

基于语料库的听说综合性任务语言特征研究*

徐鹰 廖天华 韩苏 王亚琪

(华南理工大学, 广州 510641)

摘要:本文以故事复述任务为例,采用语料库方法对影响听说综合性任务表现的语言错误进行描写和诊断分析。根据中国学习者英语语料库(CLEC)言语失误标注原则,本研究对360位考生故事复述任务录音进行转写和标注。结果发现,vp(动词短语)、wd(词语)和sn(句子)是语言错误最为集中的3种特征;这3种特征能有效区分不同水平的考生;wd和sn能有效预测考生故事复述成绩。上述结果对听说综合性任务评分标准开发和考试备考提供重要参考。

关键词:听说综合性任务;故事复述;语言特征;语言错误;中国学习者英语语料库;分数预测模型

中图分类号:H319

文献标识码:A

文章编号:1000-0100(2020)01-0056-8

DOI编码:10.16263/j.cnki.23-1071/h.2020.01.008

Investigating Language Features for the Listening-to-speak Integrated Task: A Corpus-based Approach

Xu Ying Liao Tian-hua Han Su Wang Ya-qi

(South China University of Technology, Guangzhou 510641, China)

By using the story retelling task, this paper aims to describe and diagnose language errors which may impact on the performance of listening-to-speak integrated task following a corpus-based approach. A random sample of 360 test-takers' records of story retelling task were transcribed and coded within the annotation system of Chinese Learner English Corpus (CLEC). Major findings are: (1) vp(verb phrase), wd(word) and sn(sentence) are the most frequent categories of language errors; (2) they can differentiate test-takers at different ability levels; (3) wd and sn can predict test-takers' task scores significantly. Implications are drawn for rubric construction and test preparation for the listening-to-speak integrated task.

Key words: listening-to-speak integrated task; story retelling; language feature; language error; CLEC; score predictive model

1 引言

对学习者的语言错误进行分析不仅能提供学习者习得的证据以及揭示学习者采用的策略(Corder 1967:161),而且能帮助研究者发现第二语言或外语学习的一般规律(何华清 2009:2),因此对于深化外语教学意义重大。错误分析同样对于语言测试有重要参考价值,并且往往结合语料库手段对其进行研究。一个对学习者的错误进行标

注的语料库具有以下用途:(1)开发试题,比如确定考点或设计单项选择题的干扰项;(2)通过揭示区分不同能力等级的语言特征从而开发评分标准;(3)对考试内容效度进行检验(Park 2014:28, Cushing 2017:441)。目前,国外众多学者和考试开发机构纷纷构建附带错误标注的学习者语料库进行语言测试研究,比如剑桥考试中心利用剑桥学习者语料库(Cambridge Learner Corpus)设计和

* 本文系教育部人文社会科学研究青年基金项目“基于语料库的英语听说考试评分标准的研制与效度验证”(17YJC740102)、华南理工大学中央高校基本科研业务费项目重大培育项目“基于语料库的英语语音能力等级量表研制”(2018PY22)和广东省教育科学项目重大课题“计算机智能语言技术在广东省高考英语听说考试中的应用研究”(TJW2013001)的阶段性成果。

作者电子邮箱:xuying@scut.edu.cn(徐鹰)

开发考试;Hawkey 和 Barker(2004)对参加剑桥考试中心组织的 PET, FCE, CAE 和 CPE 考试考生的作文语料进行错误分析,确定不同分数等级的区分性特征,最后构建面向剑桥英语能力系列考试写作部分的统一评分标准。国内采用学习者语料库进行考试设计和效度研究的案例相对比较匮乏(邹绍艳 2016:112)。较有代表性的研究是穆惠峰(2011:66)采用自建语料库对大学英语四级考试完形填空题的内容效度进行验证。

综合性任务被普遍认为能更有效地测出考生的真实语言能力,并且能产生更积极的反拨效应,已经被越来越多大规模、高风险测试(如 TOEFL)所采用(Purpura 2016:190)。借助语料库手段对听说综合性任务进行语言错误分析的实证研究主要有以下两项。为了探讨 TOEFL 口试听说综合性任务所测构念的本质,Brown 等(2005:2)收集 200 位考生的语音样本并进行转写和标注。其中语法错误的标注种类包括 5 种特定类型错误:时态错误、主语为第三人称单数时动词/系动词变化错误、名词复数错误、冠词使用错误、介词错误,以及 1 种总体准确度特征:无错误 T 单位百分比。10% 的考生由 2 位标注人分别进行标注,这 6 种错误的标注信度在 0.71-0.99 之间。方差分析结果表明 5 个不同考生分数等级在每种错误上都有显著差异,高分等级(4 级和 5 级)之间差异明显,低分等级(1 级到 3 级)之间差异不明显;5 种特定类型错误的效应量分别为 0.15、0.08、0.17、0.14 和 0.07,总体准确度的效应量值最大(0.22)。根据 Muijs(2004:195)的标准,虽然这 6 种错误特征能区分不同能力考生,但是对考生分数差异的解释力度较弱或有限。王华等(2018:413)通过对 110 位参加某大型英语听说考试的考生故事复述任务口语产出录音进行标注,探讨听说综合性任务的口语能力区分性特征。其中在语言准确性方面通过标注考生动词过去式使用错误率来表征具体错误,选用无错误小句比(error-free clauses ratio)判断总体准确性。多元回归分析结果显示无错误小句比能显著预测考生口语能力($p < 0.05$)。因此,听说综合性任务的口语能力构念应该包括语言准确性。

尽管这两项研究对影响听说综合性任务的语言因素进行过分析,但是对语言错误的描写则不够全面。作为一种常用的听说综合性任务,故事复述已经应用在全国英语专业四级(TEM-4)口试和广东省高考英语听说考试中。这两种考试采用的评分标准都包括语音语调、复述内容以及语法和词汇等维度,其中语法和词汇维度都把语言错

误作为评判的主要依据之一。那么哪些语言错误能有效区分不同能力考生呢?这是本文要解决的核心问题。鉴于此,本研究拟采用基于语料库的方法对影响故事复述任务表现的语言错误进行全面描写和诊断,从而回答以下研究问题:

- (1)中国学习者故事复述任务口语产出的语言错误有什么特征;
- (2)不同能力考生的语言错误是否有差异;
- (3)哪些语言错误能够有效预测考生的故事复述分数。

2 研究设计

2.1 研究工具

广东省高考英语听说考试总分为 60 分,包括模仿朗读、角色扮演和故事复述 3 个任务。故事复述任务要求考生先听一段大约两分钟的独白(通常为记叙文)两遍,然后复述所听的内容。考生应尽可能使用自己的语言复述,而且复述内容应涵盖尽可能多的原文信息。

本研究的语料来自 2013 年广东省高考英语听说考试 A 卷考生录音。故事复述任务总分 24 分。评分标准包括内容和综合两部分。内容部分按信息点评分,一共包括 10 个信息点,每个 1.5 分,总分 15 分。综合部分总分 9 分,要求对考生的总体表现进行评分,具体包括 3 个档次:(1)如果考生口语产出忠实原文内容,语言得体性好、合乎规范,表达流利,且语音语调不影响理解,应该给 7-9 分;(2)如果考生基本复述原文的主要内容,语言得体性较好、基本合乎规范,比较流利,尽管出现一些语音语调的错误,但基本不影响理解,应该给 4-6 分;(3)如果考生不能复述原文的主要内容,语言得体性较差,出现较多的语言错误,不够流利,且语音语调影响理解,应该给 0-3 分。A 卷故事复述任务文本见图 1。

2.2 研究对象

本研究通过分层随机抽样的方法选取 360 位考生录音。首先,把考生按照考试总分排序,然后平均分成高分组($M = 46.93, SD = 5.50$)、中分组($M = 37.07, SD = 6.75$)和低分组($M = 24.10, SD = 9.15$),每组 120 位考生。单因子方差分析发现 3 组考生总分组间均值差异显著 [$F(2, 357) = 295.867, p < 0.001, \eta^2 = 0.624$]。Scheffe 事后检验表明,相邻组别的总分均有显著差异($p < 0.001$)。按照上述方法随机抽取 30 份考生录音用于培训,3 个水平组各 10 位考生。

The Postman

Tom was a postman. Every time he delivered a letter to Mrs. Brown, he had to finish his work quite late because she lived far away. However, Tom did not mind this, for Mrs. Brown always kindly invited him in for a drink. When Tom entered the gate of Mrs. Brown's house yesterday, he didn't see her working in the garden. He was surprised, because she usually spent afternoons there when the weather was fine. Tom thought she might be in the kitchen. Then he went to the back of the house. But the door was locked. So he returned to the front of the house and knocked hard on the door. There was no answer. Tom thought that this was very strange because he knew Mrs. Brown seldom left the house. Just then he noticed that her milk, delivered early in the morning, was still at the door. If Mrs. Brown had not taken in her milk, perhaps she was ill. Tom walked around the house and found an open window. He went inside and found Mrs. Brown was lying on the floor of the living room. Tom realized that there was little he could do. He called for an ambulance.

图₁ 2013年广东省高考英语听说考试A卷故事复述文本

2.3 研究过程

考生语料的处理包括转写和编码两个阶段。两位研究助手(拥有文学硕士学位)逐词将考生录音转写成文字。转写后的总体和3组考生录音文本单词量的描述性统计如下: $M_{\text{总体}} = 62.71$, $SD_{\text{总体}} = 23.39$; $M_{\text{高分组}} = 79.04$, $SD_{\text{高分组}} = 15.79$; $M_{\text{中分组}} = 61.53$, $SD_{\text{中分组}} = 17.86$; $M_{\text{低分组}} = 37.80$, $SD_{\text{低分组}} = 19.45$ 。单因子方差分析发现3组单词量组间均值差异显著 [$F(2, 357) = 162.964, p < 0.001, \eta^2 = 0.477$]。Scheffe 事后检验表明,相邻组别的单词量均有显著差异($p < 0.001$)。

转写好的文本参照《中国学习者英语语料库》(CLEC)(桂诗春 杨惠中 2003:1)的言语失误

标注原则进行编码。该标注原则一共包括11大类(词形 fm, 动词短语 vp, 名词短语 np, 代词 pr, 形容词短语 aj, 副词 ad, 介词短语 pp, 连词 cj, 词语 wd, 搭配 cc, 句子 sn)和61小类,能够较全面地描写中国学习者英语产出的错误。通过对用于培训的30位考生语音转写文本进行分析,我们发现词语 wd 分类较多(7种),其中一些类型未在转写文本中出现(如 wd1(词序)、wd6(重复)),而某些常见错误(如表₁中 wd1 例子 hardly 在语料库中出现19次)又难以归为现有的类型。因此,为便于标注和分析,把 wd 的标注分类数量从7减少为2,并且重新进行定义,具体见表₁。

表₁ 词语类型错误(wd)的定义和分类

码	分类	类别	说明	例子
wd1	word	词汇	误用意义不相关词(用了错误的词,且错误的词和正确的词没有意义关系。)	So he returned to the front of the house and knocked hardly [wd1] on the door. wd1 是词用错了,正确的单词是 hard.
wd2	word	词汇	误用意义相关词(用了错误的词,但错误的词和正确的词意义相关,如同根词。)	He was surprising [wd2], because she usually spent afternoons there when the weather was fine. wd2 是词用错了,考生混淆了 surprising 和 surprised.

修改后的标注框架一共包括10大类(删除词形 fm)和53小类(删除词形 fm 的3小类,词语 wd 的7小类减少到2)。两位研究助手采用该框架对全部考生语料进行标注。在标注培训时,他们首先学习该标注体系,并共同讨论用于培训的15份考生样本,然后独立对另外15份考生样本进行标注,接着相互查看各自标注文本并讨论其中差异,最后达成共识。培训结束后,两位研究助手分别对185位考生录音文本进行标注(其中10

位是共同考生)。由于标注框架包含的语言错误种类非常全面,因此不同标注者的信度难以达到较高水平(Brown et al. 2005:51)。在本研究中,这10位考生一共产出244个错误,两位标注者相同的错误数量为196个(80.33%),说明标注结果可信。

2.4 数据分析

首先采用 AntConc 3.2.4w 统计各类特征频数。由于不同水平考生口语产出的单词量存在显

著差异,为做进一步的推断统计,将所有特征频数转化成每产出100个单词出现该错误的标准频数。然后,采用SPSS 18.0对每个特征进行描述性统计分析,并对可能的区分性特征进行单因素方差分析。最后,把所有考生按照分数排序并平均分为奇数样本(训练集)和偶数样本(测试集),采用多元线性回归分析奇数样本得到分数预测模型,并用偶数样本验证效度。

3 结果

3.1 中国学习者故事复述任务口语产出的语言错误特征

3.1.1 大类语言错误特征

全部3组考生产生的语言错误原始频数总数为7,828个,具体见表₂。

表₂ 各组语言错误大类特征原始频数

分组	大类特征										汇总
	vp	np	pr	aj	ad	pp	wd	cj	Cc	sn	
低分组	758	223	54	5	8	74	375	0	15	555	2067
中分组	1195	211	122	4	10	83	612	1	2	619	2859
高分组	1289	250	92	2	17	115	546	0	23	568	2902
总体	3242	684	268	11	35	272	1533	1	40	1742	7828
百分比(%)	44.42	8.61	3.17	0.07	0.59	3.96	18.81	0.00	0.79	19.57	100.00
零错误考生数	3	95	227	349	328	180	33	359	328	9	

在大类特征中, vp, wd 和 sn 是语言错误最为集中的3种,总和达到82.8%,且零错误考生数相对较少(最多不超过10%)。其他7种特征相对较少,其中4种特征(aj, ad, cj 和 cc)百分比不足1%,且零错误考生数相对较多。这一结果说

明 vp, wd 和 sn 是不同水平考生都会出错的特征,具有普遍性,可能是区分性特征,因此需要对它们做进一步的分析。

经过标准化处理后的 vp, wd 和 sn 这3种大类特征标准频数的描述性统计结果见表₃。

表₃ 各组 vp, wd 和 sn 标准频数描述性统计结果

指标	分组	平均值	标准差	最大值	最小值
vp	高分组	14.02	6.47	34.88	1.87
	中分组	17.01	7.44	44.44	1.19
	低分组	18.85	12.64	91.67	0.00
	总体	16.63	9.44	91.67	0.00
wd	高分组	6.03	4.70	17.19	0.00
	中分组	8.61	6.38	31.03	0.00
	低分组	8.15	6.21	25.93	0.00
	总体	8.00	5.90	31.03	0.00
sn	高分组	6.51	4.93	24.49	0.00
	中分组	9.24	6.22	28.57	0.00
	低分组	16.45	16.03	100.00	0.00
	总体	10.74	11.12	100.00	0.00

总体而言, vp, wd 和 sn 这3种特征的标准频数都随着考生水平增加而降低。值得注意的是中分组的 wd 错误平均值大于低分组。根据 Ellis (1999:76)提出的中介语理论,中介语的特征之一体现为可预测性和内部一致性。一些学习者将

这一规则泛化,比如在 hard 后面加上 -ly 后缀构成副词。因此,造成这一现象的可能原因在于中分组考生尚未完全掌握某些目的语规则,把“形容词加后缀 -ly 变成副词”的规则过度使用,忽略不适用这个规则的特殊情况。此外,学习者的中

介语还具有动态性和开放性,因此中分组考生可能在使用某些目的语规则时出现前后不一致的情况,比如混用动词的过去分词和现在分词,因此会出现中分组 wd 错误数多于低分组。

3.12 小类语言错误特征

对于 vp 和 sn 而言,高分组的错误标准频数平均值和标准差最小,低分组错误标准频数平均值和标准差最大,说明低分组的内部差异较大。vp 错误特征的原始频数是 3,242,其中 vp6(时态)频数为 2640(81.43%)。在 9 种 vp 小类错误特征中,只有 vp6(时态)出现 5 个考生没有犯错误,其他 8 种小类错误的零错误考生数量至少超过 180 人(50%)。以上结果说明 vp6(时态)可能是区分不同能力考生的小类特征。

sn 错误特征的原始频数是 1,742,其中 4 种小类错误(sn3(垂悬修饰语)、sn4(比较不符合逻辑)、sn5(主题突出)和 sn6(并列))的原始频数为 0。sn8(结构缺陷)频数为 1,414(81.17%),只有 35 位考生没有犯这一错误。sn1(不断句)、sn2(片段)、sn3(主从)错误频数和比例分别为 104(5.97%)、202(11.60%)和 22(1.26%),且没有犯这 3 种错误的考生人数分别为 276、235、341。上述结果说明 sn8(结构缺陷错误)可能是区分不同能力考生的特征。根据 CLEC 言语失误标注原则,结构缺陷主要指一个句子中语法结构的错误,包括 improper splitting(不合理分隔)、pattern shifting(句式转移)和 confusing structure(结构混乱)等主要类型。考生常犯的 sn8 错误如下:

① So he open [vp6] the door run [vp6] into the living room. [sn8]

② She knocked on the door hardly [wd1]. The door has [vp6] no answer. [sn8]

③ Tom was very strange [sn8], because she seldom go [vp6] away.

对于 wd 而言,虽然高分组的错误标准频数平均值和标准差最小,但是中分组错误标准频数平均值和标准差最大,且中分组和低分组的差别不明显。这个结果说明中国学习者中介语发展过程不是直线上升,而是呈螺旋式上升的态势(崔艳嫣 王同顺 2005:7)。wd1(误用意义不相关词)和 wd2(误用意义相关词)都是容易混淆的词语类型错误,考生常犯的 wd1 错误如下:

④ She must be in the chicken [wd1].

⑤ a lady lying in the flower [wd1].

⑥ Brown was lying on the door [wd1].

考生常犯的 wd2 错误如下:

⑦ He deliver [vp6] a letter to Miss [wd2] Brown.

⑧ a glasses [np6] of milk was broke [wd2] on the floor.

⑨ Mrs. Brown was so worry [wd2] and he could do little. . .

这两种错误的差别在于错误的词是否和正确的词有意义上的联系。wd 错误特征的原始频数是 1,533,其中 wd1 和 wd2 频数分别为 243(15.85%)和 1,290(84.15%),且分别有 192 和 60 位考生没有出现 wd1 和 wd2 错误。因此,wd2 错误可能是区分不同能力考生的特征。

经过标准化处理后的 vp6、wd2 和 sn8 这 3 种小类特征标准频数的描述性统计结果见表₄。

表₄ 各组 vp6、wd2 和 sn8 标准频数描述性统计结果

指标	分组	平均值	标准差	最大值	最小值
vp6	高分组	11.33	5.87	28.38	0.00
	中分组	13.97	6.82	38.89	1.19
	低分组	15.57	11.20	75.00	0.00
	总体	13.62	8.45	75.00	0.00
wd2	高分组	5.18	4.51	17.07	0.00
	中分组	7.25	5.70	27.78	0.00
	低分组	6.48	5.87	25.00	0.00
	总体	6.30	5.45	27.78	0.00
sn8	高分组	5.18	5.14	24.49	0.00
	中分组	7.89	6.58	26.09	0.00
	低分组	14.28	16.65	100.00	0.00
	总体	9.12	11.38	100.00	0.00

3.2 不同能力考生的语言错误特征的差异

在删除异常数据后,以考生水平为自变量,分别以 vp, wd 和 sn 标准频数为因变量进行 3 次单因素方差分析,结果发现这 3 个特征都能区分不同水平考生:

Fvp 标准频数(2,351) = 7.537, p < 0.001, η² = 0.041;

Fwd 标准频数(2,354) = 5.742, p < 0.01, η² = 0.031;

Fsn 语调分数(2,332) = 37.274, p < 0.001, η² = 0.183。

Scheffe 事后多重比较表明, vp 和 wd 错误能有效区别高分和中、低分考生,但是无法区别中分和低分考生;sn 错误能有效区别高分、中分、和低分考生。

由于 vp6、wd2 和 sn8 可能是区分不同能力考生的小类特征,我们继续以 vp6、wd2 和 sn8 标准频数为因变量,以考生水平为自变量进行单因素方差分析。在删除异常数据后,发现这 3 个特征都能区分不同水平考生:

Fvp6 标准频数(2,351) = 7.895, p < 0.001, η² = 0.043;

Fwd2 标准频数(2,354) = 4.030, p < 0.05, η² = 0.022;

Fsn8 语调分数(2,332) = 32.282, p < 0.001, η² = 0.165。

Scheffe 事后多重比较表明, vp6 和 wd2 错误能有效区别高分和中、低分考生,但是无法区别中分和低分考生;sn8 错误能有效区别高分、中分、和低分考生。

上述结果证实 Brown 等(2005:82)的研究结论:时态错误能够区分不同水平考生,且效应量不大。此外,sn8 错误能有效区别高分、中分和低分考生,其原因在于 sn8(结构缺陷错误)主要是对句子的总体准确性进行判断,在一定程度上和 Brown 等(2005:52)的无错误 T 单位百分比以及王华等(2018:416)无错误小句比的定义类似,因此也能够区分不同水平考生。

3.3 故事复述任务分数预测模型

首先采用独立样本 t 检验对奇数样本和偶数样本考生成绩进行分析,结果显示两组考生分数无显著差异(t = -.153, df = 358, p > 0.05)。然后采用 Mahalanobis 距离整理奇数样本发现 23 个异常值,删除后剩下 157 位考生。最后以 vp, wd 和 sn 标准频数作为预测变量,考生分数作为因变量采用逐步进入法建立回归方程,结果得到两个分数预测模型(表₅)。

如表₅所示,两个模型都能预测考生成绩。模型 1 的预测变量只有一个,因此容忍度和 VIF 值都为 1。模型 2 的两个预测变量的容忍度都是 0.979, VIF 值都是 1.021。上述结果说明两个模型都不存在共线性问题。模型 1(考生故事复述成绩(24 分制) = 17.092 - 0.460 × sn 标准频数)能解释考生分数 23.2% 的方差;模型 2(考生故事复述成绩(24 分制) = 18.390 - 0.435 × sn 标准频数 - 0.217 × wd 标准频数)能解释考生分数 27.2% 的方差,拟合优度更高。因此,模型 2 是较好模型。在 10 大类言语失误中,句子 sn 对考生成绩的预测能力最强,其次是词语 wd,其他特征无法预测考生分数。

表₅ 多元线性回归结果

编号	预测变量	R	R ²	Adjusted R ²	F	B	Beta	t
模型 1	常数					17.092		24.752***
	sn	0.487	0.237	0.232	48.234***	-.460	-.487	-6.945***
模型 2	常数					18.390		22.200***
	sn					-.435	-.460	-6.620***
	wd	0.522	0.272	0.263	28.802***	-.217	-.189	-2.717**

注: * p < 0.05, ** p < 0.01, *** p < 0.001

第二步验证分数预测模型。首先根据模型 2 计算出偶数样本的考生成绩,然后和考生的故事复述任务高考分数进行配对样本 T 检验。结果显示,模型 2 预测分数和考生高考分数相关系数

达到 0.826(p < 0.001),且没有显著差异(t = -0.559, df = 179, p > 0.05)。因此,模型 2 能有效预测考生的故事复述成绩。

4 讨论

本研究发现 vp6(时态)、wd2(误用意义相关词)和 sn8(结构缺陷)能有效区分不同水平的考生,主要原因如下。

首先, vp6(时态)错误可能是受到母语(汉语)迁移的影响。语言迁移是一种普遍存在的现象,是指学习者掌握的多种语言之间的相互影响(蔡金亭 2015:57)。在学习者第二语言未达到自动化之前,无法随着时间的推移和学习者水平的提高而完全消失。根据标记性和母语迁移的关系原则(Ellis 1994:206),母语中的无标记项在二语中的对应项如果有标记,那么母语的无标记特征会迁移到中介语中。由于汉语缺乏时态系统(即无标记性),但是英语具有时态系统(即有标记性),那么汉语的无标记特征会迁移到中介语当中,即学习者的中介语也会出现没有时态的情况。此外,由于考生在考试时普遍存在紧张、焦虑的情绪,这种语际影响表现得更为明显。

其次是 wd2(误用意义相关词)错误,包括两种类型:(1)误用正确单词的同根词,导致词性错误,例⑧和例⑨都是这种情况,其主要原因也是受母语迁移的影响。Odlin(1989:79)指出,当两种语言的词汇在语义上对等而形态上不具相似性时往往发生词汇层面的迁移。Larsen - Freeman 和 Long(2000:107)进一步指出:母语中无标记的语言特征比有标记的特征更容易发生迁移。由于汉语的词性没有明显标记,因此考生容易出现误用同根词的情况。(2)大量考生(频数为316)把 Mrs. Brown 错听成 Miss Brown(例⑦),可能原因是部分考生过于紧张,且 Mrs 和 Miss 发音和意义都较为接近,容易出现混淆。

最后是 sn8(结构缺陷)错误。例①是一个典型的连动句或连动结构(serial verb construction, 简称 SVC),指的是充当单一述语、并且不带有任何并列、从属或其他句法依附性显性标记的动词序列(尚新 2009:5),其特点是连动短语中的每个动词项都与一个主语联系。这种结构在汉语中广泛存在,但在英语中普遍缺失(郑学丹 2018:454)。根据标记性理论(Eckman 1977:315),可以认为这种限定性动词和非限定性动词的区别对汉语而言是一种无标记性,对英语则是一种标记性,因此会产生母语负迁移。例②和例③则是一种典型的词对词的翻译错误,即考生没有掌握英语句子的表达方式,只是按照汉语语序造句,也可以看成是一种母语迁移产生的错误。

5 结束语

通过构建一小型语料库并采用 CLEC 言语失误标注原则进行标注,本研究对影响故事复述任务表现的语言错误进行分析,结果发现 vp(动词短语)、wd(词语)和 sn(句子)是语言错误最为集中的3种大类特征, vp6(时态)、wd2(误用意义相关词)和 sn8(结构缺陷)是语言错误最为集中的3种小类特征,它们都能区分不同水平考生。本文的研究成果主要有以下意义。

首先,有助于开发基于数据的评分标准(徐鹰 2018:75)。TEM-4 口试和广东省高考英语听说考试故事复述任务采用的都是基于专家直觉的评分标准。尽管评分细则中提到语言错误,但是较为宽泛,既没有指明语言错误包括哪些错误类型,也没有界定多少语言错误算少或算多、或什么样的语言错误算严重等问题。这就导致评分员只能凭借自身的经验去解读和使用评分标准,容易导致评分员误差,最终威胁考试的效度。本研究发现的3种大类错误和3种小类错误能有效区分不同水平考生,且不同类型错误的效应量有差异,这就为充实评分标准的语言维度的等级描述语、明确语言错误的严重程度提供了重要参考,从而帮助评分员形成对评分标准统一的解读,进而为分数意义提供合理的解释。同时,本研究发现上述错误特征能区分不同能力考生,为评分标准语言维度各分数等级之间的线性关系提供实证数据的支撑,从而增强评分标准的效度。

其次,有助于帮助师生明确备考的重点。在备考故事复述这一题型时,也许考生会犯各种各样的语言错误。但是,他们不应对所有类型的语言错误等量齐观,而应有所侧重,重点解决 sn8、vp6 和 wd2 3 种类型错误。同时,教师应该在备考辅导中放弃“全面纠错”的范式(Chandler 2003:267)以求对考生所有的错误提供反馈,而应采用“选择性纠错”的方式(Sheen 2007:255),重点帮助考生改正上述3种类型错误,从而实现备考效果的最大化。

本研究的不足之处在于只采用一个任务的数据,结论的可推广性和概括性还有待增强。下一步的研究可以采用不同的任务进行验证。

参考文献

- 蔡金亭. 在二语产出中判断母语迁移的比较——归纳方法框架[J]. 解放军外国语学院学报, 2015(5). ||
Cai, J. -T. Comparison-induction Methodological Framework for Judging L1 Transfer in L2 Production [J].

- Journal of PLA University of Foreign Languages*, 2015 (5).
- 崔艳嫣 王同顺. 基于 CLEC 的中国学习者英语派生词缀习得分析[J]. 国外外语教学, 2005(2). || Cui, Y.-Y., Wang, T.-S. A CLEC-based Study of Chinese EFL Learners' Acquisition of Derivational Affixes [J]. *Foreign Language Teaching Abroad*, 2005(2).
- 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. || Gui, S.-C., Yang, H.-Z. *Chinese Learner English Corpus* [M]. Shanghai: Shanghai Foreign Language Education Press, 2003.
- 何华清. 非英语专业学生写作中的词汇错误分析——一项基于语料库的研究[J]. 外语界, 2009(3). || He, H.-Q. An Analysis of Lexical Errors in Non-English Major's Writing: A Corpus-based Study[J]. *Journal of The Foreign Language World*, 2009(3).
- 穆惠峰. 基于语料库的大学英语四级完型填空测试内容效度验证研究[J]. 外语电化教学, 2011(4). || Mu, H.-F. Corpus-based Research on Content Validity of CET4 Cloze [J]. *Technology Enhanced Foreign Language Education*, 2011(4).
- 尚新. 时体、事件与汉语连动结构[J]. 外语教学, 2009(6). || Shang, X. Aspect, Event and Chinese Serial Verb Constructions [J]. *Foreign Language Education*, 2009(6).
- 王华 范婷婷 曾用强. 听说综合性任务下口语能力本质研究[J]. 现代外语, 2018(3). || Wang, H., Fan, T.-T., Zeng, Y.-Q. Investigating the Construct of Speaking Proficiency under the Listening-to-speak Integrated Task [J]. *Modern Foreign Languages*, 2018(3).
- 徐鹰. 语言做事测试评分标准的开发和效度验证: 回顾与反思[J]. 外语教育研究前沿, 2018(2). || Xu, Y. Developing and Validating Rating Scales in Language Performance Assessments: Retrospect and Reflection [J]. *Foreign Language Education in China*, 2018(2).
- 郑学丹. 连动句的时态解读机制[J]. 现代外语, 2018(4). || Zheng, X.-D. The Tense of Serial Verb Constructions [J]. *Modern Foreign Languages*, 2018(4).
- 邹绍艳. 语料库在语言测试中的应用: 回顾与反思[J]. 中国海洋大学学报(社会科学版), 2016(6). || Zou, S.-Y. The Application of Corpora in Language Assessment—Review and Reflection [J]. *Journal of Ocean University of China (Social Sciences)*, 2016(6).
- Brown, A., Iwashita, N., McNamara, T. An Examination of Rater Orientations and Test-taker Performance on English-for-Academic-Purposes Speaking Tasks (TOEFL Monograph Series Report RR-05-05) [R]. Princeton: ETS, 2005.
- Chandler, J. The Efficacy of Various Kinds of Error Feedback for Improvement in the Accuracy and Fluency of L2 Student Writing [J]. *Journal of Second Language Writing*, 2003(3).
- Corder, S. P. The Significance of Learners' Errors [J]. *International Review of Applied Linguistics*, 1967(4).
- Cushing, S. T. Corpus Linguistics in Language Testing Research [J]. *Language Testing*, 2017(4).
- Eckman, F. Markedness and the Contrastive Analysis Hypothesis [J]. *Language Learning*, 1977(2).
- Ellis, R. *The Study of Second Language Acquisition* [M]. Oxford: Oxford University Press, 1994.
- Ellis, R. *Understanding Second Language Acquisition* [M]. Shanghai: Shanghai Foreign Language Education Press, 1999.
- Hawkey, R., Barker, F. Developing a Common Scale for the Assessment of Writing [J]. *Assessing Writing*, 2004(2).
- Larsen-Freeman, D., Long, M. *An Introduction to Second Language Acquisition Research* [M]. Beijing: Foreign Language Teaching and Research Press, 2000.
- Muijs, D. *Doing Quantitative Research in Education with SPSS* [M]. London: Sage, 2004.
- Odlin, T. *Language Transfer: Crosslinguistic Influence in Language Learning* [M]. Shanghai: Shanghai Foreign Language Education Press, 1989.
- Park, K. Corpora and Language Assessment: The State of the Art [J]. *Language Assessment Quarterly*, 2004(1).
- Purpura, J. Second and Foreign Language Assessment [J]. *The Modern Language Journal*, 2016(1).
- Sheen, Y. The Effect of Focused Written Corrective Feedback and Language Aptitude on ESL Learners' Acquisition of Articles [J]. *TESOL Quarterly*, 2007(2).

定稿日期: 2019-12-10

【责任编辑 陈庆斌】