

基于语料库的第二语言学习者汉语介词习得特征考察

郑琬铃

广州理工学院 人文与教育学院, 广东 广州 510540

DOI:10.61369/HASS.2025070040

摘 要： 本文基于 150 万字的学习者中介语语料，对第二语言学习者汉语介词的使用情况进行了考察与分析。考察结果发现，在不分母语背景的学习者语料中，前五高频介词为：在、跟、对、从、帮，若将学习者语料区分母语背景进行观察，不同语别的学习者前五高频介词会出现差异，如：日语母语者为“在、跟、对、从、到”；韩语母语者为“在、跟、对、从、为了”。不管不同语别学习者前五高频介词有何差异，一样都有“在、跟、对”三者，同时前五高频介词对于学习者而言较易习得，大部分在 A2 阶段即可熟练使用，且大部分使用率高于目的语母语者，其中以印尼学习者的泛用情况最为明显。从偏误类别来看，遗漏类偏误最多，尤其是日、韩学习者。可见介词对于第二语言学习者来说，要掌握其使用规则具有一定难度。

关 键 词： 第二语言习得；介词；语言特征

A Corpus-Based Study on the Characteristics of Chinese Preposition Acquisition of Second Language Learners

Zheng Wanling

Guangzhou Institute of Science and Technology, College of Humanities and Education, Guangzhou, Guangdong 510540

Abstract： Based on a learner interlanguage corpus of 1.5 million words, this paper investigates and analyzes the usage of Chinese prepositions by second language learners. The findings reveal that among the learner corpus without distinguishing native language backgrounds, the top five most frequently used prepositions are: 在 (zài), 跟 (gēn), 对 (duì), 从 (cóng), and 帮 (bāng). When the learner corpus is differentiated by native language backgrounds, variations emerge in the top five most frequently used prepositions among learners of different languages. For instance, Japanese native speakers predominantly use 在 (zài), 跟 (gēn), 对 (duì), 从 (cóng), and 到 (dào), while Korean native speakers primarily use 在 (zài), 跟 (gēn), 对 (duì), 从 (cóng), and 为了 (wèile). Regardless of the differences in the top five most frequently used prepositions among learners of different languages, 在 (zài), 跟 (gēn), and 对 (duì) consistently appear. Furthermore, these top five prepositions are relatively easy for learners to acquire, with most learners becoming proficient in using them by the A2 level, and their usage rates often surpass those of native speakers of the target language, with Indonesian learners exhibiting the most pronounced overuse. In terms of error types, omission errors are the most prevalent, particularly among Japanese and Korean learners. This indicates that mastering the usage rules of prepositions poses a certain degree of difficulty for second language learners.

Keywords： second language acquisition; preposition; language feature

引言

我们统计自建的学习者中介语语料库各个偏误类别的数量，词汇偏误共有 25664 个，语类偏误有 3570 个，句型类偏误有 2313，混合类偏误有 1537 个。词汇类偏误明显高于其他三类许多，由此可见词汇习得对于二语学习者的重要性。进一步依据王还（1995）的分类，将名词、动词、能愿动词、形容词、数词、量词、代名词、指示词 / 定词、疑问词、复数词缀归类为实词；介词、连接词、副词、句末助词、时貌助词、助词归为虚词^[1]。其中学习者实词偏误有 13790 个，虚词偏误有 11874 个，两者的偏误次数并没有太大的差距。但是实词作为开放的词类，其偏误数量与封闭词类虚词相差无多，本身就能说明一定问题，可见汉语虚词的习得对于二语学习者而言具有一定难度。

对外汉语虚词教学应该筛选出最常用的虚词作为教学的重点。特别是那些虚实兼用的词是外国学生更难掌握的，更是重点和难点（李如龙、杨吉春，2004）^[2]。介词也是虚词的一种，并且大部分的介词都是由动词虚化而来的，因此某些介词的语法特征与动词还有相似之处。汉语的介词虽然数量不多，自成一个相对封闭的小类，但是在句中的语法、语用和语义功能甚为突出，且某些词在功能上虚实兼有，因此对于二语学习者来说，汉语介词的习得是其二语习得过程中的难点之一。

一、二语学习者介词使用特征考察

（一）不区分母语背景的二语学习者介词使用特征考察

在学习者中介语语料库中，学习者共计使用37195次介词（包括遗漏使用介词的503次），“在、跟、对、从、帮”²依使用次数排序为学习者使用前五的高频介词。表1为前五高频介词在各个等级³中的偏误数据。这里的介词统计数据是针对介词词汇选用、遗漏、冗赘的部分，如果是介词组成的介词词组的序误偏误，则不在此统计范围内。

观察下表可以发现，学习者前五高频介词的使用基本上没有什么问题，偏误率都在百分之十以内，特别是“帮”，学习者的偏误率极低。猜测可能是因为“帮”与其他介词相较，虚化程度较低，一般在教材中也是作为动词处理，因此对于学习者来说，使用上反而不会有太大困难。除此之外，学习者对每个介词的偏误表现皆有所不同，“在、对、从”在B2和C1之间的偏误率有显著差异，在其他等级间则没有；“跟”在A2、B1、B2、C1四个等级间偏误率都有显著差异；“帮”在A2和B1之间，B1和B2之间偏误率有显著差异。

表1 前五高频介词偏误率

	A2	B1	B2	C1
在	5.35%	4.22%	4.39%	1.43%
跟	1.15%	2.05%	1.13%	0.00%
对	3.42%	3.51%	2.68%	1.79%
从	3.77%	4.72%	4.12%	2.11%
帮	0.00%	0.11%	0.00%	0.00%

（二）区分母语背景的二语学习者介词使用特征考察

将学习者中介语语料分母语背景做观察，选择其中语料量排行前五的不同母语背景学习者语料进行分析。发现日语母语者前五高频介词为“在、跟、对、从、到”；英语母语者为“在、跟、

对、帮、从”；韩语母语者为“在、跟、对、从、为了”；越南语母语者为“在、跟、对、把、到”；印尼语母语者为“在、跟、对、从、把”。五个不同母语背景的学习者群体虽前五高频介词的使用略有不同，但都有“在、跟、对”三者，而英语母语者前五高频介词则与不分母语背景的语料相同，皆是“在、跟、对、从、帮”。表2为五个不同母语背景的学习者群体前五高频介词在各个等级中的偏误率。从表中数据可以看出，前五高频介词对于五个不同母语背景的学习者而言，大都是比较容易习得，可以在A2就习得。除了印尼语母语背景学习者对“把”的使用之外，他们直到B2才习得“把”。

介词偏误在五个不同母语背景的学习者群体中有较显著的差异，对日语母语者来说，“跟、到”的偏误率在A2~C1四个等级间差异明显，“在、对”的偏误率差异则体现在A2至B2与C1两个等级区块之间，“从”则是在A2和B1至C1两个区块间。对英语母语者来说，“跟、对、从”的偏误率在A2~C1四个等级间差异明显，“在”的偏误率差异在A2至B2和C1两个区块间有明显体现。对韩语母语者来说，“跟、对、从”的偏误率在A2~C1四个等级间差异明显，“在”在A2、B1至B2、C1三个等级区块中有显著差异，“为了”则在A2、B1和B2至C1三个区块中有所差异。对越南语母语者来说，“在、跟、到”的偏误率在A2~C1四个等级间差异明显，“对”在A2和B1至C1两个区块中差异明显，“把”则在A2、B1至B2、C1三个等级区块中有显著差异。对印尼语母语者来说，“跟、从、把”的偏误率在A2~C1四个等级间差异明显，“在、对”的偏误差异体现在A2、B1至B2、C1三者之间。虽然五个不同母语背景的学习者群体高频介词的偏误情况有所不同，但与不分母语背景的学习者语料一样，“跟”的偏误情况在学习者的四个语言水平等级之间皆有显著差异。

表2 五个母语背景学习者群体前五高频介词偏误率

日语					英语					韩语				
	A2	B1	B2	C1		A2	B1	B2	C1		A2	B1	B2	C1
在	5.14%	5.58%	6.84%	3.70%	在	3.94%	3.03%	3.84%	2.53%	在	8.40%	5.73%	4.08%	0.00%
跟	0.97%	1.91%	0.87%	0.00%	跟	2.22%	1.99%	1.16%	0.00%	跟	1.00%	3.34%	1.57%	0.00%
对	7.14%	5.57%	4.04%	0.00%	对	2.17%	3.82%	1.48%	0.00%	对	4.35%	1.28%	2.94%	0.00%
从	3.92%	5.14%	3.70%	3.85%	帮	0.00%	0.00%	0.00%	0.00%	从	0.00%	2.44%	6.82%	0.00%
到	1.20%	6.11%	2.53%	7.14%	从	0.00%	8.05%	1.69%	5.56%	为了	0.00%	4.17%	0.00%	0.00%
越南语					印尼语									
	A2	B1	B2	C1		A2	B1	B2	C1					
在	3.51%	1.60%	4.24%	1.96%	在	5.18%	3.41%	4.03%	1.47%					
跟	0.00%	1.33%	0.73%	0.00%	跟	0.00%	1.65%	3.53%	0.00%					
对	0.00%	2.99%	2.75%	2.44%	对	0.00%	1.88%	1.44%	0.00%					
把	0.00%	6.35%	5.34%	10.00%	从	0.00%	3.85%	5.88%	0.00%					
到	3.39%	6.25%	10.00%	0.00%	把	23.08%	13.89%	2.78%	0.00%					

从表3学习者前五高频介词正确形式使用率与表4汉语母语者介词使用率对比的数据来看，五个不同母语背景的学习者群体对前五高频介词的使用都是呈现多用的情况，特别是在A2和B1阶段。在这五个不同母语背景的学习者群体中，尤以印尼语母语者对高频介词的多用情况最为明显，尤其表现在“在”的使用上。虽然每个不同母语背景的学习者在“在”的使用上都高于汉语母

语者许多，使用率平均落在50~60（使用频率为万分位）之间，但是印尼语母语者“在”的平均使用率高达78，明显又高于其他四者不少。除了“在”之外，印尼语母语者“跟”的使用也相对较高，具体原因可能与其母语语言系统有关。

表3 五个母语背景学习者群体前五高频介词正确形式使用率

日语					英语					韩语				
	A2	B1	B2	C1		A2	B1	B2	C1		A2	B1	B2	C1
在	61.21	56.23	61.38	48.57	在	74.80	68.92	67.41	42.88	在	63.98	58.61	55.46	41.61
跟	42.33	50.92	35.43	14.94	跟	49.12	49.61	25.41	4.45	跟	58.11	46.84	36.88	6.12
对	10.79	22.71	31.39	18.68	对	12.56	19.89	29.89	21.72	对	12.91	25.04	24.34	14.68
从	10.17	10.07	10.10	15.57	帮	17.30	15.22	7.92	0.00	从	12.33	8.63	6.05	18.36
到	17.01	6.10	5.98	8.09	从	14.79	9.84	8.67	18.93	为了	1.76	7.45	10.47	9.79
越南语					印尼语									
	A2	B1	B2	C1		A2	B1	B2	C1					
在	58.16	69.07	63.04	38.79	在	91.42	83.12	79.31	59.61					
跟	69.79	59.43	29.21	5.43	跟	55.45	46.04	15.16	12.46					
对	14.80	26.02	30.39	31.03	对	7.99	24.26	37.90	30.25					
把	7.40	9.48	13.32	6.98	从	16.98	11.59	11.83	10.68					
到	30.14	9.64	6.77	18.62	把	5.00	4.79	19.41	7.12					

（注：正确形式使用率 = 正确使用次数 / 该母语背景学习者各级语料总量，频率是万分位。）

表4 汉语母语者介词使用率

在	跟	对	从	到	帮	为了	把
0.81	0.64	0.13	0.15	0.2	0.54	0.02	0.06

（注：使用率 = 使用次数 / 总字数，频率是万分位）

二、二语学习者介词的基本使用情况考察

我们从语料库中提取了日语（8572例）、英语（6016例）、韩语（4025例）、越南语（4570例）、印尼语（3775例）等五个不同母语背景的学习者群体使用介词的全部用例，共26958例（该处总和及以下数据皆不包含遗漏使用介词的例子）。其中使用正确的有26473例，占98.2%，使用错误的有485例，占1.80%，另外遗漏偏误有388例。

肖奚强等（2015）^[3]统计了各类介词的偏误，得出时间介词的偏误类型从多到少的排列顺序为：遗漏——误代——错序——误加；空间介词为：遗漏——错序——误加——误代；对象介词为：误代——遗漏——误加——错序；依据介词为：误代——遗漏——误加；缘由介词为：误代——错序——误加——遗漏；并列介词为：误代——遗漏——误加——错序；选择介词为：误代——遗漏——误加；递进介词为：误代——错序——遗漏——误加。从肖文对介词偏误类型的统计结果中可以看到，二语学习者介词偏误类型较多的为误代和遗漏。我们将从学习者语料库中得出的学习者使用介词的数据列于表5，统计出学习者介词偏误类型从多到少为：遗漏——误加——误代——错序，其中错序的偏误例子仅有两例，基本可以忽略不提。虽然我们的统计结果与肖文的数据稍有不同，但是在两个结果中，学习者介词的

遗漏偏误都占很大比重，由此也可知道，对于二语学习者来说，汉语介词使用的难点在于对介词的使用规则掌握不够，不知何时该使用介词。

另外，从我们的统计结果中还可以看到，日语、韩语母语者介词偏误情况稍比其他三者高，且遗漏偏误在这两个学习者群体全部介词偏误中所占的比例也明显较其他三者要高；英语母语者除介词错序偏误外，其他三个偏误类型的分布都很平均；印尼语母语者介词偏误从多到少的排序稍有不同，以误加偏误为最多，其次是遗漏偏误。造成日语、韩语母语者介词偏误率较其他母语背景的学习者高，且遗漏介词率也较高的原因，可能是在这两个语言中没有介词，相反的它们有丰富的格助词，接续在名词性成分后面，用以在句子中标明句子成分或者指示名词与动词间的关系。比较令我们意外的是，在前人对二语学习者介词的偏误统计中，一般认为误加偏误是欧美学生介词偏误中最多的类型，如崔希亮（2005）^[4]提到“对于欧美学习者来说，过度使用介词是一个突出的问题，很多不需要介词的地方都加上了介词”。但是在我们的统计数据中，英语母语者介词的误加偏误并不特别明显，基本和误代、遗漏偏误数持平。另外，崔希亮（2005）还在文章中提到“欧美学生在使用介词的时候是有自己的特点的，这种特点首先表现在使用频率上。与日本学习者、朝韩学习者以及以汉语为母语的中国人相比，欧美学生使用介词的频率要高得多”。根据之前的研究中对五个不同母语背景的学习者群体使用介词情况的统计发现，确实五个不同母语背景的学习者群体汉语介词的使用频率都高过汉语母语者，但是并未发现英语母语者介词的使用率有明显高于其他四者的迹象，反而是印尼语母语者对高频介词的多用情况在五者中是最为明显的。且从表5的结果来看，印尼语母语者介词误加偏误率也较其他四者高，具体原因需要回到学习者母语和汉语的对比中去寻找答案。

表5 五个母语背景学习者介词偏误类型统计

	使用次数	偏误小计	偏误率	遗漏次数 %		误加次数 %		误代次数 %		错序次数 %	
日语	8572	359	4.19%	186	51.81%	85	23.68%	88	24.51%	0	0.00%
英语	6016	163	2.71%	56	34.36%	56	34.36%	51	31.29%	0	0.00%
韩语	4025	142	3.53%	70	49.30%	35	24.65%	37	26.06%	0	0.00%
越南语	4570	110	2.41%	44	40.00%	32	29.09%	34	30.91%	0	0.00%
印尼语	3775	98	2.60%	32	32.65%	43	43.88%	21	21.43%	2	2.04%
平均					41.62%		31.13%		26.84%		0.41%

三、结束语

通过对学习者中介语语料与汉语母语者语料的对比,我们发现不管学习者的母语背景为何,大部分的高频介词在 A2 阶段即可习得,高频介词的使用并没有想象中困难。

除此之外,一般认为欧美学习者汉语介词的使用存在较明显的误加现象,但是经过我们对英语母语者介词使用的整体观察,发现其介词的误加情况并不突出,与误代、遗漏持平。同时,也不存在欧美学生使用介词频率比其他母语背景的学习者要高得多的现象,反而是印尼语母语者高频介词的多用情况是在五个不同母语背景的学习者群体中最为明显的,其误加类偏误也较其他四者多。

从二语学习者介词使用的偏误情况来看,介词的遗漏是最多的。同时在五个不同母语背景的学习者群体中,日语、韩语母语者的介词偏误情况是较为明显的,尤其体现在遗漏类偏误上,究其原因,可能是因为在日语、韩语中没有与介词相对应的词类,汉语介词在句子中所代表的语法功能,在日语、韩语中一般由格助词或其他语法结构替代,因此对于这两个母语背景的学习者群体来说,建立汉语介词的使用概念要较其他母语背景的学习者群体难。

总之,通过语料库对二语学习者介词使用进行总体观察,我们不仅发现学习者介词使用的不均现象,也更新了不同母语背景

的学习者群体实际的介词使用情况,对二语学习者的中介语图像有更清楚的认知,在未来的汉语教学中,可以更好的考虑不同母语背景的学习者汉语介词的教学安排和设计,以减低学习者汉语介词的习得难度。

注释

1.王还(1995)根据词的语法特征和句法功能对现代汉语的词做分类,共分为十三类。名词、代词(也就是本文中的“代名词”)、数词、量词、动词、助动词(本文中的“能愿动词”为助动词的一部分)、形容词为实词;副词、介词、连词(连接词)、助词(包含有本文的“句末助词、时貌助词、助词”三类)、叹词、象声词为虚词。

2.“帮”在我们使用的词类标记中具有两种词性,分别为动词和介词。“帮”的动词被标记为 VC,也就是动作单宾述词,顾名思义在语意上是属于动作述词,需要两个参与论元;在语法上,这两个参与论元都是名词组(包括经名物化的述词组),分别充当述词的主语和宾语,例如:我帮你、请你帮个忙。“帮”做介词用的语法表现和扮演的语意角色与“替”相同,都是 benefactor、goal 和 theme,例如:哥哥帮妈妈洗碗。

3.语料库中的等级划分以《欧洲语言共同参考框架》为准,分为 A2、B1、B2、C1 四个等级。

参考文献

- [1]王还.对外汉语教学语法大纲[M].北京:北京语言学院出版社.1995.
- [2]李如龙,杨吉春.对外汉语教学应以词汇教学为中心[J].暨南大学华文学院学报,2004,(04):21-29+49.
- [3]肖溪强.外国留学生汉语偏误案例分析[M].,北京:世界图书出版公司.2015.
- [4]崔希亮.欧美学生汉语介词习得的特点及偏误分析[J].世界汉语教学,2005,(03):83-95+115-116.