

基于语料库的留学生因果关联词习得研究

沙博翰¹, 任晓霏²

¹南京师范大学国际文化教育学院, 江苏 南京

²江苏大学文学院, 江苏 镇江

收稿日期: 2024年5月16日; 录用日期: 2024年7月11日; 发布日期: 2024年7月24日

摘要

运用中介语语料库的频数为分析单元, 比较分析留学生因果关联词的应用特征与差异原因, 探究留学生因果关联词习得与汉语“意合”特征的匹配度。结果显示, 留学生存在过度使用因果关联词现象, 包括成分多加、语义重复、语序错位、语用不当、逻辑混乱等。研究结果对汉语教学与汉语教材编写具有一定参加价值。

关键词

语料库, 因果关联词, 形合, 意合

Corpus-Based Study on the Acquisition of Cause-Related Words by International Students

Bohan Sha¹, Xiaofei Ren²

¹International College for Chinese Studies, Nanjing Normal University, Nanjing Jiangsu

²School of Liberal Arts, Jiangsu University, Zhenjiang Jiangsu

Received: May 16th, 2024; accepted: Jul. 11th, 2024; published: Jul. 24th, 2024

Abstract

Using the frequency of interlanguage corpus as the unit of analysis, this paper compares and analyzes the application characteristics and differences of overseas students' cause-related words, and explores the matching degree between overseas students' acquisition of cause-related words

文章引用: 沙博翰, 任晓霏. 基于语料库的留学生因果关联词习得研究[J]. 现代语言学, 2024, 12(7): 358-364.

DOI: 10.12677/ml.2024.127567

and Chinese “parataxis”. The results show that the phenomenon of excessive use of causal related words by overseas students includes excessive composition, semantic repetition, misplacement of word order, improper pragmatics, and logic confusion. The results of this study are valuable for Chinese teaching and compiling Chinese textbooks.

Keywords

Corpus, Causal-Related Words, Hypotaxis, Parataxis

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

汉语和英语是世界上使用人数最多与使用人数最广的语言，且分属于汉藏与印欧两种差异较大的语系。从话语组织角度来说，汉语重视“意合”，英语重“形合”。所谓“意合”指不借助语言形式手段而借助词语或句子所含意义的逻辑联系来实现它们之间的连接；所谓“形合”指借助语言形式手段实现词语或句子的连接，前者注重语义的连贯，后者注重形式的对应[1]。随着英语在中国的日渐普及，英语重形式的语法结构特征开始影响汉语，这种现象在人们平常的话语中初见端倪。汉语重意合，“句子与句子的关系让对话人意会而不用连词”。虽然汉语中也有表示形态的语法，但是主要还是以意合为主[2]。在汉语教学中，无论是教师的课堂教学效果还是教材知识的呈现，像翻译那样将语言项目一一对应并不明智，而是应该教授符合中国人说话习惯和认知方式的汉语知识。汉英翻译中各个语言层次的对应应当遵循适当适应和灵活对应，在个别对应中求得整体对应[3]。尽管留学生说汉语时运用大量“形合”标记成分组成句子，并不妨碍中国人理解话语，但是会让中国人感觉不地道，进而影响交际效果。

2. 研究设计

(一) 研究问题

基于国家语委主持的现代汉语语料库和 HSK 动态作文语料库分别检索因果关联词在各语料库中的频次。通过研究，本文试图验证以下问题：留学生对因果关联词的运用有何特征？是否符合汉语一贯的表达习惯？造成这种现象的原因是什么？《高等学校外国留学生汉语长期进修教学大纲》要求留学生掌握因果关联词类型有 9 种(见表 1)，《汉语水平等级标准与语法等级大纲》列举的因果关联词有 20 个(见表 2)。

Table 1. “Syllabus for Long-Term Chinese Education for Foreign Students in Colleges and Universities”

表 1. 《高等学校外国留学生汉语长期进修教学大纲》

等级 关联词	初级 (一)	初级 (二)	中级	高级
因果关系 (9 个)	(因为)…… 所以……	由于……, ……、 ……, 因此……、 既然……就……	之所以…… 是因为…… ……以致……	(因为)……, 故……、……, 以至于……、所以……, 所以……, (还)因为……、……因……

Table 2. “Chinese Language Proficiency Level and Grammar Level Outline”
表 2. 《汉语水平等级标准与语法等级大纲》

等级 \ 关联词	甲级	乙级	丙级	丁级
因果关系 (20 个)	……, 因此……、 因为……所以……、 ……只好……	由于……, ……、 ……是由于……、 ……, 因此……	之所以……是因为……、 由于……以致……、 由于……的缘故、……可见……、 由于……(所以)……、由于……就……、 既然(既)……只好(只有、一定)……、 ……以致……、……不免……、	(因为)……, 故……、……, 以至于……、所以……, 所以……, (还)因为……、……因为……

将以上两表的因果关联词归纳整理, 例如: “由于……, ……”、“……因此……”合并改写为“由于……因此……”; “由于……, ……”、“……以致……”合并为“由于……以致……”; “既然(既)……只好(只有、一定)……”与“……, 只好……”合并为“既然……只好……”。最后总结《大纲》中因果关联词类型见表 3:

Table 3. Table of causal connective word types
表 3. 因果关联词类型表

(因为)……所以、(故)……	……, 因此……	既然……就……	之所以……是因为……
(由于)……以致……	……, 以至于……	所以……, 所以……, 还因为	既然……只好……
……是由于……	由于……的缘故	……, 可见……	由于……就……
……, 不免……	……因……		

(二) 研究语料

HSK 动态作文语料库是母语非汉语的外国人参加高等汉语水平考试(HSK 高等)作文考试的答卷语料库, 收集了 1992~2005 年的部分外国考生的作文答卷。国家语委主持建立的现代汉语语料库作为通用型语料库, 作为与 HSK 相对应的参照语料库。

(三) 研究过程

通过 HSK 动态作文语料库与现代汉语语料库检索出 14 个因果关联词分别在两个语料库中出现的频数。然后比较语料库检索项目的频数的差异, 用卡方检验方法检验差异显著性。最后分析检索项目在两个语料库中使用的特征, 探讨导致这些差异和特征的原因, 从而为汉语教学与教材编写提供一些建议。

3. 数据分析及讨论

以标记理论[4]为指导将上述 14 种因果关联词进一步分为一般因果式和特殊因果式。一般因果式主要表现为两种及以上因果标记的形式, 主要有三种情况: 一是需要使用关联词; 二是可以加上关联词; 三是不需要使用关联词[5]。例如同时出现“因为”和“所以”的复句, “因为”和“所以”可以省略其中一个, 同样可以表达因果关系。特殊因果式主要表现为只有一种因果标记的形式, 如“……, 因此……”。在分类的基础上将 13 种因果关联词放入语料库中进行检索, 检索前对有些因果关联词的词性做出区分。经过整理得到以一般因果式与特殊因果式为区分的因果关联词频数表(表 4)。

Table 4. Causal association word frequency and Chi-square test table
表 4. 因果关联词频数与卡方检验表

Corpus Size	Corpus 1		Corpus 2		
	50,000,000	4,240,000	Chi-Square	Significance (<i>p</i>)	
Word	Freq in Corpus 1	Freq in Corpus 2	Chi-Square	Significance (<i>p</i>)	
因为	8969	8667	41812.4888	0.000***	-
所以	7519	10,088	59835.5292	0.000***	-
因为/所以	807	1488	10355.0275	0.000***	-
因为	8969	8667	41812.4888	0.000***	-
故	1267	69	13.0440	0.000***	+
因为/故	70	5	0.1378	0.000	+
既然	839	145	65.3658	0.000***	-
就	44,145	11,784	13645.2560	0.000***	-
既然/就	318	35	2.1560	0.142	-
之所以	540	226	499.9535	0.000***	-
是因为	1607	818	2260.1229	0.000***	-
之所以/是因为	167	1	12.1594	0.000***	+
由于	8074	1227	372.9649	0.000***	-
以致	351	58	22.9861	0.000***	-
由于/以致	49	5	0.1559	0.693	-
既然	839	145	65.3658	0.000***	-
只好	997	225	190.3765	0.000***	-
既然/只好	10	1	0.0248	0.875	-
……, 因此……	7416	2304	3404.9381	0.000***	-
……, 以至于……	86	10	6.8280	0.009**	-
……, 是由于……	1171	818	3062.5248	0.000***	-
由于……缘故	86	10	0.9003	0.343	-
……					
……, 可见……	273	82	115.0444	0.000***	-
……, 不免……	312	29	0.2235	0.636	-
……因……	3525	1934	5775.8084	0.000***	-

由表 4 可以看出, 卡方值(自由度为 1 时)如果大于临界值 3.83、6.64 和 10.83, 则表明该值在 0.05、0.01 和 0.001 的显著水平上是有意义的, 即参与比较的两个数值具有显著性差异。其次卡方检验工具提供了与卡方值对应的显著性水平的 P 值, 并按所处的置信区间标定星号(*), 在 0.05、0.01 和 0.001 显著性水平上分别标为*、**、***表示其重要程度。最后一列的“-”表明“因为、所以”在现代汉语语料库

中的使用频数要少于其在 HSK 中的频数, 即使用不足。反之“+”表示过多使用。

(一) 一般因果式

1) 因为……所以……

经统计, “因为”在现代汉语语料库中出现的频数是 8969 次, “所以”的频数是 7519 次, “因为/所以”共现的频数只有 807 次; “因为”在 HSK 动态作文语料库中出现的频数是 8667 次, “所以”的频数是 10,088 次, “因为/所以”共现的频数是 1488 次。

鉴于“因为/所以”是使用最频繁的一组因果式[6], 为了具体比较两个语料库中三个项目, 统计出三个项目分别在语料库中的标准化频率(见表 4)。标准化频率即基于一个统一基准得出的频率。

表 5 中, “因为”在现代汉语语料库的标准化频率是 0.17938, “所以”的标准化频率是 0.15038。而在 HSK 动态作文语料库中, “因为”和“所以”的标准化频率分别是 2.044 与 2.38, 即每一千词出现 2 次“因为”、“所以”。可以看出留学生较中国人更经常使用“因为”与“所以”, 组织话语倾向于强调语句前后的逻辑性。最后“因为/所以”共现的标准化频率在现代汉语语料库中仅为 0.016, 而在 HSK 动态作文语料库中是 0.35, 高于现代汉语中的频率。这说明留学生更倾向在组织话语时使用“因为……所以……”这一关联词, 可能是因为汉语教学强化了关联词连接句子的作用, 经过大量的操练使得留学生固化了这一关联词的用法, 导致“因为/所以”共现使用的频数的提高。

Table 5. Standardized frequencies in both corpora because/so

表 5. 因为/所以在两个语料库中的标准化频率

	现代汉语语料库	HSK 动态作文语料库
因为	0.17938	2.044103774
所以	0.15038	2.379245283
因为/所以	0.01614	0.350943396

结果显示, “因为、所以、因为/所以”三个项目在现代汉语语料库中出现的频数与在 HSK 动态作文语料库中有均有显著差异, 说明留学生学习“因为/所以”时倾向于使用更具“形合”特征的语法结构组织话语, 在话语中过度使用大量因果标记形式。观察语料后总结出以下 3 个特征:

a) 语义重复: 现在我可以写汉字, 因为小时候每天都写汉字, 所以就不会忘记汉字怎么写, 按笔顺, 就写对了。

b) 成分多加:

因为中国的历史很长有五千年的文化, 人们很友好所以值得来。

因为哈尔滨的风景, 特别是冰灯漂亮得不得了, 而且导游对我很热情, 所以我高兴极了。

c) 语法误用: 因为没有别的人, 所以不能靠别人, 于是他自己要到山底下挑水喝。

例(a)“所以”和“就”同时承担了结果标记的功能, 导致重复使用。

例(b)可以将“因为”去掉, 弱化原因强调结果。

例(c)留学生将“不能靠别人”当成“没有别人”的结果, 而真正的结果是要承担自己下山挑水。“没有别人”和“不能靠别人”是“自己下山挑水”共同的原因。“没有别人”与“不能靠别人”之间有内在的因果关系, 可以用“也”来表示。

2) 因为……故……

表 5 所示: “+”表示“故”在现代汉语语料库中比在 HSK 动态作文语料库中过度使用, “因为/故”共现时也是如此, 这说明中国人比国外留学生更经常使用“因为……故……”这种形式。

3) 既然……就……

“就”的使用频数很高，“既然”反而很低，反映在语料库中两个检索项的频数差异依旧显著，表明留学生使用的频数没有中国人高。但是“既然/就”共现的频数在两个语料库中均反映不足且差异并不显著，这说明我们一般认为的“既然”后会跟“就”在实际使用中并不常见。

4) 之所以……是因为……

“之所以”、“是因为”卡方检验结果显示“-”则表示其差距具有显著性且此检索项的使用频率更倾向于留学生，相比中国人留学生更倾向于单独使用“之所以”、“是因为”。但是考察检索项共现的频数发现，中国人较留学生更倾向于将检索项组合在一起使用。

5) 由于……以致……

根据卡方表可以发现中国人和留学生都倾向于将其拆开单独使用，中国人单独使用此因果式的频率依旧没有留学生使用的高。反观它们共现的卡方值可以证明中国人和留学生都不经常将两个关联词放在一起共现使用。

6) 既然……只好……

从频数表可以看出它们共现的频数都很少，卡方检验不存在显著差异，因此这一组因果式关联词在使用中跟“由于……以致……”一样倾向于拆分使用。

(二) 特殊因果式

特殊因果式是只有一个语法标记的关联词形式。特殊因果式的卡方 P 值均呈现显著状态，说明留学生使用特殊因果式比中国人使用的频率高，因为特殊因果式没有“形合意合”转换的问题，从语法层面来看其只是起连接作用的语法标记形式，具有“形合”特征，这符合留学生的学习与认知习惯。下面的句子是留学生使用特殊因果式的实例，主要特征有“成分多加”、“语义重复”、“语序错位”、“逻辑混乱”等。

1) ……，因此……

- a) 自七岁那年起，我就读了本地一所有名的小学，而且学习成绩很好，因此认识了很多朋友。
- b) 由于我们学校的关系，因此父亲一个人去外地做工作。

例 1 中“我认识很多朋友”跟“我在有名的小学读书”和“我成绩很好”之间并无因果关系，“因此”完全可以省略；“学校的关系”前所用的“由于”和后面“因此”重复，可以删去“因此”。

2) ……，以至于……

- a) 现在我在北京工作，以至于想扎根。
- b) 也就是说，要读懂，要透彻地了解，以至于能灵活应用。

例 2 中“我在北京工作”是一个陈述性的表述，与“想扎根”构不成强烈的因果关系。第二句话“能灵活运用”与“要读懂”和“要透彻了解”之间逻辑混乱，“以至于”用得并不恰当。

3) ……是由于……

- a) 目前最严重的问题是由于别人在旁边吸烟因此自己不由得吸他的烟气，所谓“二手烟”。
- b) 是由于我个人在周围的探察，如：由亲戚朋友的介绍，那是以他(她)们的理想、眼光来判断。

例 3 中的“是由于”均为过度使用，第一句没有必要加“是由于”；第二句“是由于”出现在句首，而“是由于”一般出现在句中。

4) 由于……缘故

几十年来，由于职业的关系，再加上兴趣的缘故，所读过的书的数量是颇为可观的！

例 4 对由于……缘故用法很熟练，知道可以将其拆开来说，大部分留学生都将“由于……缘故”放在句首，或者会在“由于”前加“这是”成“这是由于……缘故”。

5) ……，可见……

a) 由此可见，政府规定严厉惩罚在公共场所吸烟的人是正确的决定。

b) 可见，父母忽视教育责任的后果是很严重的。

例 5 中除了“……，可见……”这一形式外，留学生经常会和“由此”一起组合成“由此可见”这一固定搭配，或者直接在句首使用“可见”，再引出所要说的结论。

6) ……，不免……

a) 当初虽说我己是有心理准备的，然而还是不免吃了一惊！

b) 首先，在子女的立场而言，父母始终是自己的父母，有些时候，父母啰嗦是少不免的。

例 6 留学生对“不免”的位置掌握不到位，错将其放在句末，并不符合汉语语法；

7) ……，因……

a) 幼时因跟随父母离乡谋生，在第二个弟弟未出世前，我们等于是半流浪式地四处为家。

b) 过后我难过了几天，因体会到她是为了省一笔钱，留下做菜钱的。

例 7 第一句“跟随父母离乡谋生”和“半流浪式四处为家”之间不构成直接的因果关系，又加了一句“在第二个弟弟未出世前”，更加弱化了因果关系；第二句留学生将原因至于结果后。

4. 结语

根据以上对留学生用例的分析，其因果关联词的过度使用的特征主要可以总结为：成分多加、语义重复、语序错位、语用不当、逻辑混乱五种。造成过度使用的原因主要有两个：一为留学生母语的负迁移，主要是英语为母语的留学生受母语的影响更大，将英语作为第二外语的留学生可能也会受到干扰；二为目的语的泛化，教师在备课时应注意教学材料的语言项目是否体现“形合意合”特征，教学时应按照学生水平，有的放矢地说明汉语中“意合”的特征，尽量避免体现语言结构标记项目的教学，鼓励学生运用“意合”法去组织话语。

基金项目

【2017】国家社科基金重点项目“中国古代蒙学典籍在海外的传播及其影响研究”(编号:17AZS012)。

参考文献

- [1] 潘文国. 汉英语对比纲要[M]. 北京: 北京语言大学出版社, 1997: 5.
- [2] 王力. 中国语法学初探[M]. 北京: 商务印书馆, 2018: 8.
- [3] 任晓霏. 从形合和意合看汉英翻译中的形式对应[J]. 中国翻译, 2003(3): 56-60.
- [4] 董佳. 汉语因果复句的原型表达[J]. 陕西师范大学学报(哲学社会科学版), 2012(3): 141-145.
- [5] 黄戈恒, 黄兰堞. 汉英关联词语隐现分析[J]. 牡丹江大学学报, 2018(6): 63-65.
- [6] 陆琳. 中高级水平留学生“因为……所以……”习得情况研究[J]. 语文学刊, 2010(2): 9-12.