

基于语料库的中国英语学习者写作中因果连接词研究

张雨欣

浙江工商大学外国语学院, 浙江 杭州

收稿日期: 2024年4月3日; 录用日期: 2024年5月21日; 发布日期: 2024年5月31日

摘要

因果逻辑关系作为四种基本逻辑关系之一, 对英语写作的逻辑性和流畅性起着至关重要的作用。因此, 本研究基于ETS和LOCNESS语料库, 通过R和Wordsmith对比分析中国英语学习者与英语本族语者在因果连接词使用上的异同点以及不同成绩段的中国英语学习者的特点。研究发现, 中国英语学习者整体使用因果衔接词的频率高于英语本族语者, 尤其是“so”和“because”, 这体现了他们对口语化衔接词以及显性衔接使用上的偏好。相比之下, 他们在“since”、“hence”和“as a result”等词的使用上与英语本族语者存在差异, 但并不如“so”和“because”明显。而对于“therefore”、“consequently”和“due to”等词的使用, 中国学习者能较好地接近英语本族语者的习惯。此外, 研究还指出中国英语学习者在不同成绩段中因果连接词的使用上存在明显的层次性。高成绩段的学习者显示出较成熟的语言运用能力, 而低成绩段的学生则表现出对口语化以及简单直接表达的偏好。中等成绩段的学生则处于两者之间。最后在对中国英语学习者因果连接词的错误分析中概括出了因果连接词误用、因果连接词过度使用以及因果连接词缺失这三个错误分类, 以期为教师加强对学生使用因果连接词的指导, 从而为有效提升他们的英语写作水平提供帮助。

关键词

因果连接词, 中国英语学习者, 语料库, R语言

A Corpus-Based Study of Causal Connectives in the Writing of Chinese English Learners

Yuxin Zhang

School of Foreign Languages, Zhejiang Gongshang University, Hangzhou Zhejiang

Received: Apr. 3rd, 2024; accepted: May 21st, 2024; published: May 31st, 2024

Abstract

Causal logic, as one of the four fundamental logical relations, plays a pivotal role in ensuring coherence and fluency in English writing. Consequently, this study draws upon the ETS and LOCNESS corpora to conduct a comparative analysis of similarities and differences in the usage of causal connectives between Chinese English learners and native English speakers, as well as identifying distinctive features among Chinese learners across varying proficiency levels. The findings reveal that Chinese English learners overall employ causal connectives more frequently than native speakers, particularly favoring colloquial connectives, such as *so* and *because*. In contrast, their usage of terms like *since*, *hence*, and *as a result* differs from that of native speakers, though not as markedly as with *so* and *because*. However, they demonstrate a closer alignment with native speaker habits in the usage of words like *therefore*, *consequently*, and *due to*. Additionally, the study highlights a clear gradation in the use of causal connectives among Chinese English learners across different performance levels. High-achieving learners exhibit a more sophisticated command of language, while low-achieving students show a preference for colloquial and straightforward expressions. Middle-achieving students fall somewhere in between. Finally, the study summarizes three categories of errors related to causal connectives among Chinese English learners: misuse, overuse, and omission of causal connectives. This categorization aims to offer instructors insights for better guiding students on the use of causal connectives, thereby helping to enhance students' English writing proficiency.

Keywords

Causal Connectives, Chinese English Learners, Corpus, R Language

Copyright © 2024 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

逻辑连接词的运用可以提高语篇的连贯性与清晰度,进而促使读者迅速把握文本的中心思想及结构脉络以实现有效的沟通[1]。因果逻辑关系,作为四种基本逻辑关系之一[2],其在写作中的恰当应用对于确保作文逻辑性与流畅性有着重要影响。在英语写作中,因果逻辑关系多通过一系列因果连接词来准确传递有关句子之间或段落之间的各种语义联系和逻辑联系。

国内学者对连接词的研究颇丰,其中一些研究从认知视角出发探究连接词[3],还有学者借助语料库研究中高中生英语写作逻辑连接词的使用以探讨二语习得[4]。尽管如此,在现有文献中,针对因果连接词的研究较为有限,大多数研究只是单纯地通过语料库来检索和分析英语学习者的因果连接词使用[1][5][6][7][8]。然而在语料库中借助 R 语言进行可视化分析的研究还相对罕见。此外,国内研究多为横向比较,缺乏对学习者在语言习得过程中因果连接词使用情况变化的纵向研究。

鉴于此,本研究旨在运用“ETS”(非母语英语写作语料库)和“LOCNESS”(英语母语写作语料库)并结合 R 以及 wordsmith8.0,比较分析中国英语学习者与英语本族语者在使用因果连接词的异同点以及不同成绩段的中国英语学习者在使用因果连接词上的特点,以期有关因果连接词方面的英语教学以及中国英语学习者在英语写作中正确运用因果连接词提供借鉴。

2. 研究设计

2.1. 研究问题和方法

- 1) 与本族语者相比，中国英语学习者写作中因果连接词的使用存在着什么异同点？
- 2) 不同成绩段的中国英语学习者写作中因果连接词的使用呈现了什么特点？

本研究基于 Halliday & Hasan 的衔接理论并参考和借鉴了莫俊华对因果连接词的分类标准[2] [5]从语法和功能的角度将因果连接词分为连词短语、副词短语和介词短语三类，之后筛选和删除了语料库中频次小于 10 的因果连接词，最终选取了以下因果连接词，如下表 1 所示。

Table 1. Classification of causal connectives
表 1. 因果连接词的分类

词性	示例
连词(短语)	So; because; since
副词(短语)	So; therefore; thus; hence; as a result; consequently; accordingly
介词(短语)	due to; because of

2.2. 语料来源

本研究采用了“ETS”(非母语英语写作语料库)和“LOCNESS”(英语母语写作语料库)。ETS 语料库是由非英语母语人士撰写的 12,100 篇文章的集合，这些文章是作为托福考试的一部分。该语料库包括来自 11 种不同国家的非英语为母语的学习者撰写的文章，并覆盖了一系列主题以及不同分数段(高中低)的考生作文。基于本研究的研究问题，将会从 ETS 语料库中筛选出中国学生的作文用于研究。

LOCNESS 语料库集结了由英语为母语的大学生撰写的一系列作文，总计超过 320,000 字。该语料库的实用性在于它提供了一个可靠的基准来分析英语母语者的写作模式，这对于研究英语作为第二语言的学习者的写作习得尤为重要。因此，可以利用该语料库来进行比较研究，探索母语和非母语之间在写作风格、结构、词汇使用等方面的差异。通过这种比较，可以发展和提炼教学方法，帮助非英语母语者提高他们的写作技能。

2.3. 研究步骤

第一步，收集语料。本研究的初步步骤涉及将 ETS 语料库导入 R Studio 中并从中提取标注为 ZHO(中国)的语料作为研究问题一的分析基础。其次将 ZHO 的语料分别筛选出高、中、低三个不同成绩段学生的英语作文用于研究问题二的分析。相关代码如图 1 所示。

```
library(dplyr)
library(readr)
setwd("D:/ETS_RAW/RAW")
ets <- readtext("*.txt", text_field = "text")
ets_df <- ets %>% separate_wider_delim(doc_id,delim = " ",
                                     names = c("grade","region","topic",NA),
                                     cols_remove = F)

# 过滤出 region 为 ZHO 的数据
ets_zho <- filter(ets_df, region == "ZHO")

# 创建三个子集: HIGH, MEDIUM, LOW
ets_zho_high <- filter(ets_zho, grade == "HIGH")
ets_zho_medium <- filter(ets_zho, grade == "MEDIUM")
ets_zho_low <- filter(ets_zho, grade == "LOW")

# 分别导出 HIGH、MEDIUM、LOW 子集于 D:/ETS_RAW/RAW 中
write.table(ets_zho_high, file = "ets_zho_high.txt", sep = "\t", row.names = FALSE, quote = FALSE)
write.table(ets_zho_medium, file = "ets_zho_medium.txt", sep = "\t", row.names = FALSE, quote = FALSE)
write.table(ets_zho_low, file = "ets_zho_low.txt", sep = "\t", row.names = FALSE, quote = FALSE)
```

Figure 1. R code for filtering different score ranges in the ZHO corpus
图 1. 筛选 ZHO 不同成绩段的 R 代码

继而得出语料详情如表 2，表 3 所示。

Table 2. ETS_ZHO corpus and LOCNESS corpus
表 2. ETS_ZHO 语料库和 LOCNESS 语料库

Corpus	Tokens
ETS_ZHO	359,287
LOCNESS	326,384

Table 3. ETS_ZHO corpus for high, medium and low performance
表 3. ETS_ZHO 高中低成绩段语料库

Corpus	Tokens
ETS_ZHO_HIGH	103,864
ETS_ZHO_MEDIUM	233,300
ETS_ZHO_LOW	22,123

第二步，检索语料库中所有的因果连接词。鉴于“so”在英语中可能作为程度副词修饰形容词或者作为短语“so that”和“so as to”的组成部分出现，或者“since”在英语中除了“因为”还有“自……以来”等意思。因此，本研究将步骤一中筛选的语料库导出，并根据表 1 的分类标准，应用于 Wordsmith 8.0 中进行索引行检索并通过手工识别和删除不符合因果连接词定义的实例。最终筛选得到的数据被整理入 Excel 表格，以便进一步分析。此外，鉴于 ETS_ZHO 语料库中不同成绩段(高、中、低)的语料库规模不平衡，因此，本研究对这三个库中检索出的因果连接词的频率进行了标准化处理，转换为每 10 万词的相对频数呈现。

经过手工筛选排除非因果连接词后，将 excel.csv 格式转入 R Studio 形成数据框，并将 excel 中的宽数据格式转换为长数据格式，使之更容易在 R 中进行数据处理分析和绘制图表[9]。ETS_ZHO 语料库与 LOCNESS 语料库中因果连接词的使用频次(部分)如表 4 所示；ETS_ZHO 语料库中高、中、低成绩段的三个语料库中因果连接词的使用相对频数(部分)如表 5 所示。相关代码如图 2 所示。

```
library(tidyr)
LOCNESS_df <- read.csv("C:/Users/82050/Desktop/ETS and LOCNESS.csv")
mydf_three <- read.csv("C:/Users/82050/Desktop/高中低因果连接词 word.csv")

# 将宽格式数据转换为长格式
LOCNESS_df <- pivot_longer(LOCNESS_df, cols = -corpus, names_to = "Connective", values_to =
"Frequency")
View((LOCNESS_df))
mydf_three <- pivot_longer(mydf_three, cols = -Grade, names_to = "Connective", values_to = "Relative
Frequency")
View((mydf_three))
```

Figure 2. R code for filtering causal connectives
图 2. 筛选因果连接词的 R 代码

Table 4. Causal connectives in the ETS_ZHO and LOCNESS
表 4. ETS_ZHO 与 LOCNESS 中的因果连接词

	Corpus	Connective	Frequency
1	ETS_ZHO	so	834
2	ETS_ZHO	because	1015

续表

3	ETS_ZHO	since	187
4	ETS_ZHO	therefore	332
5	ETS_ZHO	thus	151
6	ETS_ZHO	hence	51
7	ETS_ZHO	as a result	99
8	ETS_ZHO	consequently	32
9	ETS_ZHO	accordingly	10
10	ETS_ZHO	due to	116
11	ETS_ZHO	because of	172
12	LOCNESS	so	305
13	LOCNESS	because	665
14	LOCNESS	since	82
15	LOCNESS	therefore	304
16	LOCNESS	thus	139

Table 5. Causal connectives across different score ranges in ETS_ZHO
表 5. ETS_ZHO 不同成绩段中的因果连接词

	Grade	Connective	Relative Frequency
1	High	so	152.12
2	High	because	223.37
3	High	since	77.03
4	High	therefore	124.20
5	High	thus	61.61
6	High	hence	19.26
7	High	as a result	36.57
8	High	consequently	13.48
9	High	accordingly	1.96
10	High	due to	44.29
11	High	because of	33.70
12	Medium	so	254.18
13	Medium	because	301.76
14	Medium	since	45.01
15	Medium	therefore	82.73
16	Medium	thus	35.58

第三步, ETS_ZHO 语料库与 LOCNESS 语料库因果连接词的卡方检验分析。本研究的第一个研究问题旨在通过比较 ETS_ZHO 语料库中中国学生的英语写作和 LOCNESS 语料库中英语母语者的写作中因果连接词的使用频率, 探索两者在因果连接词使用上的差异与共性。为此, 本研究采用卡方检验这一统

计方法，定量评估两个语料库中每个因果连接词使用频率的显著性差异。第四步，对 ETS_ZHO 语料库中高中低三个成绩段的中国学生的英语写作中因果连接词的使用频次进行可视化分析并以折线图的形式呈现，旨在探究不同成绩段的中国英语学习者在因果连接词使用上的特点。

3. 结果与分析

3.1. ETS_ZHO 语料库与 LOCNESS 语料库因果连接词分析

3.1.1. ETS_ZHO 与 LOCNESS 中因果类连接词的使用频率和分布情况

检索两个库中的所有的因果连接词后，通过 R studio 以可视化的形式展示，代码及柱形图如图 3，图 4 所示。

```
library(tidyr)
library(ggplot2)

LOCNESS_df <- read.csv("C:/Users/82050/Desktop/ETS and LOCNESS.csv")

# 将宽格式数据转换为长格式
LOCNESS_df <- pivot_longer(LOCNESS_df, cols = -corpus, names_to = "Connective", values_to = "Frequency")
View(LOCNESS_df)

#创建柱状图
ggplot(LOCNESS_df, aes(x = reorder(Connective, -Frequency), y = Frequency, fill = corpus)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  labs(x = "Connective", y = "Frequency") +
  theme_minimal() +
  scale_fill_manual(values = c("ETS_ZHO" = "#E7B800", "LOCNESS" = "#FC4E07")) +
  theme(axis.text.x = element_text(angle = 15, vjust = 0.8))
```

Figure 3. R code for the bar chart of causal connective frequencies in ETS_ZHO and LOCNESS

图 3. ETS_ZHO 与 LOCNESS 中因果类连接词的柱形图的 R 代码

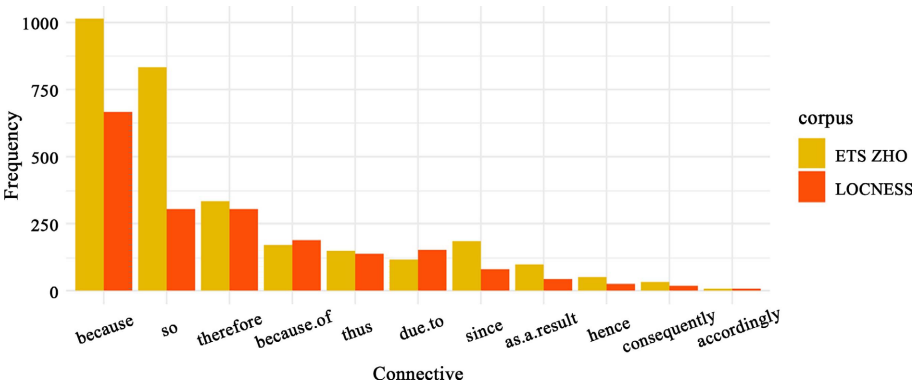


Figure 4. The bar chart of causal connective frequencies in ETS_ZHO and LOCNESS

图 4. ETS_ZHO 与 LOCNESS 中因果类连接词的柱形图

图 4 表明，从整体情况来看，中国英语学习者写作中因果连接词的使用频率整体高于英语本族语者，由此可见，中国英语学习者在表述因果关系时倾向于使用显性衔接手段，即连接词。或许这与英语教学中教师过分强调连接词在语篇中的衔接连贯作用有关系。学生可能因此形成意识上的偏颇，认为连接词是达到语篇衔接连贯的唯一方法，却忽略了其它隐性衔接手段[7]。具体来看，中国英语学习者和英语本

族语者在因果连接词的选择上有显著的相似性，但在使用的频率上却存在较大差异，中国英语学习者在使用“because”和“so”这两个因果连接词时频率显著高于英语本族语者，这可能暗示了中国英语学习者在英语写作中倾向于使用这些连接词来明确表示因果关系。

相反，在使用“therefore”，“because of”和“thus”等其他连接词时，两者使用频率较为接近，表明在某些情况下，中国英语学习者能够接近英语本族语者的使用习惯。此外，例如“hence”，“consequently”和“accordingly”等词的使用频率在两个语料库中都较低，这可能表明这些连接词在议论文写作中的使用度较低。

3.1.2. ETS_ZHO 与 LOCNESS 中因果类连接词的使用情况分析

为了更进一步探究中国英语学习者和英语本族语者在因果连接词上的使用情况，本研究通过对 ETS_ZHO 与 LOCNESS 中检索出的 11 类因果连接词进行卡方检验，代码如图 5 所示并得出以下结论如表 6 所示。

```
library(tidytr)
LOCNESS_df <- read.csv("C:/Users/82050/Desktop/ETS and LOCNESS.csv")

rownames(LOCNESS_df)=LOCNESS_df[,1]
LOCNESS_df=LOCNESS_df[,-1]
#进行卡方检验
chisq.test(LOCNESS_df[,1])
chisq.test(LOCNESS_df[,2])
chisq.test(LOCNESS_df[,3])
chisq.test(LOCNESS_df[,4])
chisq.test(LOCNESS_df[,5])
chisq.test(LOCNESS_df[,6])
chisq.test(LOCNESS_df[,7])
chisq.test(LOCNESS_df[,8])
chisq.test(LOCNESS_df[,9])
chisq.test(LOCNESS_df[,10])
chisq.test(LOCNESS_df[,11])
```

Figure 5. R code for chi-square test on causal connectives in the ETS_ZHO and LOCNESS

图 5. ETS_ZHO 与 LOCNESS 中因果类连接词卡方检验的 R 代码

Table 6. Chi-square test on causal connectives in the ETS_ZHO and LOCNESS

表 6. ETS_ZHO 与 LOCNESS 中因果类连接词卡方检验

Causal Connectives	X-squared	p-value	
so	245.69	<2.2e-16	+
because	72.917	<2.2e-16	+
since	40.985	= 1.534e-10	+
therefore	1.2327	= 0.2669	-
thus	0.49655	= 0.481	-
hence	8.1169	= 0.004385	+
as a result	20.25	= 6.795e-06	+

续表

consequently	2.283	= 0.1308	-
accordingly	0	= 1	-
due to	5.0892	= 0.02408	-
because of	0.80055	= 0.3709	-

表 6 显示了中国英语学习者(ETS_ZHO)与英语本族语者(LOCNESS)在使用不同因果连接词的频率上的差异。其中,“so”和“because”的卡方值较高,且其 P 值远小于 0.05,这表明在这两个词的使用频率上,中国英语学习者与英语本族语者之间存在着显著差异。结合柱状图中的信息得出,中国英语学习者频繁使用“so”和“because”这两个因果连接词的原因可能是这两个词相对于其他词更简单,更易习得。此外值得注意的是,“so”的卡方值位列第一,说明相比于英语本族语者,中国英语学习者在这 11 类因果连接词中最偏爱使用“so”,而“so”具有较强的口语色彩[5] [10],经常出现在口语语篇中而非书面语中,由此可以看出中国英语学习者在写作中存在着明显的口语化倾向。

此外,对于“since”,“hence”和“as a result”,虽然 P 值也小于 0.05,显示了统计学上的显著差异,但其卡方值较“so”和“because”要小得多,这表明虽然存在差异,但其使用频率的差距不如“so”和“because”那么悬殊。由此,比起英语本族语者,中国英语学习者更多地使用这三个因果连接词的原因可能有两点:其一,在中国的英语教学中,老师可能更倾向于教授这些特定的连接词作为表达因果关系的工具,导致学生在写作中更频繁地使用它们。其二,中国英语学习者有时会过度泛化某些语言点的使用,特别是当他们认为这些表达可以帮助其写作文本看起来更加“学术”时。而且“since”,“hence”和“as a result”比起“consequently”,“accordingly”等词更易习得,因此,这可能导致这三个因果连接词的使用频率异常高。

另一方面,“therefore”,“thus”,“consequently”,“accordingly”,“due to”以及“because of”的 P 值大于 0.05,表明这些词在两个语料库中的使用频率没有显著差异。这可能反映出这些词在写作中具有较为统一的使用频率。

综上所述,中国英语学习者整体使用因果衔接词的频率高于英语本族语,这体现了他们对显性衔接使用的偏好,这种趋势可能源于英语教学中对连接词作为语篇连贯关键手段的过分强调,引导学生偏向于认为连接词是实现语篇连贯的主要手段。此外,中国英语学习者在写作中尤其依赖“so”和“because”,显著超出英语本族语者的使用频率,这反映出他们对口语化连接词的偏好。尽管“since”,“hence”和“as a result”等词与英语本族语者的使用频率也存在显著差异,但远不及“so”和“because”的使用频率差异显著。而“therefore”,“consequently”,“accordingly”,“due to”以及“because of”等词使用频率相似,由此显示了某些情况下学习者能接近英语本族语者的使用习惯。

3.2. 高中低不同成绩段的中国英语学习者因果连接词使用情况分析

鉴于 ETS_ZHO 语料库中不同成绩段(高、中、低)的语料库规模不平衡,因此,本研究对这三个库中检索出的因果连接词的频率进行了标准化处理,转换为每 10 万词的相对频数呈现。结果呈现如表 7 所示。

为了更直观地看出三个成绩段的区别,本研究将此表放入 R Studio 中进行可视化操作,代码如图 6,折线图如图 7 所示。

3.2.1. 整体趋势分析

从表 7 以及图 7 中可知,从整体上来看,好成绩段的中国英语学习者在使用因果连接词比中、低成绩段的学习者节制,这可能是因为他们更加注重语义的准确和逻辑的严密。他们的写作不单纯依赖显性

Table 7. Standardized frequencies of causal connectives in different score ranges sub-corpus of ETS_ZHO

表 7. ETS_ZHO 不同成绩段语料库中因果类连接词的标准化频率

Causal Connectives	High	Medium	Low
So	152.12	254.18	375.18
Because	223.37	301.76	357.09
Since	77.03	45.01	9.04
Therefore	124.2	82.73	45.2
Thus	61.61	35.58	18.08
Hence	19.26	12.86	4.52
As a result	36.57	25.29	9.04
Consequently	13.48	7.29	4.52
Accordingly	1.96	3.34	0
Due to	44.29	27.43	27.12
Because of	33.7	56.58	0
Total	787.59	852.05	849.79

```
library(tidyr)
mydf_three <- read.csv("C:/Users/82050/Desktop/高中低因果连接词 word.csv")

# 将宽格式数据转换为长格式
mydf_three <- pivot_longer(mydf_three, cols = -Grade, names_to = "Connective", values_to = "Frequency")
View(mydf_three)
#创建折线图
library(ggplot2)
ggplot(mydf_three, aes(x = Connective, y = Frequency, group = Grade, color = Grade)) +
  geom_line() +
  geom_point() +
  theme_minimal() +
  labs(x = "Causal Connective", y = "Frequency") +
  theme(axis.text.x = element_text(angle = 30, hjust = 0.7))
```

Figure 6. R code for line plots of the standardized frequencies of causal connectives across different score ranges sub-corpus of ETS_ZHO

图 6. ETS_ZHO 中不同成绩段中因果类连接词的折线图的 R 代码

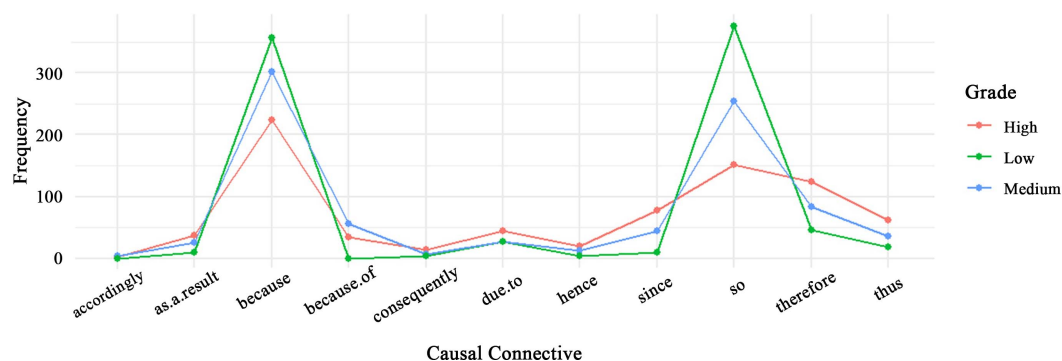


Figure 7. Line plots of the standardized frequencies of causal connectives across different score ranges sub-corpus of ETS_ZHO

图 7. ETS_ZHO 中不同成绩段中因果类连接词的折线图

的连接词，也不会盲目地使用过多的因果连接词，而是综合利用内在逻辑和上下文的隐性衔接，表现出较高的语篇组织能力。中等成绩段的学习者在因果连接词的使用上介于高分和低分学生之间，显示了适中的使用频率和一定的词汇多样性，但相比高分学生，可能在复杂连接词的使用上略显不足。而低成绩段的学习者对因果连接词的掌握和使用比较单一，主要表现出对“so”和“because”这两个因果连接词的过度依赖，两者的使用频率显著高于其他因果连接词，对“therefore”的使用次之。导致的原因可能一方面是受制于其较弱的英语语言水平，另一方面是由于学习者在写作过程中采取了“回避策略”，有意识地避免使用不熟悉或没把握的词汇。

3.2.2. 具体因果连接词使用分析

从细节来看，对于“so”和“because”这两个因果连接词，在所有成绩层次中均是使用频率最高的连接词，但随着成绩等级的降低，其使用频率不断增高，尤其是“so”的使用在低成绩层次中最为突出，这一点再次佐证了中国英语学习者在书面写作中的口语化倾向，尤其是对于低成绩段的学生来说。

对于在写作中常见的“since”，“therefore”和“thus”等连接词在好成绩学习者中使用频率较高，中等成绩段的学生次之，低成绩段的学生使用最少，这能显示出相比于其他成绩段的学习者，高分学生能够灵活运用这些词汇来构建复杂的逻辑关系。与此趋势相同的还有“hence”，“as a result”，“consequently”和“due to”等使用频率较低的连接词，可能因为这些词汇在学习英语的过程中出现的频率不高，学生不够熟悉，从而导致使用度低。

综上所述，中国英语学习者在不同成绩层次中因果连接词的使用存在明显的层次性。好成绩段的学习者表现出更为成熟的语言运用能力，而低成绩段的学习者的写作则体现出对口语化以及简单直接表达的偏好(即更多地使用“so”和“because”)。因此在教学中，老师应该要加强对高级连接词的教学和实践，以促进学生能力的提升，特别是对低成绩学生的指导，以减少对“so”和“because”等口语化连接词的过度依赖。

3.3. 有关中国英语学习者因果连接词的错误分析

本研究对因果连接词的错误使用进行人工识别和分类后可以发现，部分中国英语学习者在写作中不能恰当或正确地使用因果连接词。通过相关索引行的分析可以看出中国英语学习者在写作中使用因果连接词会犯一系列错误。例如，部分学习者由于对连接词用法的了解有限而错误地使用因果连接词。除此之外，有的学生在写作中倾向于过多使用不必要的因果连接词，这往往会导致写作晦涩难懂，而另一些学生则很少使用连接词，导致作文不合逻辑，难以阅读。因此本研究将这些连接词的使用错误主要分为三类：因果连接词误用、因果连接词过度使用以及因果连接词缺失。

3.3.1. 因果连接词误用

例 1: People didn't change job too much once they were get used to on they own job, so that is why people make different types money and how successful their will be. **Therefore**, people were being successful it is because they knew how to change their life by change their own job. (ETS_ZHO_LOW)

忽略句中的语法错误，该例句中前后两句表达的是相同的意思，因此用“therefore”来表达因果关系是不合理的。因此如要用“therefore”来连接句子的话，后句的内容应该重点突出这种行为所导致的结果或影响。

3.3.2. 因果连接词过度使用

例 2: Maybe they have much more important things need to do and share their times, so they consider they are very

bussy (busy), so they aren't willing to pay more time to helping their communities. (spend more time helping...)
(ETS_ZHO_LOW)

这三句话中的衔接词连续出现了两次“so”，因果连词“so”的过度使用不仅会导致语义上的混乱，而且会影响句子的流畅性。因此在这段话中应该删除一个“so”，以使其符合逻辑且易于阅读。如果需要表达多个因果关系，可以使用不同的连词，如“therefore”，“thus”，“hence”等，这样可以使语义更加清晰，同时丰富句子的表达。

3.3.3. 因果连接词缺失

例 3: For example, nowadays our conversations contain more knowledge and have become broader than before. We must learn more from anywhere (everywhere) so that we can talked (talk) with others. (ETS_ZHO_LOW)

在这个例句中，应该在两个句子间加“therefore”，句子的逻辑关系才能通顺。“therefore”是因果连接词，它表明了一个结果或后果。第一句描述了现状，指出现代交谈中包含的知识更多，范围也更广。而“therefore”引导了因果关系，说明因为交谈包含更多知识并且范围更广，所以必须更广泛地学习，以便能够与他人交流。通过加入“therefore”，两个句子之间的逻辑关系得以清晰地表达出来。如果缺失了因果连接词，则会导致句子之间缺乏合理的过渡。

综上所述，中国英语学习者在正确运用因果连接词方面存在很大困难，他们常常以多种方式误用这些连接词。鉴于此，正确使用因果连接词在写作教学中的重要性不容忽视。为了减少写作中的错误，教师应当加强对使用因果连接词的指导，有效提升他们的英语写作水平。

4. 结论

本文基于语料库并结合 R 试图探究“ETS”(非母语英语写作语料库)和“LOCNESS”(英语母语写作语料库)中中国英语学习者与英语本族语者在使用因果连接词的异同点以及不同成绩段的中国英语学习者在使用因果连接词上的特点，研究结果如下。

首先，在对 ETS 和 LOCNESS 语料库中中国英语学习者与英语本族语者的因果连接词分析中发现，中国英语学习者整体使用因果衔接词的频率高于英语本族语者，这体现了他们对显性衔接使用的偏好，这种趋势可能源于英语教学中对连接词作为语篇连贯关键手段的过分强调。此外，中国英语学习者在写作中尤其依赖“so”和“because”，显著超出英语本族语者的使用频率，这反映出他们对口语化连接词的偏好。尽管“since”，“hence”和“as a result”等词与英语本族语者的使用频率也存在显著差异，但远不及“so”和“because”的使用频率差异显著。而“therefore”、“consequently”，“accordingly”，“due to”以及“because of”等词使用频率相似，由此显示了某些情况下学习者能接近英语本族语者的使用习惯。因此在英语教学中，老师要加强对因果连接词使用上的指导，让学生认识到除了显性连接词外，还有其他多种文本内在连贯手段。

其次，在对高中低不同成绩段的中国英语学习者因果连接词使用情况分析中发现，中国英语学习者在不同成绩层次中因果连接词的使用存在明显的层次性。好成绩段的学习者表现出更为成熟的语言运用能力，而低成绩段的学习者的写作则体现出对口语化以及简单直接表达的偏好(即更多地使用“so”和“because”)。中等成绩段的学习者在因果连接词的使用上介于高分和低分学生之间。因此在教学中，老师应该要加强对高级连接词的教学和实践，以促进学生能力的提升，特别是对低成绩学生的指导，帮助他们克服对基础连接词的过度依赖，引导他们掌握更多样化的连接词，以提升其整体的写作能力。

最后，在对中国英语学习者因果连接词的错误分析中发现，部分中国英语学习者在写作中不能恰当或正确地使用因果连接词。这些连接词的使用错误主要分为三类：因果连接词误用、因果连接词过度使

用以及因果连接词缺失。因此，正确使用因果连接词在写作教学中的重要性不容忽视。为了减少写作中的错误，教师应当加强对使用因果连接词的指导，有效提升他们的英语写作水平。

参考文献

- [1] 戴俊红. 非英语专业学生限时议论文写作中的因果连接词使用研究[J]. 牡丹江大学学报, 2013, 22(12): 176-180.
- [2] Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English*. Longman, London.
- [3] 吴晓春, 杨忠. 认知视角与外语学习者逻辑连接词使用错误[J]. 东北师大学报(哲学社会科学版), 2011(1): 139-142.
- [4] 凌梦琪. 基于语料库的高中生英语写作中连接词的使用状况调查研究[D]: [硕士学位论文]. 赣州: 赣南师范大学, 2023.
- [5] 莫俊华. 中国学生在议论文写作中使用因果连接词的语料库研究[J]. 外语教学, 2005(5): 45-50.
- [6] 杨建生. 基于语料库的英语专业学生议论文中因果连接词使用研究[J]. 河北联合大学学报(社会科学版), 2012, 12(2): 82-85.
- [7] 刘西娟, 刘秀蓉. 非英语专业本科生议论文中的因果连接词研究[J]. 西安石油大学学报(社会科学版), 2015, 24(6): 100-106.
- [8] 韩存新, 陈京娜. 中学生英语议论文中的因果连接词研究[J]. 基础外语教育, 2020, 22(1): 22-29+109.
- [9] 施雅倩, 雷蕾. 语言数字人文与 R 语言实践[M]. 上海: 上海交通大学出版社, 2023.
- [10] 文秋芳, 丁言仁, 王文字. 中国大学生英语书面语中的口语化倾向——高水平英语学习者语料对比分析[J]. 外语教学与研究, 2003(4): 268-274+321.