

## Self- built Small Corpus Based Research on Senior Two School Students' English Writing

Xiaowe Li<sup>1</sup>, Yan Zhuang<sup>2</sup>

<sup>1</sup>School of Foreign Languages, Southwest Petroleum University, Chengdu 610500, China;

<sup>2</sup>State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Chengdu 610500, China.

**Abstract.** Based on self- built small corpus, this paper makes English writing error analysis on senior two school students, and classifies the errors into: differences between Chinese and English language, L1 transfer, overgeneralization, internal control factors and evaluation system. I hope this article can enlighten some English writing teaching method and can enhance students' self- awareness and self-correction, so as to improve their English writing.

**Keywords:** Self built corpus; English writing; error analysis; senior two school students.

## 基于自建小型语料库的高二英语写作研究

李晓微<sup>1</sup>, 庄严<sup>2</sup>

1. 西南石油大学外国语学院, 成都 中国;

2. 西南石油大学石油与天然气工程学院, 成都 中国

**摘要:** 本文基于自建小型学习者语料库对高二学生英语写作进行错误分析, 将错误归因为: 英汉语言差异、母语迁移、过度概括、内控因素和评价体系。本文希望能够对教师英语写作教学方法有所启示, 使学生能自我意识和自我改正、学习、提高英语写作。

**关键词:** 自建语料库; 英语写作; 错误分析; 高二学生

### 1. 研究背景

英语写作能力是衡量英语输出效果的重要指标, 长期以来英语写作研究受到广泛关注。英语写作课程、兴趣小组、各种比赛等用于推进英语写作的同时, 随着互联网飞速发展, 大数据、云计算等科技日益影响人们的生活、学习。学生英语学习方式日益多样化, 如: 使用句酷批改网写作文、百词斩背单词, 利用国内外各大网站资源查询英语相关学习内容。英语写作的形式不断多样化, 但学生的写作能力还未达到预期目标。

首先, 笔者分别以“英语写作”“高中英语写作”“基于语料库的高二英语写作研究”为关键词, 检索了 2000 年至 2016 年间 CNKI 所刊载的英语写作类文献, 发现: 1) 学生英语写作研究整体呈上升趋势, 但是研究主体主要涉及大学生英语专业及非英语专业, 涉及高三英语写作研究较少; 2) 高中英语写作研究内容围绕分析学生写作中的问题和错误, 提出相应的对策; 以及英语写作教学反思和建议这两大主题; 3) 高中英语写作研究方法多样, 如: 范文模方法, 任务型教学法, 过程体裁教学法, “四位一体”教学法, 图式理论, 合作学习, 篇章理论, 词块/语块教学, 同伴反馈和教师反馈等。但基于语料库的高三英语写作研究鲜有涉及。分析还表明, 高二学生英语写作研究鲜有涉及, 高中英语写作亦存在重复研究、研究失衡、教学实证性研究比例偏低等问题。高二学生处于英语基础知识积累的重要阶段, 其英语写作规范的养成以及英语习惯表达更应该受到密切关注。本文自建小型语料库, 分析高二学生英语写作出现的错误, 并对错误进行归因, 并给出指导意见, 旨在对教学有所启发, 提高学生的写作水平。

## 2. 理论基础

### 2.1. 错误分析

C. Frices 和 R. Lado 等一些美国语言学家和语言学家提出了对比分析理论 (contrastive analysis)。他们认为：通过对学习者的母语 (mother tongue) 和目的语 (target language) 对比研究，可以预见学习者在学习该目的语中可能会犯的错误，但是局限性也逐渐的暴露了出来。未来对学习者的错误进行系统性的分析研究，以确定其错误的来源，为教学过程中消除这些错误提供依据，一种分析语言错误的新理论“错误分析理论 (error analysis)”产生了。外语学习者语言的错误并不意味着语言学习的失败，有些错误是不可避免的，这样经过分析，可以了解学习者的语言学习状况，找寻错误的来源，进而有意识进行纠正，提高外语学习水平。

### 2.2. 过程评价

从 20 世纪 70 年代起，瑞典人马顿 (Marton) 与塞里欧 (Saljo) 开始研究学生的学习方式。在他们之后，还有英国人恩特韦斯特尔 (Entwistle) 和澳大利亚人比格斯 (Biggs) 的相关研究。比格斯通过研究发现，学生的学习方式 (learning approach) 与学习动机、学习策略紧密相关：学习动机决定学习的总体方向，而学习策略则帮助实现学习的总体目标。进一步研究发现，学生通常表现出来的学习方式主要有三种，即：表层式学习方式 (surface approach)、深层式学习方式 (deep approach) 和成就式学习方式 (achieving approach) (Biggs J B, Telfer B: 149)。过程性评价有两个重要特征：其一，关注学习过程；其二，重视非预期结果。过程性评价的分类：（一）依据评价主体划分，可以将过程性评价分为学生自评、同学互评和教师点评三类；（二）依据评价层次来划分，可以分为教师对小组的评价和小组对个人的评价；（三）依据评价的规范程度来划分，可以分为程序式评价与随机式评价；（四）依据具体的评价方式来划分，可以有轶事记录、课堂观察、成长记录、个别交流、态度调查、辩论演讲、作文比赛、模型制作等。

### 2.3. 归因理论

人力资源管理和社会心理学的激励理论之一，归因是指观察者为了预测和评价被观察者的行为，对环境加以控制和对行为加以激励或控制，而对被观察者的行为过程所进行的因果解释和推论。美国心理学家伯纳德·韦纳 (B. Weiner, 1974) 认为，人们对行为成败原因的分析可归纳为以下六个原因：能力，努力，任务难度，运气，身心状态，其他因素。其中，能力、努力和身心状况属于内控因素，其它属于外控因素。韦纳认为：1、人的个性差异和成败经验等影响着他的归因。2、人对前次成就的归因将会影响到他对下一次成就行为的期望、情绪和努力程度等。3、个人的期望、情绪和努力程度对成就行为有很大的影响。

## 3. 研究设计

语料库通常指为语言研究收集的、用电子形式保存的语言材料，由自然出现的书面语或口语的样本汇集而成，用来代表特定的语言或语言变体 (杨惠中, 2004)。语料库可以帮助人们观察和把握语言事实，分析和研究语言系统的规律。从语料库语言学的发展历程可以看出，语料库的用途主要体现在对语言现象以及语言本质的研究，如：语言统计、词典编纂、词汇搭配和自然语言处理等方面 (顾曰国, 1988)。现在语料库也开始逐渐应用到语言教学研究中。随着语料库语言学的发展，学习者语料库的建设已经成为了当今语料库语言学研究的重点之一。大型语料库的出现和发展在研究语言的用法、翻译、教学、教材编写和词典编纂等方面发挥重要的作用，但正因为其内容包罗万象，在研究基础英语写作方面缺乏针对性，实用性不强。本文将建立针对高二学生写作的语料库来研究高二学生的英语写作，以提高其总体水平。本文随机选取四川省成都市某高中二年级第二学期第二次月考作文 100 份为原始语料，建立高二学生英语写作小型语料库。

### 3.1. 研究问题

针对所收集的语料库特点并参考相关文献, 本文希望回答以下几点问题:

(1) 怎样自建小型语料库? 又如何进行语料库分析、检索?

(2) 从自建小型四川省某高中高二年级第二学期第二次月考作文语料库中, 可以知道高二学生的英语写作常见错误有哪些?

(3) 这些错误大致是由什么原因导致的? 应如何避免?

### 3.2. 分析工具

(1) 本论文采用 Antconc3.2.4w 软件分析。Antconc 是由日本早稻田大学科技学院 Laurence Anthony 编写的一款绿色, 跨平台语料处理软件。该软件具有索引, 词表生成, 主题词计算, 搭配和词簇提取等多功能。该软件为免费软件, 几乎所有的语言都可以用 Antconc 来进行处理。

(2) 本文采用网络在线赋码器 (Free CLAWS WWW Trial Service)。CLAWS 主要有兰卡斯特大学计算机语料库研究中心 UCREL (University center for Computer Corpus Research on Language) 研制开发, 该词性赋码器正确率比较高, 在 96%-97% 之间。CLAWS 为商用软件, 其码集含有近 140 种标注码 (Baker, etl. 2006:45-46, 161)。

网页链接: <http://ucrel.lancs.ac.uk/claws/trial.html>.

### 3.3. 语料库建立

#### 3.3.1 搜集和标注语料

笔者从四川某高中高二学生第二学期第二次月考中, 随机抽取 100 份试卷, 将试卷分两部分, 由两人录入电脑, 之后相互审查, 以确保语料真实完整录入。将每位考生的电子作文单独保存为 word 文档, 存放在“原始语料”文件夹, 再将其另存为 UTF-8 文件编码纯文档格式, 存放在“编码语料”文件夹, 之后使用网络在线赋码器 (Free CLAWS WWW Trial Service) 进行赋码, 存放在“已标注语料库”, 以供计算机处理。试卷内容如下:

#### 书面表达 (满分25分)

假如你是李华, 你的朋友 Johnson 沉溺于网游 (online games) 一段时间了, 近来他想戒掉网瘾, 写信向你求助。请根据以下提示给 Johnson 写回信。

要点: 1、制定计划, 并努力执行;

2、培养其他爱好, 以便转移注意力;

3、取得进步, 则自我奖励;

4、陷入困境, 可寻求帮助。

词数: 80 左右

Dear Johnson,

Glad to hear you have decided to give up online games, \_\_\_\_\_

\_\_\_\_\_

I'm sure you can break away from the bad habit.

Yours,

Li Hua

#### 3.3.2 处理语料

下载安装 Antconc3.2.4w, 打开主界面, 在文件中打开“已标注语料库”文件夹。

#### 3.3.3 统计与分析

下载使用 WordCount 对“已标注文件夹”进行批量统计, 发现字数为 18076, 随机抽取的 100 篇文章中, 有 3 篇是空白试卷, 其余 97 份中作文的开头和结尾已给出, 因此这些情况不计入分析范围内。

## 1. 用词分析

本篇作文为任务型作文，目的是帮助好友 Johnson 戒掉网瘾，根据提示要点写作文。搜索 2006-2016 近 10 年的四川省高考英语作文范文，笔者发现作文类型均为书信格式的半任务型写作，内容比较贴近生活，旨在考查学生用英语解决问题的能力。笔者在 Antconc 中 Concordance 搜索界面，在 search term 中分别输入 “I, You”，得出的 Concordance Hits 分别为 387, 1042，可以看到学生应该以给出建议为主，符合文章的主旨。但是第一人称，尤其是第二人称使用频率太高，容易使文章流于口语化。同样，分别输入 “should, help”，情况也不容乐观，Concordance Hits 分别为 84, 139。查询任一汉英词典，“应该”的相应表达应至少有 “should, ought to, must, be supposed to”；“帮助”应有 “help, assist, assistance, aid”，但事实是，学生大都偏向第一种表达。最后，点击进入软件右上方 word list，可以更科学、直观地了解学生的整体用词频率。以下截图排名前 15 的高频率词汇（图 1）：

Word List Results 1			
Word Types: 815		Word Tokens: 10421	Search Hits: 0
Rank	Freq	Word	Lemmas Word Form(s)
1	987	you	
2	507	to	
3	396	can	
4	305	i	
5	257	games	
6	247	online	
7	246	the	
8	227	give	
9	223	have	
10	200	up	
11	167	and	
12	139	a	
13	133	some	
14	127	help	
15	114	bad	

图 1 排名前 15 的高频率词汇

整体来看，高二学生在第二学期，即将进入高三综合学习阶段，英语写作用词单调，缺乏多样化。

## 2. 词语搭配

点击 cluster 界面，将 cluste, mini size 设为 3, max 设为 3, 在左下角 search term 输入 “ask”，可得到相关搭配使用情况，部分截图如下（图 2）：

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List			
Total No. of Cluster Types 18		Total No. of Cluster Tokens 24	
Rank	Freq	Range	Cluster
1	3	3	ask for help
2	3	3	ask your teacher
3	2	2	ask others personal
4	2	1	ask someone for
5	1	1	ask for some
6	1	1	ask help for
7	1	1	ask help. all
8	1	1	ask one for
9	1	1	ask others for
10	1	1	ask people for
11	1	1	ask somebody for
12	1	1	ask someone to

图 2 相关搭配使用情况

以 ask 这一常见搭配为例，可发现 ask 词语搭配使用出现错误。一般来说，ask for sb, 请/叫某人过来；ask sb./someone sth. 向某人要某物；ask sb. for help. 想某人寻求帮

助，这类搭配用法较多，但是不少学生并没有熟练掌握，且不用说变换使用多样化的表达法，可见，学生的英语基础水平急有待提高。

### 3. 语法分析

写作不是单词的简单堆砌，而是要遵守写作规则，就使用单词表达来讲，要遵守英语语法规则，笔者将从以下三个方面分析：

#### ① 情态动词的使用

根据英语语法规则，情态动词后直接加动词的原有形式或 be 再加相应形式。所收集的学生作文中，情态动词 can 的使用频率最高，但使用并不规范。以下选取典型的错误用法：

(1) Do some other things, for example, you can running or watching TV or reading books.

(2) Decided to give up online games, you can morning to evening to do play.

(3) You can your paln or your mother and father.

(4) You can to help for your teacher.

情态动词在英语学习中起举足轻重的作用，高中阶段涉及情态动词的用法并不少见，主要有情态动词的一般用法，即取情态动词本身的意义；情态动词的推测用法，即情态动词前加 could/should/would/might+情态动词或 could/should/would/might+情态动词+have done，表示对现在或过去情况的推测。作文中，主要涉及情态动词 can 的一般用法，属于较简单用法，但是学生并没有牢固掌握其用法，才导致作文犯低级的错误。

#### ② 词簇分析

进入 cluster 界面，点击 N-Grams，将 Grams size, mini size 设为 3，可以得到 N 字语词频表（图 3），以下截图排名靠前的结果：

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Total No. of N-Gram Types 5642		Total No. of N-Gram Tokens 10225				
Rank	Freq	Range	N-gram			
1	162	97	give up online			
2	161	97	up online games			
3	138	97	to give up			
4	109	97	decided to give			
5	106	97	have decided to			
6	105	97	you have decided			
7	104	97	break away from			
8	103	97	away from the			
9	103	97	hear you have			
10	102	97	glad to hear			
11	102	97	to hear you			
12	100	97	can break away			

图 3 N 字语词频表

通过以上表格分析数据不难发现，学生在表达“放弃”时，偏向使用“give up”，“决定做某事”时，用“decide to do sth.”；“摆脱某事”用“break away”。调查中，几乎所有学生都用到了这些表达，甚至不止一次使用，因此教师可在平时单词用法讲解时，多扩展相同意思的表达，学生也应在平时学习或做题中多加积累。

整体来看，在本次抽样建库的学生中，英语写作存在的重要问题是：词汇拼写错误、词汇使用不理想，语法基础薄弱，基本的情态动词使用和动词搭配会出错，这些应该引起高二学生和老师的重视。

### 4. 错误原因

写作常常伴随错误而存在，即使是母语学习者，写作或多或少也会出现各种各样的错误，找出错误所在，对错误进行归因，才能使外语学习者更清楚地意识到错误，进而进行克服。

本文将从英汉语言差异、母语迁移、过渡概括、内控因素和评价体系这五个方面来分析学生错误产生的主客观原因。

#### 4.1. 英汉语言差异

英汉语言差异既反映英汉语言思维方式的不同,又反映在英汉语言表达上。英语重形合,汉语重意合。美国翻译学家 Eugene A. Nida 在其 *Translation Meanings* (1983) 一书中曾深入浅出地说明了英汉这一差异:

就汉语和英语而言,也许在语言学上最重要的一个区别就是形合和意合的对比,在英语以及大多数的印欧语言中,句子的从属关系大多是用连接词如, *although, because, when, in order to that, so* 及 *so that* 等词明确地表达出来。但是,这用一概念,我们用意合的方法基本上也可以表达出来;那就是说,将两个句子放在一起并无连接词表明其相互关系,而从句子本身的意思中体现出来。这种形合和意合的差异反应在英语写作上,最明显的特点是学生作文多为简单句,流水式的写法。如:

- (1) You don't could give up online games. You should keep sleep.
- (2) You trouble in online games you might look for help.
- (3) I hope my advices can help you and I hope you can break away.
- (4) You can study history, you can go to the zoo in the 4:00 a.m. and you can see movies in the night.
- (5) I like play with my friends. I hope you can too. I enjoy listening to music. You can try. Try listening music.

以上例句均来自此次学生作文。显然,有的学生还没有习惯英语的表达习惯,作文缺少文章连接词,句子多为简单句,并且表达单一,如 *I hope* 和 *You can* 重复使用,英语语言忌重复,重替换,强调语言使用的多样性。

#### 4.2. 母语迁移

以教育心理学中的“迁移律”来进行分析,即:任何新知识、新熟巧、新技能的获得,都要受已知知识、熟巧、技能的影响;影响又分为积极影响和消极影响两种,即知识、技能、熟巧的正迁移和负迁移。显然,汉语在其后的英语学习中主要构成负迁移,即母语干扰。常见的错误有:名词单复数变化、动词的曲折变化、主谓一致等。笔者认为母语迁移中,影响较大的是母语思维迁移,以本次作文为例:

- (1) This world don't have difficult things, only have heart. people.
- (2) Because we is friends, and is good friends. So I hope you can decided to give up online games.
- (3) you can this think is very good.
- (4) you can morning to evening to do play.

这些句子我们都明白其中的含义,但是其英文表达欠妥,由于长期受母语思维方式的影响,写作中难免不由自主融入母语的表达,但是需要注意的是,应用符合目标语的表达来传递母语思想。学生可在平时写作中有意识避免这一现象。

#### 4.3. 过渡概括

过渡概括往往是造成语言错误的又一个原因。泰勒 (Taylor, 1986) 对语言学习错误进行过调查,发现初学者犯的母语干扰错误多,而中等程度学习者犯的过渡概括错误多,表现出过渡概括的使用与英语水平成正比,而语间转移的使用与英语水平成反比。如学生作文中出现的 *sheep*, 的复数形式写为 *sheeps*, “let me to help you” 以及 “You can playing the basketball” 等都是过渡概括的表现, *sheep* 的名词变复数不规则,为单复数同形,学生运用学习的一般名词变复数的规则,因此出现了错误。同样学生耳熟能详的是常见的 *want sb. to do sth.* 固一遇到 *let* 就套用了 *want* 的用法,出现错误。从本质上看,过渡概括属于语内迁移,其心理过程与语间转移一样均依赖先前的学习经历,进而影响学习新的内容。

#### 4.4. 内控因素

学生作为英语学习的主题，学习的主人和使动者，学生学习的积极主动性，势必是影响英语学习的主要原因之一。笔者在录入原文时，发现作文中仍存在大量拼写错误，以及句首首字母小写，标点符号尤其是英文句号乱使用等不规范书写，不少学生还存在字数不够，卷面不整洁，勾画乱涂，使用汉语拼音代替英语单词等现象。这些不良的写作习惯应该从日常的写作练习中有意识改正。学生应该端正学习态度，认真对待每一次考试，同时也应该认真对待课堂或课下的写作任务，养成良好的书写习惯。提高英语写作水平需要学生日积月累，只有重视英语写作，努力克服自身缺点，才能从根本上改善英语写作。

#### 4.5. 评价体系

教师作为学生学习的启发者和引领者，应完善对学生的评价体系。传统的学生评价体系受传统智能理论和应试教育的长期影响，过分强调甄别和选拔，很少关心考试反映出来的自身存在的问题和被评价者在各个时期的进步状况及努力程度，不能很好地发挥评价的改进、激励和发展；在评价内容上过于量化，如学生的成绩，使学生和教师在教育活动中表现和培养的创新精神、实践能力、心理素质、行为习惯等综合素质评价相对忽视；被评价对象基本上处于被动的检查、比批判的地位，其自尊心、自信心往往得不到很好的保护，对评价往往持一种冷漠、应付逃避的态度，甚至发生欺骗、弄虚作假等行为，使得评价效果适得其反。除此之外，教师应多完善过程评价模式，开展小组互助学习，观察、记录学生的点滴进步，不时对学生的情况反馈，多给予鼓励，增强学生学习的自信心。英语考试写作是平时写作的集中反映，教师适当选取生活中学生感兴趣的话题，进行发散式思维写作，亦可进行英语演讲比赛，锻炼学生的英语表达，这些形式不仅有助于提高英语写作，更使学习变得有用、有趣，使学生更积极主动融入学习。

### 5. 总结

学生英语学习，尤其是英语写作，不可避免会出现错误。影响学生写作水平的因素可能有：英汉语言差异，母语迁移，过渡概括，学生的内控因素和教师的评价体系。笔者认为英语写作水平的提高不是一蹴而就的，需要落实到每个单词的拼写，每个标点的正确使用，认认真真的书写，进而慢慢克服其它因素的干扰。本研究还存在不足，语料收集范围较小，未必能说明其它学校高二学生的学习英语写作学习情况。

### References

- [1] Biggs J B, Telfer B. The Process of Learning (Second Edition) [Z]. Prentice Hall of Australia.1987.149.
- [2] Baker, P., A. Hardie, T. McEnery. A Glossary of Corpus Linguistics [M]. Edinburgh: Edinburgh University Press, 2006:45-46,161.
- [3] Taylor, G.C. Errors and Explanations [J]. Applied Linguistics, 7 2, 1986:144-66.