



第四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

Dì-sì Jiè Hànyǔ Zhōngjiěyǔ Yǔliàokǔ Jiànshè yǔ Yīngyòng Guójì Xuésù Tàolūnhuì Lúnwén Xuānjí

张亚军 肖奚强 张宝林 林新年 主编

第四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

张亚军 肖奚强 张宝林 林新年 主编

世界图书出版公司

北京·广州·上海·西安

图书在版编目 (CIP) 数据

第四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 / 张亚军等编. —北京: 世界图书出版有限公司北京分公司, 2018.11

ISBN 978-7-5192-5033-1

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

中国版本图书馆 CIP 数据核字 (2018) 第 188656 号

书 名 第四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-SI JIE HANYU ZHONGJIEYU YULIAOKU JIANSHE
YU YINGYONG GUOJI XUESHU TAOLUNHUI LUNWEN XUANJI

主 编 张亚军 肖奚强 张宝林 林新年
责任编辑 梁沁宁

出版发行 世界图书出版有限公司北京分公司
地 址 北京市东城区朝内大街 137 号
邮 编 100010
电 话 010-64038355 (发行) 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 27.75
字 数 542 千字
版 次 2018 年 11 月第 1 版
印 次 2018 年 11 月第 1 次印刷
国际书号 ISBN 978-7-5192-5033-1
定 价 82.00 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：张亚军 肖奚强 张宝林 林新年

编委：（以姓氏拼音为序）

陈 莉 陈 晟 崔希亮

梁丹丹 林新年 肖奚强

颜 明 张宝林 张亚军

目 录

第四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

..... 张亚军 陈 莉 颜 明 (001)

大会报告

迈开汉语拼音走向世界的新步伐

..... 冯志伟 (017)

国际汉语教材语料库的建设与应用

..... 周小兵 薄 巍 王 乐 李亚楠 (024)

英语学习者语料库研究的六张面孔

..... 许家金 (039)

论汉语中介语语料库建设的基本问题

..... 颜 明 肖奚强 (048)

言语管理及其意义

..... 刘运同 (061)

关于汉语中介语语料库标注规范研究的新思考

——兼谈“全球汉语中介语语料库”标注规范的设计

..... 张宝林 崔希亮 (069)

中介语语料库建设研究

- 从离合词的偏误调查看“HSK动态作文语料库”的补充及其教学
对策
..... 王鸿滨 (085)
- 中日中介语语料库语料采集对比研究
..... 何 佳 (100)
- 多模态语料库建设研究
..... 邢晓青 (108)
- 关于汉语中介语语料库语料标注代码的思考
..... 赵焕政 林君峰 (115)
- 汉语中介语词语误用研究及对偏误标注建设的建议
..... 杨 帆 (132)

基于中介语语料库的汉语语法习得研究

- 韩国学生汉语中介语各句长句子补语复杂度发展研究
..... 黄自然 (157)
- 韩国学生汉语“自己”宾位照应习得研究
..... 肖奚强 宁倩倩 (167)
- 韩国留学生汉语中介语补语系统实证研究
..... 徐开妍 (181)
- 外国留学生双项并列宾语习得研究
..... 程仕仪 肖奚强 (191)
- 英语背景的美国学生汉语语序偏误分析
——基于“HSK动态作文语料库”的研究
..... 张 硕 (209)

基于中介语语料库的汉语字词习得研究

音形义学习顺序对留学生汉字习得的影响

..... 杨婉晴 梁丹丹 (219)

基于中介语语料库的留学生汉字偏误研究

..... 祁 峰 张慕昭 (231)

汉语成语教学的数量、维度及相关问题

..... 郭圣林 (245)

汉语二语学习者多义词“放”深度习得研究

..... 陈 明 宋春阳 (256)

越南学习者汉语“了”的习得研究

..... 彭 臻 唐文成 (269)

基于语料库的汉语语素义与词汇习得研究

——以“只见”“反而”为例

..... 赵春秋 (286)

暨南大学书面语语料库中“挑选”与“选择”的使用考察

..... 郭曙纶 (300)

基于汉语中介语语料库学习者“帮”“帮助”“帮忙”混淆情况的
考察

..... 乔 侠 (311)

基于语料库“突然”“忽然”“猛然”“猝然”的状语差异性特征
研究

..... 林琳琳 (321)

外国留学生交互义同义副词习得研究

..... 颜 明 (332)

基于中介语语料库的语音语义及其他相关研究

基于口语语料库的外国留学生语义偏误类型考察

..... 王永星 (351)

理解性练习、产出性练习与致使重动句习得

..... 李贤卓 (362)

注疏文献中原文与引文的句子对齐研究

..... 梁社会 (375)

我国汉语中介语语料库研究现状与热点透视

——基于 CiteSpace 的可视化分析

..... 蔡 武 郑通涛 (389)

新世纪以来汉语作为第二语言习得研究成果分析

——基于四份 CSSCI 中国语言学来源期刊文献的统计

..... 毕 晋 肖奚强 程仕仪 (403)

搭配语义韵视角下的翻译共性研究

..... 冯浩达 (420)

总结报告

汉语中介语语料库建设与应用的深入持续发展

——第四届汉语中介语语料库建设与应用学术研讨会总结报告

..... 肖奚强 (431)

第四届汉语中介语语料库建设与应用国际学术研讨会参会论文目录

..... (434)

第四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

张亚军 陈 莉 颜 明
(扬州大学文学院)

1. 引言

“第四届汉语中介语语料库建设与应用国际学术讨论会”由北京语言大学、南京师范大学、福建师范大学、扬州大学联合主办，扬州大学文学院承办，于2016年11月11日至14日在江苏扬州召开。会议受到学界广泛的关注与支持，共一百余位专家学者、研究生参会。会议共收到投稿70余篇，实际宣读论文56篇（含大会报告6篇）。论文集收录其中32篇论文全文及大会总结。分组汇报的论文共分四个议题：汉语中介语语料库建设研究、基于中介语语料库的汉语语法习得研究、基于中介语语料库的汉语字词习得研究、基于中介语语料库的语音语义及其他相关研究。下面对所收文章按大会报告、分组专题讨论分别加以综述^①。

2. 大会报告

大会报告分为两场，每场三位教授发言。

教育部语言文字应用研究所冯志伟教授做题为《迈向汉语拼音走向世界的新步伐》的学术报告，介绍了国际标准ISO 7098的修订过程，指出该国际标准修订的两大特点：一是把汉语拼音按词连写的规则引入国际标准中，二是把汉字—拼音转写的自动译音方法引入国际标准中。

中山大学周小兵教授报告了与团队合作的《国际汉语教材语料库的

① 部分论文题目较参会时有所改动，综述和论文集目录以改稿为准。

建设与应用》，讨论国际汉语教材语料库的建设和应用。结合具体研发成果，从教材标准研制、教材评估与难度测定、测评软件研制和应用等方面，考察语料库对汉语教学与习得研究的促进作用；通过案例分析，展示教材语料库与其他语料库（如目标语语料库、学习者母语语料库、双语语料库、中介语语料库等）配合使用的积极效应；针对教材语料库的问题，提出解决方案，强调教材语料库在教学上的应用效果和转化效率。

北京外国语大学许家金教授报告《英语学习者语料库研究的六张面孔》，将英语学习者语料库研究按发展阶段粗略分为前后两个10年，并按研究视角细分为6个方面。报告认为，前一个10年可视为英语中介语语料库研究的经典面孔，其分析思路偏重中介语的结构特征；后一个10年可视为英语中介语语料库研究的后现代面孔，其分析思路具有社会认知和社会建构方面的特色。报告指出“大数据视野下中介语语料库的动态性和复杂性研究”将成为今后一段时期英语中介语语料库研究的发展方向。

南京师范大学肖奚强教授报告了与颜明合作的《论汉语中介语语料库建设的基本问题》，报告集中讨论了汉语中介语语料库建设中互相关联的四个基本问题：1）语料有不同种类，应该对入库语料的种类和性质有清醒的认识；2）中介语语料的自然性不是不证自明的，应该对其自然性有所认识和把控，语料的连续性各不相同，应该有收集的规划、规则和说明；3）应注意语料的全面性和代表性，大而全的全面性语料收集法面临分级对齐的巨大难题，局部的代表性有其不可替代的价值；4）中介语语料库在展开语言要素的习得研究方面确有优势，但难以直接进行认知层面的探究。语料的标注与否各有利弊，基于语料库的中介语研究应与习得者母语、目的语及教学输入语言相比勘方能更具解释力。

同济大学刘运同教授的报告《言语管理及其意义》认为，在口语研究与口语教学中有以下几个方面值得我们进一步反思和改进。首先，需要改变观念，口语中的一些现象应放在口语的框架下进行评价。其次，在进行口语评价时，要区别正常的言语管理与不正常的非流利现象，仅仅采用定量的标准可能扭曲学习者的口语能力。再次，在语料库建设中对口语语料应按照口语自身的特点进行分析、转写和标注，不要用书面语的标准来要求口语。

北京语言大学张宝林教授报告了与崔希亮教授合作的《关于汉语中介语语料库标注规范研究的新思考——兼谈“全球汉语中介语语料库”标注规范的设计》，指出语料标注在基于语料库的语言教学与研究中共

有重要意义。就通用型汉语中介语语料库的语料标注而言,应贯彻全面性、有限性、渐进性、准确性、系统性、简洁性、开放性、自动化等 8 项基本原则。在标注内容方面,应从教学与研究的实际出发,以问题为导向,进行字、词、短语、句、篇、语体、语义、语用、修辞、标点符号等 10 个层面的标注,以满足教学与研究的多方面需求。在标注方法方面,可以采取多版标注、“基于语料库”和“语料库驱动”两种研究范式理念相结合的标注、部分层面的自动标注等方法,以提高标注效率,提升语料库建设水平。

3. 分组讨论一: 汉语中介语语料库建设研究

本组共宣读 12 篇论文,其中,多模态语料库建设研究 3 篇,汉语中介语语料库建设具体问题研究共 5 篇,汉语中介语语料库建设综合性问题的探讨共 4 篇。

3.1 多模态语料库建设(主要总结多模态语料库建设的历史与成果,分析其研究和应用的现状与价值,并展望未来,针对现有问题提出解决建议)

渤海大学陈雷的《汉语中介语多模态语料库建设与研究的现状》梳理了多模态语料库、基于多模态语料库的话语研究和语料库建设研究在近二十年所取得的成果,分析了多模态语料库建设的研究现状、使用工具、存在问题及发展趋势。

南京师范大学张小峰的《第二语言多模态学习语料库建设初探》详细介绍了基于计算机的信息技术对于第二语言教学的根本意义,在此基础上提出:语言学习应有多模态知识的支撑,应从图、文、声、像多维度开发多模态学习语料库,获取语言的多模态知识,创造真实而丰富的教学环境,从而实现多媒体语料库支持的汉语学习。

北京语言大学邢晓青的《多模态语料库建设研究》将现有的多模态研究从理论上归为社会符号学角度、系统功能语言学角度、社会互动分析角度三类,并指出多模态语料库应用于第二语言习得和教学研究的价值。该文提出目前多模态研究存在的一些局限性,并提出尝试性解决方法。

3.2 汉语中介语语料库建设具体问题研究（主要讨论语料收集、语料标注及区域语料库建设问题）

建设中介语语料库的第一步是收集语料，扬州大学陈莉的《汉语中介语语料库语料收集问题再思考》提出：中介语语料收集要为目的服务，语料收集要保证同质性，大数据的语料收集要考虑学习者的学习方式、伦理及海量信息的筛选问题。南通大学何佳的《中日中介语语料库语料采集对比研究》则通过中日两国三种中介语语料库语料采集的对比，提出完善中介语语料采集的八点建议。包括：丰富和细化语料的产出者类别，同一学习者的口语与书面语语料共同采集，汉语作文与母语翻译一同收录，收录汉语母语作文，增加纵向语料的收集等。

语料标注是语料库建设中最重要、最复杂的问题之一，本次会议如下3篇论文对此进行了讨论。西北师范大学赵焕改、福建师范大学林君峰合作的《关于汉语中介语语料库语料标注代码的思考》指出：目前汉语中介语语料库标注代码设计没有统一标准，不同语料库的分词及词性标注代码和偏误标注代码不一致等问题造成了语料共享困难及标注工具的重复开发。中国传媒大学高玮的《中介语语料库的语篇偏误标注探讨》以30余万字留学生期末作文试卷为原始语料，对语料中出现的各类篇章偏误类型进行了细分；对语料进行了元数据标注和篇章偏误项目标注，设计了一个篇章偏误标注框架，针对篇章偏误标注的特殊性提出了“关注整体，凸现偏误”的标注原则。北京语言大学杨帆的《汉语中介语词语误用研究及对偏误标注建设的建议》对“首都留学生汉语文本语料库”与“HSK动态作文语料库”的“词语误用”进行考察，归纳出十一个类型并总结出“词语误用”的分布特征。在此基础上，该文还提出对汉语中介语语料库偏误标注建设的两点建议：区分词语标注的三个层面，重新整合并细化汉语中介语语料库词语偏误标注层级体系。

3.3 汉语中介语语料库建设综合性问题的探讨（涵盖语料库建设理念与实践的考察、专门语料库建设、中介语语料库如何更好地为研究服务等内容）

北京语言大学靳继君的《汉语语料库的建设考察及对建设汉语中介语语料库的建议》考察两岸三地的汉语语料库，提出开放语料库、设计易于使用的语料库界面、统一语料库的标注规范和标准、分离文本与标注符的标注方式等建议。

南开大学冉启斌、顾倩的论文《“口音汉语在线：世界典型汉语语音偏误数据库”的制作与问题》介绍了“汉语口音在线”的制作情况：收录世界范围不同母语背景说话人说汉语时的口音表现，并提供与口音汉语有关的资料、资源。数据库的目标是收录世界上主要国家/语言的汉语语音偏误数据。辽宁大学李媚乐的《汉语偏误成因的国别化研究与区域性中介语语料库的建设》一文分析总结了目前偏误成因国别化研究的概况，就区域性语料库建设问题提出了一些设想。

语料库建设的重要目的之一是为相关研究服务，针对如何更好地实现这一目的，北京语言大学王鸿宾的《从离合词的偏误调查看“HSK 动态作文语料库”的补充及其教学对策》通过对中介语语料库中离合词偏误的调查分析发现，由于学生“回避”等因素的影响，调查的结果不能代表学生在真实语境中的使用情况，为此在 HSK 动态作文语料库的偏误分析基础上，总结并补充了离合词的使用偏误类型，对不同阶段的留学生设计了相应的问卷，进行了进一步调查。最后，该文将两种调查的结果进行综合，对教学中离合词的教学提出具体的补救建议。

4. 分组讨论二：基于中介语语料库的汉语语法习得研究

本组共宣读 12 篇论文，其中，句型、句式的习得研究共 5 篇，篇章研究共 3 篇，其他语言项目的习得研究 4 篇。

4.1 句型、句式的习得研究(既有对某一特殊句型、句式的习得研究，也有对于中介语句式系统发展的研究)

共有 3 篇论文分别考察了形容词谓语句、连谓句和“把”字句的习得状况。北京语言大学何星慧的《基于语料库的留学生形容词谓语句习得情况考察——以美国学生、泰国学生为例》一文发现，不同国家的留学生习得形容词谓语句时没有国别因素的影响，但是在习得某些下位类型时母语背景是影响其习得效果的因素之一。此外，论文认为，笼统地说形容词谓语句是或者不是母语或第一语言为英语的学生的学习难点是不恰当的。西北工业大学白雯的《基于 HSK 动态作文语料库的韩国学生连谓句偏误研究》对韩国学生连谓句偏误的具体用例进行了分析，并与母语使用者的使用情况进行了对比，发现韩国学生连谓句的错误主要集中在语序方面。造成这类偏误的主要原因是母语负迁移和目的语规则的

泛化。扬州大学吴海玲的《基于“HSK 动态作文语料库”的韩国留学生“把”字句使用偏误研究》将韩国学生的“把”字句偏误分为四类,对各类偏误的出现频率及产生原因进行了分析,并提出了要重视“把”字句的语义、语用教学,加强相关知识点的对比分析,“把”字句的教学要合理适度等教学建议。

黄自然和徐开妍的两篇论文基于大规模语料库的研究,尝试证实中介语系统假说,发现了一定的特点和规律。黄山学院黄自然的《韩国学生中介语各句长句子补语复杂度发展研究》基于大规模语料库的统计和对比发现,中介语各句长句子补语复杂度的发展变化具有其内在的系统性和规律性,表现在同一等级句子句法复杂度的发展具有一致性,各级变化具有动态性,中介语补语复杂度跟本族人语料存在较大差距,其发展具有不受教学输入语言制约等特点。南京信息工程大学徐开妍的《韩国留学生汉语中介语补语系统实证研究》基于大规模语料库的调查,以韩国学生的汉语中介语句式系统为例,初步证实了汉语中介语的各种特性,如中介语与本族语相比既有共性也有差异。文章的调查结论同时也肯定了中介语对比分析法的科学性。

4.2 篇章研究(均来自南京师范大学的研究团队,主要考察篇章连接成分和篇章照应的习得研究)

陈文君的《基于中介语语料库的韩国学生篇章连接成分“因为、所以”习得研究》借助自建的100万字本族人书面语语料库和100万字分为初、中、高三级的韩国留学生中介语作文语料,对“因为、所以”在使用过程中出现的偏误及其成因进行了分析和解释。刘珠的《韩国留学生连续时间连接成分“后来”的习得研究》根据统计数据认为,韩国学生汉语篇章中后时时间连接成分“后来”的所有使用情况、使用频率都低于本族语者,使用情况还比较单一。但从初级阶段到高级阶段,韩国学生汉语篇章中“后来”的各种使用情况正确用例和正确率是在逐渐增加的。肖奚强、宁倩倩的《韩国学生对汉语“自己”宾位照应习得研究》对韩国学生汉语中介语中处于宾语位置的“自己”进行了句法分布及使用状况的调查研究,并针对韩国学生的习得情况对“自己”的对外汉语教学提出了比较具体的建议。

4.3 其他语言项目的习得研究（包括具体词语、语法成分、语序的习得研究）

南京师范大学郑善美的《基于中介语语料库的韩国学生现代汉语“就”的使用现状及原因探讨》基于韩国留学生的中介语语料与韩汉双向平行语料，将汉语副词“就”与韩语相应词的使用特点进行了对比分析，总结了韩国学生跟外国留学生使用现代汉语副词“就”时的差异，并结合韩语对此原因进行了深入的分析。同样来自南师大的程仕仪、肖奚强的《外国留学生双项并列宾语习得研究》基于本族人并列宾语的排序原则，对留学生并列宾语的排序情况进行了考察。论文发现本族人使用的三大类十条原则在中介语语料中均有出现，而且在留学生语料中还发现了由小到大和个人熟悉程度两条新的排序原则。对比发现，除了显著性原则在两种语料中都强势使用外，本族人和留学生在倾向性上存在明显差异。

扬州大学张硕的《英语背景的美国学生汉语语序偏误分析——基于“HSK 动态作文语料库”的研究》对汉语语序偏误进行了研究，分析了语序偏误的类型、偏误原因，在分析偏误原因的基础上提出了比较具体的教学建议。渤海大学于本凤的《转折连词“但是”和“不过”的比较分析——基于国家语委语料库的研究》运用语料库统计检索的方法，从句法分布、语法意义、语用角度三个方面进行了比较，其结论为对比教学提供了语料和数据支撑。

5. 分组讨论三：基于中介语语料库的汉语字词习得研究

本组共宣读 13 篇论文，包括汉字习得研究 2 篇，一般词语习得研究 6 篇，近义词习得研究 5 篇。

5.1 汉字习得方面的研究

南京师范大学杨婉晴、梁丹丹的《音形义学习顺序对留学生汉字习得的影响》根据对汉字音形义三方面的排列顺序不同拟出六种汉字学习顺序，尝试寻找一种最佳的汉字学习顺序来促进对外汉语教学实践。结果显示：汉字学习效果与汉字学习顺序无关，但与学习者母语背景和汉字本身的结构相关。“汉字文化圈”背景对汉字学习有促进作用，主要体现在形义结合上。左右或左中右结构易于加快汉字的形音联系，部件多或复杂的结构会减弱汉字的形义联系。

华东师范大学祁峰、张慕昭的《基于中介语语料库的留学生汉字偏误研究》基于华东师范大学留学生汉语中介语语料库,对汉字偏误现象进行定量统计与分析。结果表明,错字与别字的数量大致相同,分别占四成左右。所有汉字偏误中,字形类偏误最为常见,其中又以部件偏误居多。在其他类汉字偏误中,日本汉字导致的偏误占一半以上。该文分析了偏误原因并给出相应的教学建议。

5.2 近义词习得研究

辽宁大学赵春秋的《基于语料库的汉语语素义与词汇习得研究——以“只见”和“反而”为例》,分析中级汉语词汇的构词语素义对词汇习得的影响。通过分析发现,由于整词语义及用法不等于语素义的简单加合,导致学习者类推策略的使用干扰了他们对这类词汇的理解和产出。

上海交通大学郭曙纶的《暨南大学书面语语料库中“挑选”与“选择”的使用考察》基于暨大语料库,具体考察“选择”与“挑选”这两个动词的使用情况。分析表明:1)“选择”比“挑选”常用得多;2)“选择”比“挑选”掌握更好;3)“选择”主要充当谓语,也能充当主语、宾语和定语,甚至能用于构词。

河南中医药大学乔倓的《基于汉语中介语语料库学习者“帮”“帮助”“帮忙”混淆情况的考察》基于本族语者语料,从句法功能、所带宾语的性质、在相同句法结构中的表义倾向性等方面,对这三个词的用法进行了比较分析。该文利用留学生中介语语料,对留学生的偏误情况进行考察,并提出教学建议。

福建师范大学林琳琳的《基于语料库“突然”“忽然”“猛然”“猝然”的状语差异性特征研究》立足于语料库,运用中介语对比分析法(CIA)、显性搭配统计测量法(MI值测量法),从词频、显著搭配、语义韵三个方面对“突然”“忽然”“猛然”“猝然”四个易混淆的词进行调查、对比、分析,揭示它们充当状语时的差异性特征。

扬州大学颜明的《外国留学生交互义同义副词习得研究》考察得出,外国留学生对交互义同义副词的使用情况与教材和大纲的编排不无关系;“互相”的高频使用不仅与单双音节有关,而且和学习与掌握的先后次序有关。这组同义词的偏误主要是误代,表现为“互”、“相”与“互相”之间的单双音节的互代,“互相”误代“彼此”,以及与“互相”共现

的动词之间的误代。从历时发展来看,偏误率与汉语水平的提高成正比,且近 66% 的偏误是误代类偏误,说明单双音节的语内规则的相互迁移比语际迁移更难克服。

5.3 其他词语习得研究

南京师范大学郭圣林的《汉语成语教学的数量、维度及相关问题》认为,现有部分汉语教学大纲初、中、高三个阶段成语数量安排得不够合理,对于成语应该当作词汇项目还是语法项目也存在分歧。从“HSK 动态作文语料库”的语料来看,对于成语偏误的判定需要考虑多种可能性。

上海交通大学陈明和宋春阳的《汉语二语学习者多义词“放”深度习得研究》主要考察与探讨教学大纲对二语学习者多义词不同义项习得的效应。通过对中介语作文语料库中不同水平学习者对“放”不同义项使用率和正确率两方面的统计,发现教学大纲的义项安排并非影响其习得效应的决定性因素。文章认为,人脑普遍的认知机制激活了“放”的语义网络和学习者习得“放”的一般的认知路径;由于功能动因,实用性需求最强的义项即输入及输出频率高的义项被优先习得。

中山大学李捍华的《基于中介语语料库的留学生汉语生造词偏误分析》利用中山大学的中介语语料库,对留学生的生造词偏误进行考察发现:甲级、乙级词的生造词偏误数量多于超纲、丁级、丙级词;汉语水平越高越容易出现生造词偏误;母语为越南语、泰语的学生比其他母语背景的学生更容易出现生造词偏误。该文对生造词偏误的类型和原因进行了分析。

北京语言大学王艺璇的《基于语料库中高级汉语学习者产出性词汇发展趋势与特点研究》以“HSK 动态作文语料库”中 360 篇作文作为文本,参考借鉴了国外学者提出的有关“词汇丰富性”的研究框架,考察中高级汉语学习者产出性词汇的发展趋势及特点。

广西民族大学彭臻、唐文成的《越南学习者汉语“了”的习得研究》利用广西民族大学的越南留学生作文以及北语 HSK 动态作文语料库研究发现:“了₂”的习得情况比“了₁”好;“了₁”和“了₂”最主要的偏误类型都是遗漏;习得最好的是“Adj+了₂”“不……了₂”,其次是“V+了₂”“V+O+了₂”,再次是带“了₁”的结构。

台湾师范大学张莉萍的《初探华语学习者名词词组的习得情况》利用本族人语料库和学习者语料库,考察学习者和本族人名词词组的使用