主义。所以，我们应当关注哲学中的经验主义，运用哲学思想来提升我们的研究水平。

不过，在关注哲学中的经验主义的同时，我们也不能忽视哲学中的

① 洛克，《人类理解论》，商务印书馆，1998年，第598页。

② 休谟，《人类理解研究》，商务印书馆，1957年，第27页。

③ 塞耶编，《牛顿自然哲学著作选》，上海人民出版社，1974年，第8页。

④ 葛力，《十八世纪法国哲学》，商务印书馆，1991年，第59页。