

# 汉语中介语口语语料库建设的现状与任务<sup>1</sup> (Constructing Spoken Corpora of L2 Chinese Learners: Where We Are and What Needs to Be Done)

张宝林  
(Zhang, Baolin)  
北京语言大学  
(Beijing Language and Culture University)  
zhangbl@blcu.edu.cn

**摘要:** 与书面语语料库建设的蓬勃发展相比, 汉语中介语口语语料库的建设相对滞后, 制约着基于语料库的口语教学与习得研究的发展。目前, 急需对口语库建设进行研究, 解决对口语库的认识问题、语料收集与标注问题。本文简要讨论了汉语中介语口语语料库研究相对滞后的原因, 并介绍了汉语中介语口语语料库需要标注的内容和标注方法。希望我们的介绍能有助于今后更多的研究出现。

**Abstract:** The construction of Chinese interlanguage spoken corpus lags behind similar efforts in Chinese written corpus, which constrains the research on spoken corpus-based Chinese language acquisition and teaching. At present, there is an urgent need to study the construction of Chinese interlanguage spoken corpus, focusing on issues such as the understanding of the basics of such spoken corpus, materials collection, and corpus annotation. In this paper, we briefly discuss the reasons for less research efforts on the construction of Chinese interlanguage spoken corpora. We also introduce the content, scope, and annotation of Chinese interlanguage spoken corpus, with the hope that our introduction will help promote more research in this field in the future.

**关键词:** 汉语中介语语料库, 口语语料库, 语料收集, 语料标注

**Keywords:** Chinese interlanguage corpus, spoken corpus, corpus collection, corpus annotation

---

<sup>1</sup> 本研究得到北京市语言资源高精尖创新中心项目(编号: KYD17004)、北京市社会科学基金项目重点项目(编号: 15WYA017)、教育部哲学社会科学研究重大课题攻关项目(批准号: 12JZD018)的资助, 谨致谢忱。

## 1. 引言

基于汉语中介语<sup>1</sup>语料库的对外汉语教学研究于上世纪90年代中期发轫，20余年来得到迅速发展，取得了众多的研究成果。仅以“HSK动态作文语料库”为例，在中国知网（CNKI）中进行查询，已发表各类论文3029篇（截至2017年12月15日）。其年度发文数量见下面图1。

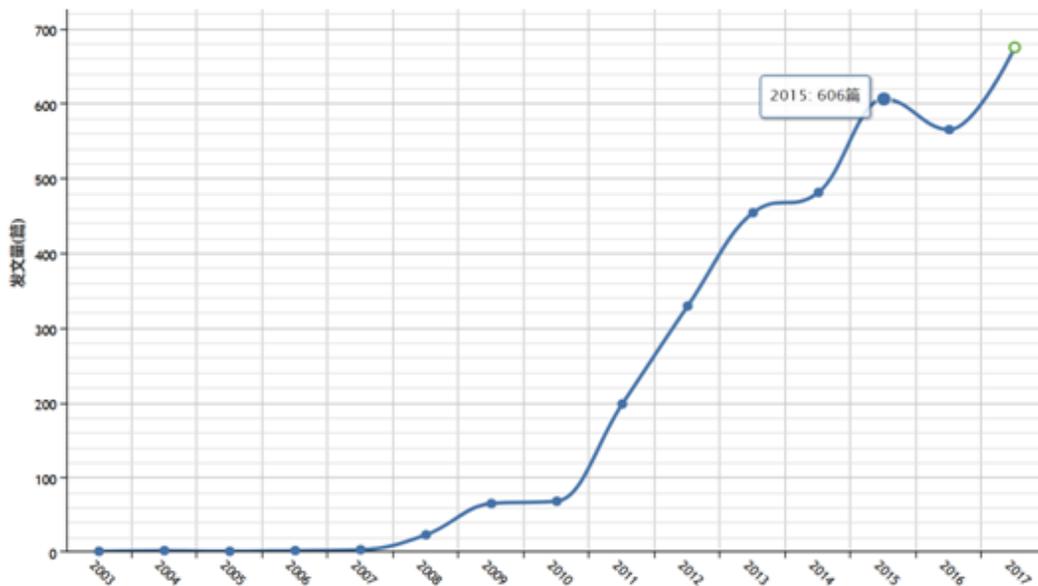


图1. 年度发文数量统计分析图（单位：篇）

图1显示，年度发文量增长的重大拐点出现于2011年，当年发文量达到198篇。此后，每年产出的论文达数百篇，从329篇、454篇、481篇、606篇、565篇，直到2017年的675篇，除2016年略有下降，年度发文量均创新高。而基于语料库的研究带来的另一变化，是促进了汉语习得研究范式的转变，即从小规模、经验型、思辨性研究逐渐发展为基于大规模真实语料的、定量分析与定性分析相结合的实证性研究，使研究结论具有了较强的客观性、稳定性和普遍意义。

由此可见，汉语中介语语料库在对外汉语教学中确实发挥了重大作用。而基于语料库的汉语偏误分析与习得研究的不断发展，又极大地促进了汉语中介语语料库的建设：从上世纪90年代北京语言大学“汉语中介语语料库系统”的一枝独秀，到本世纪前10年北京语言大学“HSK动态作文语料库”、中山大学“留学生中介语语料库”、和南京师范大学“外国学生汉语中介语偏误信息语料库”等多个语料库并存，直至今今天各种中介语语料库的蓬勃发展<sup>2</sup>，十分鲜明地勾勒出汉语中介语语料库的发展轨迹。

<sup>1</sup> 中介语指二语学习者产出的目的语，汉语中介语语料库所收集的是把汉语作为第二语言的学习者产出的汉语语料。

<sup>2</sup> 关于汉语中介语语料库建设的目前进展情况，可参见林新年、肖奚强和张宝林(2016)。

然而,这种语料库建设与基于语料库的研究蓬勃发展的繁荣局面仅限于书面语料库,口语语料库则少之又少,仅有“汉语学习者口语语料库”<sup>1</sup>、留学生中介语语料库之口语语料库<sup>2</sup>、小型外国学生口语中介语语料库<sup>3</sup>。其中开放的口语库曾有暨南大学华文学院的语料库。

与口语库建设缺乏的现实情况相对应,基于口语库的汉语口语中介语研究同样很少;对口语语料库建设的研究也不多。我们看到的仅有杨翼、李绍林、郭颖雯和田清源(2006)、陆庆和和陶家骏(2011)、刘运同(2013,2016)、刘喆(2013)、李航(2013)、周宝芯(2013)、张蕾(2016)、方淑华、陈庆华、杨惠媚和陈浩然(2016)等。

杨翼等(2006)认为,在对外汉语教学界,有关口语习得和口语测试的研究,在数量和质量上,都处于令人遗憾的状态。大多数研究方法依然是经验型的,仅靠例举少量零散的口耳直觉语料或一次性小规模抽样,得到的是不够系统的、缺乏说服力的定性结果。因此,建立具有一定规模的汉语学习者口语语料库就显得十分必要和迫切。

我们认为,这种评价是非常中肯的,从今天口语库的建设与应用情况看,仍有其现实意义。口语语料库建设已经成为汉语中介语语料库建设的瓶颈,严重制约着语料库建设水平的提高,并影响到对外汉语教学效率与水平的提升。口语库建设在汉语中介语语料库建设中有着独到的价值,在对外汉语教学研究中具有特殊的作用。因此,值得充分注意,深入研究。

## 2. 口语语料库的作用与价值

### 2.1 解决口笔语语料库建设的“倒挂”问题,为口语习得研究奠定坚实基础

在人们的日常交往中,口语<sup>4</sup>始终占据主导地位,是使用最方便、使用频率最高、使用范围最广、因而最重要的交流方式。而书面语<sup>5</sup>虽然可以打破时空限制,却仍然是使用受限、使用频率相对较低、使用范围相对较窄、因而较为次要的交流方式。在目的语环境中学习汉语的外国学习者,不论是在汉语学习中还是在日常生活中,不论是在汉语学习的课堂上还是下课之后,其使用汉语口语的几率同样远远高于书面语。然而,不论是以往传统的对外汉语教学研究,还是上世纪80年代中期以来的汉语中介语研究,都是重书面语,轻口语,并直接影响到汉语中介语语料库

<sup>1</sup> 见杨翼等,2006。

<sup>2</sup> 暨南大学华文学院建设的语料库曾在 <http://www.globalhuayu.com/corpus3/Search.aspx> 开放。目前该网址已失效。

<sup>3</sup> 见陆庆和和陶家骏,2011。

<sup>4</sup> 这里所谓口语仅指自主口头交流形式,包括独白和对话。所谓自主,指自发产生的言语材料,即不包括对书写材料的朗读。

<sup>5</sup> 这里所谓书面语仅指自主书写出来的言语材料,包括作文、日记、邮件、微博、便条等。所谓自主,指自发产生的言语材料,而不包括抄写的材料。

的建设,书面语库相对较多,口语库则极为少见。在口笔语的实际使用和语料库建设与相关的应用研究之间,存在着十分严重的“倒挂”现象。这是非常不正常的,也是十分不应该的。这种“倒挂”意味着以往学界的众多研究所关注的只是学习者使用目的的次要部分,即书面语;而忽略了其主要部分,亦即口语。倘能加强口语语料库的相关研究,推进其建设,则将为口语教学与口语习得研究奠定坚实基础,有助于解决口笔语语料库建设与应用研究的倒挂问题。

## 2.2 为口语习得研究及其与书面语习得的对比研究奠定基础,进而提升汉语中介语研究水平

在对汉语中介语口语语料库进行深入研究的基础上,建设大规模、高质量的口语语料库,首先可以为汉语口语的偏误分析与习得研究奠定坚实基础,通过基于口语语料库的中介语研究,加深我们对汉语口语中介语的认识,更好地把握其规律,为口语教学提供有益的参考。

另一方面,还可以将这种研究与已经取得众多研究成果的基于书面语语料库的相关研究相结合,进行对比分析,研究口语和书面语习得的相同点与不同点,认识二者在习得过程中的相互作用、相互影响与相互关系,全面掌握作为第二语言的汉语习得特征与规律,进而把汉语中介语的研究水平和对外汉语教学与研究的水平提升到新的高度。

## 2.3 为口语教学、教材编写等提供语料支持

口语语料库中的大规模真实语料,将提供语音、词汇、短语、句子、语篇、语体、语义、语用、修辞等各语言层面的各类偏误与正确的语言表现,经研究可以揭示汉语中介语的全貌与典型的偏误类型。口语中介语的这些宝贵信息,将为口语教学和口语教材的编写提供典型的例句、句群与篇章,从而使口语教学与教材更有针对性,促进口语教学水平与教材编写水平的提高。以把字句为例,研究表明参加高等汉语水平考试写作考试的学习者把字句的偏误率并不高,虽然在一定程度上存在回避问题,但同时也存在与之相反的泛化问题,且与回避现象相差不到一个百分点。(张宝林,2010)那么,在口语中把字句的习得情况如何?应予考察。假如同样存在类似现象,那么在教学以及教材编写中再过分强调把字句难学及其回避问题,就纯属误导,很有可能会加重其泛化问题。

## 2.4 为口语测试研究提供参考

基于口语语料库的教学与习得研究,其成果将为口语测试提供重要参考与借鉴。例如研究表明,对于参加高等汉语水平考试写作考试的学习者来说,离合词的偏误率极低,甚至达到了可以忽略不计的程度。(张宝林,2011)假如学习者口语中也呈现出同样现象,也就没必要把离合词作为测试的重点,甚至无需将其作为语法点进行测试。

## 2.5 为语料库建设探索新经验

相对于书面语语料库的建设而言,汉语中介语口语库建设是一个全新的挑战,建设实践很少,相关的研究也很少,成熟的建设经验几乎为零。汉语中介语口语语料库究竟应该如何建设?例如在选取语料方面,是只选取独白语料,还是也选取二人对话语料,乃至三人以上的多人对话语料?多人对话语料如何辨识说话人的身份?如果无法分辨说话人与话语之间的对应关系,这种语料还有无价值?能否补救?采用什么方法补救?在标注方面,是只做语音标注,还是同时进行词汇、语法、语体、语义、语用等多层面语言现象的标注?是只对小规模语料做精细标注,还是对大规模语料做浅层标注?这些问题都需要进行深入研究并得出明确、可靠、可行的结论,并付诸建库实践,逐渐积累经验,掌握相关规律,推动口语语料库建设。

## 2.6 支持教学资源库的建设

除为课堂教学提供语料支持,以及用于数据驱动学习和翻转课堂的教学之外,语料库直接应用于教学的几率并不多。语料库为教学服务的主要方式是间接的,一是通过偏误分析与习得研究为老师们的教学工作提供参考;二是通过二次开发,在语料库基础上建设各种教学资源库,例如不同母语背景的学习者使用汉语的口语词和书面语词对照表、口语句式 and 书面语句式对照表、所使用到的语法点、项的语法项目表、口语话题表,研发词汇、语法(含词法、句法、篇章语法)、语音(含声、韵、调、停顿、重音、轻声、儿化)、辞格、标点符号、语体、语义、语用的偏误信息库(包括偏误类型与数据、习得顺序、偏误原因、教学建议等)。这些教学资源库均需在建库基础上进行建设,建成后可以为教师备课提供极大方便,提高教学的针对性和教学效率,对汉语教学具有重大意义(张宝林, 2016a)。

## 3. 口语语料库建设滞后的原因与影响

### 3.1 原因

如上文所述,口语库在对外汉语教学与汉语中介语研究中具有十分重要的意义与价值。然而,其建设情况却远远落后于书面语库,基于口语库的相关研究也同样落后于基于书面语库的相关研究,其原因究竟何在呢?

我们认为,主要有如下几点:

#### (1) 语料获取难

##### a. 语料内容过于单一

从语料内容的角度看,目前收集到的语料中,与教学、测试相关的语料较多,较易收集。前者如教学实况的录音、录像,后者如高等汉语水平考试(HSK高等)

中的独白、C.TEST<sup>1</sup>考试中考生与主考官的对话，成绩考试中的独白与师生对话等。而学习者的自然交谈语料很少，谈论学习之外内容的语料很少。这固然与在校生以学习为主的校园生活内容有一定关系，但如果安排得当、引导得法，还是可以收集到学习者更多内容的口语语料的。

不同国家和地区汉语学习者语料不均是老问题，在书面语库建设中已经普遍存在，并已受到学界较多关注，例如任海波（2010）、施春宏和张瑞朋（2013）等。在口语库建设中，此问题更加凸显。整体上看，仍然是东亚、东南亚国家学习者语料多，欧美国家学习者语料少，非洲、南美国家学习者语料更少。新问题是某些国家学习者样本数量太少，在这样的基础上，很难形成具有普遍意义的结论。例如欧洲一所大学可以通过网上辅导和语伴的形式提供其口语语料乃至视频语料，但只有寥寥数人，难以形成大规模真实语料，研究结论的普遍性依然难以保证。

#### b. 语料质量欠佳

所收录的部分口语语料声音不清晰，音量太小；多人谈话分不清说话人身份，难以分辨说话人和话语之间的对应关系。这些问题一方面是语料收集者对摄录设备操作不当或缺乏对语料用途的正确认识所造成的，另一方面也需要其他技术手段的支持。例如多人对话只是录音就不够了，还需要录像，收集视频语料。这也是建设多模态语料库的必要性之一。当然，视频语料除学习者的口语表达之外，还能提供表情与肢体动作等非言语信息，这对说话者话语内容的表达具有重要的辅助作用，这也是非常有研究价值的。

#### （2）语料转写难

为了对语料进行标注与检索，口语语料需要进行转写，这是口语语料与书面语料的最大不同点。其转写准确率低，转写后的语料校对也不容易，而且费用昂贵。母语语料库建设即已因此受到制约，汉语中介语语料由于语音偏误的影响，其转写难度更为突出。

#### （3）标注规范缺乏研究

口语语料无需汉字标注，这是其简便之处；但增加了语音标注，而语音标注究竟应该标什么？怎么标？并无定论，都是有待研究的问题。例如正常停顿与非正常停顿的时间长度标准如何确定，即尚需研究。这是其繁难之处。此外，词语、句子、语体、话语、语义、语用等层面是否应该标注，观点不一，也需要进行研究。

#### （4）对口语库重要性的认识不足

<sup>1</sup> C.TEST 的全称是“实用汉语水平认定考试(Test of Practical Chinese)”，它是用来测试母语非汉语的外籍人士在国际环境下社会生活以及日常工作中实际运用汉语的能力，它的主要目的是考查应试者在商务、贸易、文化、教育等国际交流环境中使用汉语的熟练程度，对应试者的汉语实际应用能力给予权威的认定。可参见：<http://www.c-test.org.cn/>。

从语料库建设与应用研究的整体表现上看,目前学界止步于书面语库的建设与应用研究,对口语库的建设认识不足,重视不够,建设积极性不高。从更深的层次考虑,如果学界满足于书面语库的建设与应用研究的已有成果,因而故步自封,止步不前,那么必将使口语库的建设与研究减缓与停滞,难以推动相关研究的进一步发展,导致更为严重的后果。

### 3.2 影响

口语库建设的停滞不前,对学术发展已造成或将造成如下影响:

(1) 对外国学习者的汉语口语表达状况与能力的研究缺乏足够的研究资料,难以形成具有普遍意义的研究结论,无法全面、准确地了解与把握学习者的口语习得状况;

(2) 无法对学习者的汉语口语和书面语表达进行对比分析,无法对其口笔语表达特征进行研究并得到准确的认识;

(3) 无法在前述研究的基础上,深入了解与认识学习者汉语口语习得与书面语习得之间的相互作用、影响与关系,进而形成学习者口语习得与笔语习得相互促进的教学方法、体系与模式;

(4) 影响汉语中介语研究、汉语作为第二语言的习得研究、汉语学习理论研究的发展,严重影响汉语教学水平与效率的提高。

因为这些或可能的影响,所以我们需要开展并大力推进口语库建设的相关研究,积极探索,切实推动口语语料库的建设。

## 4. 口语语料库的标注内容

### 4.1 口语语料的特点及影响

相比于书面语语料,口语语料的最大特点是其有声性。这决定了在标注内容方面,口语语料应增加语音标注。因为口语语料本身无文字表述,为了研究与标注而进行的转写由母语者承担,因而无需进行文字标注。而为了了解学习者口语在词汇、语法、语体、语义、语用等方面的具体运用与表现情况,为了检索与研究的方便,皆需进行标注,这和书面语语料是完全相同的。

### 4.2 语音标注策略

#### (1) 单一标注与综合标注

有研究依据口语语料的有声性特点主张对口语语料只做语音标注,而不做其他

语言层面的标注。我们认为这种认识是非常片面的。有声性固然是口语语料的重要属性，但并非其唯一属性；我们固然关注学习者的汉语语音状况，但也关注其在词汇、语法、语体等方面的实际表现，关注其口语习得的全面情况。为此，就不能只做语音标注，还须进行全面的综合性的语言标注。必须明确认识到，口语语料库并不等于语音语料库，不能像语音语料库那样只做语音层面的单一标注，而要对中介语口语语料进行多层面的综合标注。

## （2）深层标注与浅层标注

从为教学与研究提供更多帮助的角度来看，标注的内容越多、层次越深、越细，可能越有价值。例如语音标注应该把停顿的时长多少算非正常停顿、声调偏误是第几声的偏误乃至应该用第几声、声韵母的具体使用情况都标注出来。我们曾经是这样认为的，也是这样做的。然而姑且不论标注的正确性与一致性问题，即以语料库使用者的研究需求而言，真的需要做这样深入细致的标注吗？如果使用者所采用的语言知识系统与标注所依据的系统不一样怎么办？如果使用者不认同、不相信标注结果怎么办？倘真如此，所做的全部标注不仅失去了应有的价值，甚至反而成为了使用者使用语料库的障碍，这就完全背离了人们建设语料库的初衷了。

也许，“不替用户做判断”是最明智的选择。语料库的建设者只要标出在母语者看来有问题的停顿、声调、声母、韵母等即可，至于算不算偏误、停顿多长时间算不正常停顿、声韵调等为什么不对、应该用哪个声母、韵母、声调等，皆无需标注，因为这些恰恰是语料库的使用者需要研究与解决的问题，无需建设者代劳。

## 4.3 语料标注的全面性

建设语料库的根本目的是为对外汉语教学及其相关研究服务。从学习者角度看，在汉语学习过程中，在语音、词汇、语法、语体、语义、语用等各个语言层面都可能出现偏误；从教学者角度看，教师和研究者们的关注点和研究兴趣是多方面的。因此，语料标注应包括语言的各个层面，便于教师和研究人员通过语料检索，查询并了解学习者汉语口头表达的各方面情况，并在此基础上进行相应的科学研究。

这就决定了中介语口语语料库的标注必然是一种全面标注，包括语音、词汇、短语、句子、语篇、语体、语义、语用、修辞格等。原则上能体现学习者口语特征的项目都应该进行标注。

语料库的功能只是为教师和研究人员提供检索语料的方便，而不是代替他们的研究工作，并不需要把所有的语言现象都研究清楚，提供彻底而清晰的分类结果。因此，语料标注又只能是一种浅层标注，即以便于检索为原则，能查询到所需要的语料即已达到目的。（张宝林，2016b）应尽量少标，避免任何画蛇添足的行为。例如标明某词存在声调偏误即可，至于应为第几声、误读成了第几声，皆无需标注。

## 5. 口语语料库的标注方法

### 5.1 相关理念

根据目前自然语言处理和中文信息处理的研究水平,语料标注想完全采用计算机自动标注的方式是不现实的,而只能采取以人工标注为主、计算机标注为辅的“人标机助”的方式。但摒弃传统而陈旧的人自为战的离线分包方式、采取互联网思维基础之上的“人机互助、人人互助”的在线众包方式则是完全可行的。这种方式不但可以提高语料标注的速度,而且可以改变标注模式,扩大标注的开放性与社会参与度,提高语料标注的正确性和一致性,提升标注的质量与效率。

### 5.2 标注工具

传统的手工标注方式及其采用的表示方法随意性很强,缺乏相应的理据。例如表示不正常停顿,一般的不正常停顿用1个后单引号(′)表示,严重的用2个后单引号(′)表示,最严重的用3个后单引号(′)表示。停顿正常与否母语者是可以凭语感加以判断的,但其时间长度多长算一般或严重的不正常停顿,则比较主观随意。单靠人工标注,不可能保证标注结果的客观一致性。

ELAN软件则非常直观,可以显示声波图、发音时长、停顿时间,可以显示国际音标以及视频图像等,还可以进行多层标注,十分适用于口语乃至视频语料的标注。而且可以更好地体现“不替用户做判断”的理念,软件客观显示了停顿时间的长度,算不算不正常停顿,算何种等级的不正常停顿,语料库建设者都无需过问,而由研究者自己去判断与论证。这也在一定程度上简化了建库程序,降低了口语库的建设难度。

## 6. 结语

口语库有其十分重要的作用、意义与价值,与书面语库相辅相成而不可互相替代。应充分认识口语库建设的重要性、必要性、紧迫性,积极开展口语库建设,更好地为对外汉语教学与相关研究服务。

毋庸讳言,口语库建设确实有其特殊性,存在诸多问题与难题,应通过多种方式,研究并解决相关问题,促进口语库建设。例如语料来源与分布问题,可以通过与不同国家的大学中文系、中文部、孔子学院等汉语教学单位进行国际合作,解决语料分布不均的问题。应尽快研制口语语料的收集标准,规定录音的音量、声道、制式等技术标准;研究并制定口语语料的标注规范,包括标注的目的、原则、内容、方法、工具、代码等,以便于相关实际工作的开展,确保工作质量与效率。对口语语料来说,语音和语体是其显著特点,对其标注规范的研究应给予充分的重视。

为了提高语料库建设的效率与水平,应开发基于网络的语料库建设平台,包括

语料收集与转写转录系统、语料标注与审查系统、语料库管理系统、语料检索系统等,使语料库建设更加便捷、高效。

### 参考文献

- Fang, S., Chen, Q., Yang, H., & Chen, H. (2016). A corpus-based pronunciation errors analysis: A case study of Chinese vowels produced by Japanese learners. In X. Lin, X. Xiao & B. Zhang (Eds.), *Proceedings of the 3<sup>th</sup> International Symposium on the Construction and Application of Chinese Interlanguage Corpora* (pp. 231-243). Beijing: Beijing World Publishing Corporation. [方淑华, 陈庆华, 杨惠媚, & 陈浩然. (2016). 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. 林新年, 肖奚强 & 张宝林(编辑), 2016第三届汉语中介语语料库建设与应用国际学术研讨会论文集(pp. 231-243). 北京: 世界图书出版公司.]
- Li, H. (2013). The construction and functional presentation of a spoken text corpus transcript retrieval system. In X. Cui & B. Zhang (Eds.), *Proceedings of the 2<sup>nd</sup> International Symposium on the Construction and Application of Chinese Interlanguage Corpora* (pp. 119-134). Beijing: Beijing Language and Culture University Press. [李航. (2013). 口语文本语料库检索系统的建立与功能呈现. 崔希亮 & 张宝林(编辑), 2013第二届汉语中介语语料库建设与应用国际学术研讨会论文集(pp. 119-134). 北京: 北京语言大学出版社.]
- Lin, X., Xiao X., & Zhang B. (Eds.). *Proceedings of the 3<sup>th</sup> international symposium on the construction and application of Chinese interlanguage corpora*. Beijing: Beijing World Publishing Corporation. [林新年, 肖奚强, & 张宝林(编辑). 2016第三届汉语中介语语料库建设与应用国际学术研讨会论文集. 北京: 世界图书出版公司.]
- Liu, Y. (2013). Discussions about the construction of Chinese interlanguage spoken corpora. In X. Cui & B. Zhang (Eds.), *Proceedings of the 2<sup>nd</sup> International Symposium on the Construction and Application of Chinese Interlanguage Corpora* (pp. 99-106). Beijing: Beijing Language and Culture University Press. [刘运同. (2013). 汉语口语中介语语料库建设刍议. 崔希亮 & 张宝林(编辑), 2013第二届汉语中介语语料库建设与应用国际学术研讨会论文集(pp. 99-106). 北京: 北京语言大学出版社.]
- Liu, Y. (2016). Comparison of frequently-used transcription systems of spoken language. In X. Lin, X. Xiao & B. Zhang (Eds.), *Proceedings of the 3<sup>th</sup> International Symposium on the Construction and Application of Chinese Interlanguage Corpora* (pp. 199-204). Beijing: Beijing World Publishing Corporation. [刘运同. (2016). 常用口语转写系统的比较. 林新年, 肖奚强 & 张宝林(编辑), 2016第三届汉语中介语语料库建设与应用国际学术研讨会论文集(pp. 199-204). 北京: 世界图书出版公司.]
- Liu, Z. (2013). Research on the HSK Spoken Corpus with speech recognition technology. In X. Cui & B. Zhang (Eds.), *Proceedings of the 2<sup>nd</sup> International Symposium on the Construction and Application of Chinese Interlanguage Corpora* (pp. 135-142). Beijing: Beijing Language and Culture University Press. [刘喆. (2013). 基于语音识

- 别的HSK口语语料库研究. 崔希亮&张宝林(编辑), 2013第二届汉语中介语语料库建设与应用国际学术研讨会论文选集(pp. 135-142). 北京: 北京语言大学出版社.]
- Lu, Q., & Tao, J. (2011). The construction and evaluation of a small-scale foreign students' spoken interlanguage corpus. In X. Xiao & W. Zhang (Eds.), *Proceedings of the 1<sup>st</sup> International Symposium on the Construction and Application of Chinese Interlanguage Corpora* (pp. 46-55). Beijing: Beijing World Publishing Corporation. [陆庆和, & 陶家骏. (2011). 小型外国学生口语中介语语料库的建立与价值. 肖奚强 & 张旺熹(编辑), 2011首届汉语中介语语料库建设与应用国际学术研讨会论文选集(pp. 46-55). 北京: 世界图书出版公司.]
- Ren, H. (2010). On the construction of Chinese interlanguage corpus: Using the HSK Dynamic Composition Corpus as an example. *Language Teaching and Linguistic Studies*, 6, 8-15. [任海波. (2010). 关于中介语语料库建设的几点思考——以“HSK动态作文语料库”为例. *语言教学与研究*, 6, 8-15.]
- Shi, C., & Zhang, R. (2013). The balance of interlanguage corpora. *Applied Linguistics*, 2, 117-126. [施春宏, & 张瑞朋. (2013). 谈汉语中介语语料库的平衡性问题. *语言文字应用*, 2, 117-126.]
- Yang, Y., Li, S., Guo, Y., & Tian, Q. (2006). Prospects of constructing a Chinese learners' spoken corpus. *Chinese Language Learning*, 3, 58-64. [杨翼, 李绍林, 郭颖雯, & 田清源. (2006). 建立汉语学习者口语语料库的基本设想. *汉语学习*, 3, 58-64.]
- Zhang, B. (2010). Avoidance and overgeneralization—An investigation of the acquisition of bǎ-sentence based on the HSK Dynamic Composition Corpus. *Chinese Teaching In The World*, 2, 263-278. [张宝林. (2010). “回避”与“泛化”——基于“HSK动态作文语料库”的把字句习得考察. *世界汉语教学*, 2, 263-278.]
- Zhang, B. (2011). *Status quo and strategies—Research on the teaching of Chinese as a second language*. Beijing: Beijing Language and Culture University Press. [张宝林. (2011). *现状与对策——作为第二语言的汉语教学研究*. 北京: 北京语言大学出版社.]
- Zhang, B. (2016a, August). "Research Orientation" and "Teaching Orientation"—On the application of Chinese interlanguage corpora. Plenary speech presented at the 2<sup>nd</sup> International Symposium on the Construction and Application of Chinese Interlanguage Corpora, Houston, Texas. [张宝林. (2016, 8月). “研究导向”与“教学导向”——谈汉语中介语语料库的应用问题. 发表于第二届汉语中介语口语语料库国际学术研讨会. 休斯顿, 得克萨斯州.]
- Zhang, B. (2016b, November). *Reconsidering the norm of annotation of Chinese interlanguage corpora*. Plenary speech presented at the 4<sup>th</sup> International Symposium on the Construction and Application of Chinese Interlanguage Corpora, Yangzhou, China. [张宝林. (2016, 11月). 关于汉语中介语语料库标注规范研究的新思考. 发表于第四届汉语中介语语料库建设与应用国际学术研讨会, 扬州, 中国.]
- Zhou, B. (2013). A spoken corpus based prosodic study of intermediate level Chinese learners from overseas. In X. Cui & B. Zhang (Eds.), *Proceedings of the 2<sup>nd</sup> International Symposium on the Construction and Application of Chinese Interlanguage Corpora* (pp. 355-367). Beijing: Beijing Language and Culture University Press. [周宝芯. (2013). 基于口语语料库的中级阶段留学生汉语韵律研

- 究. 崔希亮 & 张宝林(编辑), 2013第二届汉语中介语语料库建设与应用国际学术研讨会论文集(pp. 355-367). 北京: 北京语言大学出版社.]
- Zhang, L. (2016). A preliminary study on the transcription rules of Chinese interlanguage spoken corpus. In X. Lin, X. Xiao & B. Zhang (Eds.), *Proceedings of the 3<sup>th</sup> International Symposium on the Construction and Application of Chinese Interlanguage Corpora* (pp. 205-215). Beijing: Beijing World Publishing Corporation. [张蕾. (2016). 汉语口语中介语口语语料转写规则初探. 林新年, 肖奚强, & 张宝林(编辑). (2016). 第三届汉语中介语语料库建设与应用国际学术研讨会论文集(pp. 205-215). 北京: 世界图书出版公司.]