

基于语料库的英语学习者歧义偏误分析

周致宇

湖南科技大学 外国语学院

摘 要：歧义是英汉语中共存的语言现象，歧义偏误是学习者语言习得过程中容易产生的常见偏误，歧义偏误的研究对于指导英语教学、提升学习者语言表达及交际能力有着重要的意义。通过对中国学习者英语语料库中的语料进行统计分析，从学习者认知水平、对比语言学和语言迁移的角度来看，中低层次学习者由于单词等基础知识的匮乏更容易产生词汇歧义偏误，高层次学习者则更倾向产生语法歧义偏误。教师要注重学习者英语思维和跨文化意识的培养、创建适宜的语言环境和并重视词汇教学，从而消除或更好地利用歧义。

关键词：中国学生；歧义；偏误分析

引言

外语学习中，学习者认知结构随新知识积累而变化，生词、一词多义等现象易引发学习困难，尤其体现在英语阅读与写作学习中。当前外语教育愈发重视读写能力，本文通过分析中国学习者英语语料库数据，探究歧义偏误成因，为英语教学提供理论支撑。

英语歧义主要分为口语歧义与书面语歧义，其中书面语歧义可以划分为以下三类：一是词汇歧义，含一词多义与词性歧义（即词汇量有限，无法实现“一词一义”）；二是语法歧义，含句法功能、成分结构及指代关系歧义；三是语用歧义，指因脱离上下文语境而导致的句子语用意义模糊。

一、理论与数据

歧义是语言自然属性，分析学习者歧义偏误可揭示二语习得规律、明确教学重难点。本章基于语料统计，针对歧义偏误产生问题进行分析。

（一）理论基础

文章采用了对比分析理论和偏误分析作为理论基础。Lado 指出，通过对两种语言的语音、形态、句法甚至文化这四个系统进行结构对结构的比较，目的是为了发现两种语言之间的相似点与不同点，最终得以预测学习者在语言习得中容易学会或者难以学会的领域（Lado R 1957）。

Corder 在《学习者错误的重要性》中区分了“错误”与“偏误”：错误近似口误，学习者可自我识别并修正；偏误是对目标语的系统、长期偏离，学习者当前阶段无法自我修正（Corder S P 1967）。Ellis 详细阐释了偏误分析的程序：选取语料样本、识别偏误、分类描述、解释偏误、评估并制定教学措施（Ellis 1989），也正

是基于此五个步骤，确定本文行文结构。

（二）数据来源与处理方式

本章中的偏误语料均来自桂诗春、杨惠中建立的中国学习者英语语料库（桂诗春，杨惠中 2003）。语料库中语料为学习者在日常学习、写作中的书面材料。针对这些语料，笔者做了以下处理：

1. 根据语言学习年限以及不同的水平，选取子库：ST2（中学生，低层次学习者）、ST5（低年级英语专业生，中层次学习者）、ST6（高年级英语专业生，高层次学习者）

2. 使用 AntConc3.5.9 软件检索已被标记为“WD7”的歧义偏误语料，之后将检索的全部语料平均分为 15 组，在每一组中分别抽取 6 个偏误项并观察其特征。

3. 由于语料库中仅仅只标注了歧义偏误的大类，为了进行详细讨论本文将歧义偏误类别继续进行细分，分为词汇歧义、语法歧义与语用歧义，其中，词汇歧义包含了词汇拼写的错误、词性原因导致的歧义偏误；语法歧义包含了在句法功能、句子成分结构以及指代关系等层面上导致的歧义偏误；语用歧义则是指在语言的语用分析层面上所导致的歧义偏误。在此分类基础上，将所获取的语料进行人工识别并且分类统计。

4. 根据统计得出的数据，从各方面对偏误进行比较分析，并提出有针对性的教学建议。

二、结果与讨论

通过使用 AntConc3.5.9，对歧义偏误 WD7 进行检索，发现不同层次的学习者在英语习得过程中出现的歧义偏误频数分布几乎没有差异，具体的学习者偏误频数详见表 1：

表 1 不同层次学习者偏误频次总体情况

歧义偏误	学习者层次	底层学习者	中层学习者	高层学习者	总计
	偏误频数	330	270	268	868

从统计的数据来看，歧义偏误的出现频数与学习者层次高低呈现负相关，随着学习者学习年限和水平的提高，歧义偏误出现的频数逐渐降低，这符合二语习得的一般规律，但是仅从偏误的频数来确定学习者歧义偏误的总体情况是不全面的，因为不同层次学习者的偏误总体频数不同，由此计算的偏误频率就有失准确，所以还需要通过计算每个层次学习者的偏误标准化频率来进行比较，再对第二语言习得规律进行说明。

（一）标准化频率计算

标准化频率计算是指基于统一标准，针对同一个语法偏误在不同总数的总体中出现的实际频次进行统一计算，计算公式如下：

标准化频率（每万 / 十万 / 百万词）= $\frac{\text{实际观测频数}}{\text{总体频数}}$

= × 10000/100000/1000000

该公式是指在每万 / 十万 / 百万词中，同一个语法现象出现的频率大小。中国学习者英语语料库子库中，ST2、ST5、ST6 中偏误语料的总体频数分别为：208088、214510、226106。根据上述公式，结合表 2 可以计算出每万词中，各层次学习者在歧义偏误上的标准化频率，如表 2：

表 2 各层次学习者歧义偏误的标准化频率

歧义偏误	学习者阶次	低层学习者	中层学习者	高层学习者
	歧义偏误频次	330	270	268
	标准化频率	15.86	12.59	11.85

从表 2 中可以看出，对比不同阶层的歧义偏误标准化频率，低层次学习者的标准化频率明显高于中层次和高层次学习者，这说明随着学习者层次的提高，学习者对于知识的掌握加深，认知水平逐渐提高，出现歧义偏误的频次因此降低。

（二）不同类型歧义偏误分布情况

在前文中，我们将书面歧义偏误分为词汇歧义、语法歧义和语用歧义三个大类，根据这一分类标准，对抽样的歧义偏误进行人工识别并分类，获得的总体偏误分布情况如表 3：

根据表 3 中所示，在第二语言习得过程中，词汇歧义偏误占据了偏误总数的 64%，这是学习者占比最大的歧义偏误类型。语法歧义偏误占据了学习者偏误

总数的 30%，学习者的语法歧义偏误有多种表现，在之后的部分会进行细致的划分和讨论。而语用歧义在不同层次的学习者当中出现的频次相近，并且所占歧义偏误总数的比例仅有 7%，可见学习者在进行书面表达时对于整体的语境和情感的把握比较精确，使得语用歧义偏误的产生较少。

表 3 不同类型偏误分布情况

学习者类型 \ 偏误类型	低层次学习者	中层次学习者	高层次学习者	总计
词汇歧义	64	62	44	170
语法歧义	18	23	41	82
语用歧义	8	5	5	18
总计	90	90	90	270

（三）不同层次学习者歧义偏误具体情况

上一个部分讨论了不同类型的歧义偏误在三个层次学习者中的整体分布情况，那不同的歧义偏误类型在不同层次学习者中的分布是否会有所不同？这里将不同的歧义偏误类型在各层次学习者所占百分比进行统计，详情如表 4：

表 4 不同层次学习偏误分布情况

类别 \ 学习者	低层次学习者	中层次学习者	高层次学习者
词汇歧义	64	62	44
	71.1%	68.9%	48.9%
语法歧义	18	23	41
	20%	25.6%	45.6%
语用歧义	8	5	5
	8.9%	5.5%	5.5%

从表 4 中可以看出，低层次和中层次学习者在词汇方面的歧义偏误要远高于高层次学习者。同时从表中也能看出，在语法歧义偏误上出现了低层次学习者犯错较少而高层次学习者犯错较多的情况。而在语用歧义方面，三个层次的学习者情况相差不大，这说明不论是低层次、中层次还是高层次学习者，都意识到了在用英语行文时要注意上下文语境，选择合适的语句进行叙述。

（四）不同层次学习者语法歧义偏误的具体分布情况

在前面的章节，根据语法歧义偏误的具体成因将其细分为：指代关系歧义、句法功能歧义和成分结构歧义。三个类别的歧义偏误在各层次学习者中分布的具体情况如表 5：

从表 5 中可以看出，三个层次的学习者在语法歧

义偏误的总体频次上有很大的区别。在低层次学习者的语法歧义偏误中,指代关系歧义占比 44.44%,相比其他两个类型的语法歧义出现率较高;中层次学习者在三个类型的语法歧义偏误中没有占比特别凸出的类型,句法功能方面的偏误相比于其他两个类型的歧义偏误占比要偏低;高层次学习者中,成分结构歧义偏误的比例超过 50%,而句法功能歧义偏误的占比较少,仅有不到 20%。

表 5 各层次学习者在语法歧义偏误方面的
具体分布情况

类别 \ 学习者	低层次 学习者	中层次 学习者	高层次 学习者
指代关系歧义	8	9	12
	44.44%	39.13%	29.27%
句法功能歧义	5	6	8
	27.78%	26.09%	19.51%
成分结构歧义	5	8	21
	27.78%	34.78%	51.22%
总计	18	23	41
	100%	100%	100%

三、歧义偏误产生的原因分析及教学建议

导致中国学习者在英语习得过程中产生歧义偏误的原因有很多,从认知水平、对比语言学和语言迁移的角度,本文结合之前章节对其进行原因分析,并提出相应教学建议。

(一) 原因

从认知水平的角度来说,人的认知规律总是从简单向复杂发展,语言习得也是如此。由于低层次学习者语言能力有限且无法构建完整语言认知模型,容易因理解偏差产生歧义偏误;随着层次提升,认知能力增强,偏误频率逐渐降低,但高层学习者会较多尝试复杂语法的输出,从而增加了语法歧义的出现频率。

对比三个不同类型的歧义偏误,词汇歧义偏误在所有的偏误中所占比例最大,通过对语料库中的具体实例进行分析,得出三个原因:一是学习者大多只记忆单词常见义项,而忽略多义性。如“I get over after a few days”中“get over”含“克服、恢复”等义,语境应指“克服”,学习者因只记住了单一义项导致歧义;二是对单词不熟悉:如混淆形似词(如“intuition fee”应为“tuition fee”,表“学费”)和拼写错误(如“heedly”应为“heedfully”,表“仔细地”);

三是进行不当缩写:如“a adv”中“adv”应为“ad”(“advertisement”缩写,表“广告”),由错误缩写致歧义。

(二) 教学建议

基于上文分析发现,学习者的语言输入量、词汇掌握深度以及语境理解能力与偏误的发生密切相关,针对这些原因提出以下教学建议。

1. 培养英语思维

引导学生明晰英汉语言结构差异,避免汉语直译;通过文学作品、新闻、电影等材料,渗透跨文化知识,减少文化差异引发的歧义。

2. 加强语言输入

提供语境丰富的语料及正确表达示范;营造英语交流环境(小组讨论、角色扮演、演讲等),即时纠正偏误并示范,帮助学生强化正确用法。

3. 重视词汇教学

词汇教学中,用同义反义对比、词根词缀法帮助学生辨析词汇;结合例句讲解多义词不同义项。针对高层学习者,侧重语法教学,规范复杂语句结构,减少语法歧义。

四、结语

通过分析中国学习者英语语料库,得出以下结论:一是随学习者层次与学习年限提升,歧义偏误概率降低;二是中低层学习者以词汇歧义为主,高层学习者以语法歧义为主;三是高层语法歧义较多,因其尝试高级语法但语言水平不足,教师需针对不同层次学习者制定精准教学策略。

参考文献:

[1] Lado R. Linguistics across cultures; applied linguistics for language teachers[J]. 1957.
[2] Corder S P. The significance of learner's errors[J]. 1967.
[3] Ellis R. Understanding second language acquisition[M]. Oxford: Oxford university press, 1989.
[4] 段满福. 谈英语一词多义现象 [J]. 内蒙古大学学报 (人文社会科学版), 2004(3): 116-120.
[5] 桂诗春, 杨惠中. 中国学习者英语语料库 [M]. 上海: 上海外语教育出版社, 2003.
[6] 章振邦, 张月祥, 赵美娟. 新编高级英语语法 [M]. 上海: 上海外语教育出版社, 2012.12.
[7] 郭嘉, 杨蕾, 王爽, 等. 母语迁移对二语写作复杂度的影响 [J]. 当代外语研究, 2022(3): 140-148.