

A Corpus-Based Error Analysis of Vocabulary Made by Korean Students

Dixiao Wu

International College of Chinese Language and Culture, Hunan Normal University, Changsha Hunan
Email: wudixiao1994@qq.com

Received: Mar. 20th, 2020; accepted: Apr. 20th, 2020; published: Apr. 27th, 2020

Abstract

The author builds a corpus of spoken language tests for intermediate Chinese language learners, marks and counts the vocabulary errors made by Korean students, and summarizes the characteristics of Korean students' vocabulary errors, that is, the largest number of confusable words errors, followed by function words errors, while the number of other types of errors such as the misuse of verb-object detachable words is small. At the same time, the author also traced the reasons by analyzing the negative transfer of the mother tongue and the target language. This study uses a self-built corpus to conduct empirical research, which is quite convincing, but the scope of the corpus and the subjects is limited, and a larger range of research can be carried out later.

Keywords

Vocabulary Error, Korean Students, Corpus

基于语料库的韩国学生汉语词汇偏误分析

吴狄潇

湖南师范大学国际汉语文化学院, 湖南 长沙
Email: wudixiao1994@qq.com

收稿日期: 2020年3月20日; 录用日期: 2020年4月20日; 发布日期: 2020年4月27日

摘要

作者通过自建中级汉语学习者口语测试语料库, 标注和统计出韩国学生口语测试中词汇偏误情况, 并归纳总结了韩国学生汉语词汇偏误的特点, 即别词类偏误数量最多, 其次是虚词类, 而其他如离合词误用等偏误类型的数量较少。同时, 作者还追寻了原因, 从母语负迁移和目的语负迁移两个方面进行了分析。本研究采用自建语料库的方式, 通过实证研究, 颇具一定的说服力, 但语料和被试范围有限, 后续可以进行更大范围的研究。

关键词

词汇偏误, 韩国学生, 语料库

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在学习汉语的外国人的中介语系统中, 词语偏误(error)是大量的, 而且是几乎随着学习的开始就发生的。随着词汇量的增加, 发生的词语偏误也越来越多。因此, 如何最大限度地、有效地防止偏误的发生, 成了汉语教师与教学法研究者十分关注的问题[1]。运用偏误分析理论, 归纳总结学生口语表达中出现的词汇偏误类型, 分析其产生原因, 并提出针对性建议, 从而有效地提高学生汉语口语表达能力, 是一个既有理论价值又有现实意义的课题。中级阶段的学生积累了一定数量的词汇和语法项目, 尤其渴望提高自己的口语表达能力。而专门针对中级水平韩国学生汉语口语词汇偏误的研究并不多见, 已有的研究也很少是基于自建口语语料库进行研究。因此, 本研究运用偏误分析理论, 基于自建口语测试语料库, 统计中级水平韩国学生汉语口语词汇偏误, 总结其特点, 追寻偏误成因。

2. 构建语料库

2.1. 语料概况

自建语料库的语料来源系湖南师范大学国际汉语文化学院 D 班语言生各学期汉语口语(三)课程期末考试。D 班语言生汉语水平为中等, 接受汉语课教学时数在 80~120 学时左右, 词汇量在 1200 词左右。因该水平中韩国学生最多, 共计 21 人, 故选做本研究的研究对象。其中前 9 人语料采录于 2012 年 1 月 4 日, 后 12 人采录于 2017 年 1 月 9 日。口语考试内容主要包括朗读句子、朗读段落及命题说话三个部分。朗读类题型不符合词汇偏误研究需求, 因此我们仅选取命题说话部分进行研究。命题说话时长 3 分钟, 内容由学生在考前事先准备, 可供选择的话题有: 《学习汉语的经验》、《我在长沙的学生活》、《说说来中国以后让你感到新奇的事情》和《谈谈旅游》等。

2.2. 建库步骤

本研究使用索尼 PCM-D100 录音笔在安静的教室进行音频录制, 录音格式为*.Wav 格式。接着笔者用音视频转写标注工具软件“语宝标注”以分句为单位进行时间标记, 逐句转写并标注, 总共转写了完全与原始音频同步的约 1 万字左右的真实文本语料, 自建了一个中级水平的韩国学生汉语口语测试语料库。除汉语词汇偏误分类表中的例句外, 本文使用的例句全部提取自这个语料库。语料库建立完成后, 笔者使用由荷兰内梅亨马普心理语言学研究所开发的跨平台多媒体转写标注软件 ELAN 的检索功能对语料库中所有词汇偏误标注代码进行检索, 得到本语料库中各类词汇偏误的数据。

3. 词汇偏误分类探析

3.1. 词汇偏误分类及标注依据

本研究采用《汉语中介语语料库标注规范研究》[2]一书中的词汇标注研究作为词汇偏误分类及标注

的依据,将词汇层面的偏误共分为 10 种,分属于 3 个大类。缺词包括缺虚词、缺实词;多词包括多虚词、多实词;错词包括生造词、离合词误用、外文词、构词语素错位、同义词叠加、别词(包括词替换、单双音节词的误用、近义词或形近词的误用、同音词或音近词的误用)。各类词汇偏误的标注代码见表 1,此处例句均为书中原例。

Table 1. Chinese vocabulary error classification
表 1. 汉语词汇偏误分类表

偏误名称	标注代码	示例
缺虚词	[QXC]	还有我最近开始踢足球【】[QXC了]。
缺实词	[QSC]	因为祥林嫂【】[QSC是]嫁过二次的女人。
多虚词	[DXC]	我【的】[DXC]家附近有亚运村。
多实词	[DSC]	你真是【个】[DSC]我的好朋友。
别词	[CCb]	他是韩国人,还有跟我一样的【岁】[CCb岁数]。
生造词	[CCz]	免费下载歌曲引起了人们不再买昂贵的 CD,而导致了光碟 CD 的【卖量】[CCz]逐渐下降的后果情况。
离合词误用	[CCh]	现在我也有机会【见面】[CCh1]他们。 我想【应推销员的聘】[CCh2]。
外文词	[CCw]	因为我不是中国人,我一点儿不懂中文和中国人的种种【culture】[CCw]。
构词语素错位	[CCx]	一起看电影,一起吃顿饭,一起商量自己的困难……代沟什么时候可以【失消】[CCx]了。
同义词叠加	[CDJ]	我就不用想就可以说我是【很非常】[CDJ]喜欢。

3.2. 词汇偏误分类及分布情况

3.2.1. 缺虚词

这一类偏误在本研究的语料库中共计出现 80 次。例如:

(1)但是,两天以后我接到了一个电话,通知我【及格】[CCb资格]被取消【】[QXC了]。

结构助词“了”在句尾表示变化或出现新的情况。例(1)缺少“了”的话,会造成句意缺失,并且句子也不够完整。

(2)但是吃……吃【】[QXC完]以后,【】[QSC发现]牛……牛蛙的肉好像【】[QXC跟]鸡肉的味道差不多。

动作“吃”之后缺少结果补语“完”;“跟”作为介词,用于引进“牛蛙的肉”比较或比拟的对象——“鸡肉”。

(3)长沙的天气跟韩国【】[QXC的]完全不一样。

“长沙的天气”和“韩国”属于不同的逻辑范畴,无法进行类比,需要加上结构助词“的”表示“韩国的天气”才能进行比较。

3.2.2. 缺实词

这一类偏误共计出现 75 次。例如:

(4)首先,大连人他们【都】[DXC]【】[QSC有]自己的方言,大连人一般说儿化。

例(4)缺少谓语动词,不能成句。

(5)我感觉到她说的是没有票,但是我【】[QXC得]按时到北京,所以要买今天晚上【要】[DSC]出发【】[QXC的]【】[QSC票]。

例(5)谓语动词后缺少宾语,无法成句。

(6)我觉得这个经验跟.....经验跟那个观点【】[QSC 是]互相呼应的。

例(6)为“是.....的”句型表示对主语的描写或说明,有加重的语气的作用,在这里缺少判断动词“是”。

3.2.3. 多虚词

此类偏误共计出现 81 次。例如:

(7)但是没有失败【的】[DXC]【】[QXC 就]没有得到【的】[DXC]。

(8)我来中国以前非常期待【的】[DXC]中国的生活。

(9)但是来到长沙以后,长沙【的】[DXC]人喜欢辣的,他们不喜欢甜的。

在多虚词这类词汇偏误中,韩国学生“的”字滥用的情况非常多。如,例(7)中“失败的”和“得到的”形容词直接充当宾语是不符合汉语语法规则的;受到母语负迁移的影响,在韩语中形容词是可以作宾语的,所以韩国学生在汉语学习中普遍存在这个问题。例(8)动词“期待”和宾语“中国的生活”之间不需要结构助词“的”。例(9)中修饰成分“长沙”和中心名词“人”之间结构紧密,不能分离,类似一个复合名词,不需要使用“的”。

(10)【因为】[DXC]第一个问题是吃的问题.....吃的问题。

(11)【可是】[DXC]其中我觉得最喜欢的爱好是国内旅游。

(12)但是她说什么.....她说什么.....什么话,【但是】[DXC]我【】[QXC 都]听不懂【了】[DXC]。

除了结构助词“的”滥用之外,关联词的过度使用情况也非常多。结合上下文,例(10)前后不存在因果关系,不需要使用连词“因为”;例(11)前后没有出现转折,不需要使用“可是”。例(12)中“但是”在句中无意义的重复,也是多余的。

3.2.4. 多实词

这一类偏误共计出现 24 次。例如:

(13)我跟朋友们一起去烟台港,票.....票船,在这我【收到】[DSC]定.....定了船票。

动词“收到”与前后文毫无联系,属于多余的成分。

(14)但是有【】[QSC 很多]机会去张家界或者凤凰【的】[DXC]【多】[DSC]。

例(14)的偏误略显复杂,属于句式杂糅。可以修改为“但是有很多机会去张家界或者凤凰”,也可以改为“但是去张家界或者凤凰的机会多”。基于最大限度保留原句结构和词汇的原则,本文更倾向于第一种修改方法,在“机会”前加修饰词“很多”,后面的结构助词“的”和形容词“有”为多余成分。

(15)我一边激动,【我】[DSC]一边害怕。

例(15)两个小句之间主语相同,后面的主语应该省略。

3.2.5. 别词

将词替换、单双音节词的误用、近义词或形近词的误用、同音词或音近词的误用归入别词,是因为这几类词误用在表现形式上都属于应用某一个词却使用了另一词的情况。这一大类在本语料库中出现的次数最多,共计 90 次。例如:

(16)第一次来这儿,经验很多.....【经验】[CCb 经历]过了很多事。

“经验”是从多次实践中得到的知识或技能,常做名词;“经历”作动词表示亲身见过、做过或遭受过,作名词指亲身见过、做过或遭受过的事。这个句子意在表达经历了许多事情,而不在于做过之后获得的知识经验,应使用“经历”。

(17)咱们老师告诉我,长沙的人【为了】[CCb 因为]天气的影响,不喜欢甜的,他们喜欢辣的。

“为了”是介词,表目的;“因为”也是介词,表示接在后面的部分是原因。这个句子是解释长沙人为什么不喜欢吃甜的而喜欢吃辣的,引出的是原因,所以应该使用“因为”。

(18)只会……只会说你叫什么名字、你……这里……我想去这里，这两【个】[CCb 句]很简单的话。

(19)这【个】[CCb 趟]火车【】[QSC 要]【多长】[DXC] 29 小时。

(20)听说在中国，蛙是一【个】[CCb 种]常见的食材

如例(18)、(19)和(20)所示，在别词这一大类中，本语料库中出现的量词误用集中表现在量词“个”的泛化。

3.2.6. 生造词

生造词是学习者自造的汉语中不存在的词。在本语料库中生造词仅出现了两例。

(21)还有因为以前我想做……想做翻译工作，所以决定去……决定去【做】[CCb 当]【交换学生】[CCz 交换生]。

(22)还有他们的【吃的味道】[CCz 口味]也完全不一样。

例(21)中，该学习者直接将名词“学生”和汉语语素“生”直接等同，将“交换生”说成了“交换学生”；例(22)结合上下文分析得出这里“吃的味道”是指“口味”的意思。

3.2.7. 离合词误用

(23)我在长沙跟很多新的朋友【见面】[CCb 认识了]，【】[QSC 是]他们帮助我适应长沙的生活。

离合词是汉语学习者的重点和难点，但在本语料库中没有出现离合词的使用偏误情况，不排除学生采取“回避策略”。例(23)是唯一一处使用了离合词而有偏误的句子，但综合句意，此处为别词偏误。

3.2.8. 外文词

外文词偏误指学习者用外文词代替汉语词，对于外文词的标注可以让研究者清楚地观察到某一阶段学习者的汉语词汇掌握情况，有哪些词是这一阶段学习者应该掌握但却未能掌握的，可以为对外汉语词汇教学提供另外一种参考。本语料库中没有出现这种偏误类型。

3.2.9. 构词语素错位

这类偏误是学习者将词中构词成分写错顺序的偏误，仅出现 1 次。

(24)然后我跟朋友们一起玩后，一起去烟台港……烟台港搭乘【】[Cqs 轮船]回【国韩】[CCx 韩国]……回国。

值得讨论的是，除词内语素错序外，词语在句子内部错序的情况也屡见不鲜。但是因为这类偏误涉及到句法层面而非词汇层面，因此本研究没有在语料库中进行标注。例如：

(25)然后，【常常】[JCx]他陪我去逛街。

“常常”是副词，修饰后面的谓语“陪我去逛街”，并放在谓语动词的前面，所以正确的语序应该是“他常常陪我去逛街”。

3.2.10. 同义词叠加

在本语料库中，同义词叠加型的词汇偏误一共出现了三次：

(26)我来中国以前，非常期待中国的生活，【还有】[CCb 还]听【过了】[CDJ]很多故事。

“过”和“了”在动词后都表示动作已完成，例(26)中任选其一即可。

(27)我通过国内旅游，我的……我享受和放松……放松的时……放松的时间【还有和】[CDJ]消除我的疲劳。

“还有”与“和”作为连词都可以表并列关系，二选一即可。

(28)从三月到四五月，为了准备上课和在中国的生活……在中国的生活，我【】[QSC 变得]【非常特别】[CDJ]忙了。

“非常”和“特别”都是程度副词，在这里都可以用来修饰后面的形容词“忙”，只需要一个即可。

Table 2. Statistics of Chinese vocabulary errors made by Korean students
表 2. 韩国学生汉语词汇偏误统计总表

偏误名称	偏误例数	偏误比(%)
缺虚词	80	22.47
缺实词	75	19.66
多虚词	81	22.75
多实词	24	6.74
别词	90	25.28
生造词	2	0.56
离合词误用	0	0
外文词	0	0
构词语素错位	1	0.28
同义词叠加	3	0.84
总计	356	100

3.3. 词汇偏误表现特点

从表 2 可知，中级水平韩国学生汉语词汇偏误有以下特点：

1) 别词类偏误数量最多，共 90 例，占总偏误比重的 25.28%。可见韩国学生在汉语习得过程中词义掌握不够准确是造成口语词汇偏误的主要原因之一。

2) 两大缺词偏误中，缺虚词有 80 例，占总偏误比重的 22.47%；缺实词类偏误数是 75 例，占总偏误比重的 19.66%。其中，助词“的”的缺失共计 15 例，占缺词偏误的 9.68%。两大多词偏误中，多虚词类的偏误一共出现了 81 例，占总比重的 22.75%；多实词类偏误数为 24 例，占总偏误比重的 6.74%。其中，助词“的”的多余多达 29 例，占多词偏误的 27.62%。在别词类偏误中，出现了 1 例将“的”和“了”混用。在这三大主要的词汇偏误类型中，助词“的”的误用共计 45 例，占总偏误比重的 12.64%，可见中级水平韩国学生对“的”字的习得还存在着很多问题。

3) 其他的偏误类型如生造词、构词语素错位和同义词叠加数量较少；离合词误用和外文词偏误没有出现。但考虑到标注过程中存在的问题和困难，如学生采取“回避措施”的情况难以统计，并且本研究构建的语料库容量太小，所以这些偏误的具体情况有待建立更大的语料库进行深入考察。

4. 词汇偏误原因分析

词汇，是一种语言里所有的词(或特定范围的)和固定短语的总和[3]。韩国的语言和文化受中国的影响古已有之，韩语的发展受汉语的影响颇深。但同一国度不同地区的方言差异都如此之大，就更不用说国别之间了。时至今日，韩语随着民族的发展和历史的演变早已有了自身的词汇语义体系。汉语和汉语之间的对应词也往往是形近意殊。不同的语言发展的条件不同，它们的词汇就形成了各自发展变化的轨迹。两种语言的词汇系统会有许多共同点，但是共同点再多，它们发展变化的轨迹也不会是重合的(鲁健骥，1987)。这就使韩国学生在汉语学习的过程中常常会因两种语言词语的差异而产生偏误。韩国学生词汇偏误产生的原因是多方面的，下面从母语负迁移和目的语负迁移两个方面进行分析。

4.1. 母语负迁移

相较于汉语和其他语言，汉、韩语之间由于地理、历史和文化等原因，关系非常密切。韩语词多由

汉字词演变而来,韩语词中的60%以上的词由汉字词组成,并且在常用的韩国语名词中的90%为汉字词[4],所以汉字词对在韩语中扮演着重要的角色。但是很多韩语中的汉字词和汉语的词汇只是词形上相同或相似,词义上却有所不同,这些差异导致了韩国学生在学习汉语时出现词汇偏误。

如例(29)中,“最高”在韩语中是“最厉害、最棒、最好”的意思,该考生在汉语口语考试中直接用了韩语词中的意思去形容日本最好的大学,正是受到了母语负迁移的影响。

(29)所以我考【】[QXC了]日本最【高】[CCb好]的大学,而且通过【】[QXC了]考试。

除了近义词或同形词的影响之外,韩语中没有或者少有的词类也会影响韩国学生学习汉语词汇。比如汉语中存在着丰富而多变的量词,是韩国学生学习的难点之一。

(30)只会……只会说你叫什么名字、你……这里……我想去这里,这两【个】[CCb句]很简单的话。

如例(30)中,量词“个”的滥用在汉语学习者普遍存在,主要原因是学习者的母语中不存在这么多对应的量词。

4.2. 目的语负迁移

随着汉语学习时间越来越长,中级水平的韩国学生习惯用已掌握的有限汉语词汇和汉语知识,去类推新词和新的语言点知识。不恰当的套用导致了偏误的出现。

(31)我在北京旅行的时候遇到了很多困难,但是旅行后获得了很多经验,所以我下次【再】[CCb还]想去旅行。

例(31)中,学习者直接使用已有的汉语知识去表达“做相同的动作或行为”,没有区分“再”和“还”的区别。“再”的基本用法是表示添加相同的动作或情况,也含有重复发生的意思。多用在将发生的事情,用在此处无疑是不恰当的。

(32)所以我说【】[QSC汉语]的时候,他常常我说错的话【改变】[CCb改正],这样帮我。

例(32)中“改变”一次只有改换、更动的意思,并无纠错的用法;“改正”才有纠正错误的意义。两个词形近义也近,造成了学生使用出现偏误,属于目的语负迁移。同时谓语动词的位置也不符合汉语的正常语序,这是受韩语SOV语言属性影响,属于母语负迁移影响在句子层面的表现。

5. 结语

本文以自建的中高级汉语学习者口语测试语料库为基础,对韩国学生的汉语词汇偏误进行分析和研究。结果发现:1)在约1万字的语料中,别词类偏误数量最多,词义掌握不够准确是造成口语词汇偏误的主要原因之一;2)虚词类偏误中,结构助词“的”和连词的误用所占比重最大;3)中级水平韩国学生汉语虚词和实词的偏误分布存在不对称性,虚词类偏误多于实词类偏误;4)其他如离合词误用等偏误类型的数量较少,有待建立更大容量的语料库进行考察。究其原因,一方面韩国学生受母语负迁移影响,错误地将发音、词形甚至词义都跟汉语相似的汉字词,直接与汉语划等号;另一方面韩国学生把所学到的有限的、不充分的汉语知识,用类推的方法,不适当地套用在新的语言现象上,从而产生了偏误。

致 谢

感谢导师李斌老师提供的学生口语测试语料以及对本研究的指导。

基金项目

项目名称:基于自建多模态口语测试语料库的韩国学生语音偏误研究;项目编号: CX2017B218;项目类别:湖南省研究生科研创新项目;项目时间:2017年1月至2020年6月。

参考文献

- [1] 鲁健骥. 外国人学习汉语的词语偏误分析[J]. 语言教学与研究, 1987(4): 122-132.
- [2] 张宝林. 汉语中介语语料库标注规范研究[M]. 北京: 北京大学出版社, 2019: 95-101.
- [3] 黄伯荣, 廖序东. 现代汉语(增订五版) [M]. 北京: 高等教育出版社, 2011: 205.
- [4] 李翊燮. 韩国语语言学通论[M]. 北京: 外文出版社, 2010: 264.