

语料库视角下的物流英语语篇中过渡词研究

柏 薇

上海海事大学外国语学院, 上海

收稿日期: 2023年2月28日; 录用日期: 2023年3月30日; 发布日期: 2023年4月7日

摘 要

过渡词是一种很好的语篇衔接方式。本文拟从词汇密度、词频、空间分布、语义分布四个角度, 对物流英语语篇中的过渡词语进行基于语料库的定性与定量分析, 揭示其在物流英语语篇中过渡词的使用特征。笔者期望本研究能为我国物流英语专业的学生在学习和运用英语方面提供一些参考。

关键词

语料库, 物流英语语篇, 过渡词

Research on Transition Words in Logistics English Discourse from the Perspective of Corpus

Wei Bai

College of Foreign Languages, Shanghai Maritime University, Shanghai

Received: Feb. 28th, 2023; accepted: Mar. 30th, 2023; published: Apr. 7th, 2023

Abstract

Transition words are a good way to connect the discourse. This paper intends to make qualitative and quantitative analysis of transition words in logistics English discourse based on the corpus from the perspectives of vocabulary density, word frequency, spatial distribution and semantic distribution, to reveal the characteristics of the use of transition words in logistics English discourse. It is hoped that this study will have implications for the learning and use of professional English in logistics.

Keywords

Corpus, Logistics English Discourse, Transition Words

Copyright © 2023 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

由于传统的语法侧重于个别的句子或词汇研究,因此,在语法体系中,句子是最上层的结构。而语篇分析则以句子关系,语篇中的过渡与连贯为重点。从篇章分析的视角来看,篇章并非单纯的语句的组合,它是由多个具有一定联系的语句所组成的一个整体。为了达到篇章的连贯性,有许多方法可供选择,其中之一就是过渡词。

2. 研究的问题

本文主要从四个角度来探讨过渡词。

2.1. 词汇密度

词汇密度(TTR)是衡量篇章难度的一个重要指标。由于 TTR 作为一个样本函数,受到文本长度的影响较大,因此笔者选择对长文本进行分段,然后对每个段落进行分析,这样可以分散文本长度带来的影响。在这篇文章中,TTR 是一个特定篇章中过渡词语的类符号与形符号的比例。通过类符/形符可以一定程度上反映文本的特征,即用词的变化性[1]。词汇的密集程度越高,其所包含的信息就越多,其表达的意义也就越严密。在外语学习中,可以用来表示文章的难易程度。所以,在编写教材的时候,要把词汇的密度问题考虑进去。

2.2. 词频

“高频词”是一种普遍使用的文体标记。高频率词汇既可以用来展示文章的内容,又可以用来做阅读前的预设“模式”。所以,很多学者都在努力构建能够用基本英语表述各种思想的“核心词汇”,或者是为以后的阅读做好准备。词汇在各种文章体裁中出现的频率是不一样的。

2.3. 空间分布

过渡词在句子中的空间位置也有变动,可以在句子的开头,中间,结尾三种情况。过渡词在所有空间位置都可以使用,但在特定空间位置使用的几率要大得多。在词位上,词位的不同会引起词义的差异,词位的选取会受到具体因素的影响。1) 句子开头。在句子的开头,可以清楚地将语篇的各个单元联系起来。过渡词可以使读者准确地了解后文与前文的关系。因此,句子的开头一般都不带标志的。2) 句子中间。Quirk [2]曾说过,“句中”是一种介于“主题”和“陈述”两个部分之间的“位置”,其实“句中位置”是一种非常灵活的划分方式。在这篇文章中,更多的具体位置将被称为“句子中间”。人们普遍认为,在学术文章中,过渡词语往往位于句中。句子结尾。句子的结尾是句子中所有必需的成分后面的位置。句子结尾处的过渡词语,大部分都是与会话有关的。在学术文章中,一般很难在句子的末尾使用过渡语。

2.4. 语义分布

许多语言学者对过渡词的语义分类进行了探讨, 包括 Halliday [3]、Hasan [3]、Quirk [2]、Biber [4] 等。通过多次对比, 作者认为, Quirk 与 Biber 等人所作的分类法较为相近, 且更符合当前的研究现状。由于已有文献资料, 因此, 本论文主要使用了 Quirk 等的分类方法。本文研究的对象如表 1 所示。

Table 1. The object of this research

表 1. 本文的研究对象

语义类	过渡词
增加	also, and, and then, too, in addition, furthermore, moreover, again, on top of that, another, first, second, third
时间顺序	now, then, before, after, afterwards, earlier, later, immediately, soon, next, in a few days, gradually, suddenly, finally
空间顺序	near (to), far (from), in front of, behind, beside, beyond, above, below, to the right/left, around, outside
比较	in the same way, just like, just as
结果和原因	because, since, so, as a result, therefore, then, thus, otherwise
目的	for this reason, for this purpose, so that
强调	actually, in fact, as a matter of fact, in effect, of course, indeed, apparently, surely, necessarily, certainly, without any doubt, truly, to repeat, above all, most important
解释说明	for example, in fact, in this case, for actually
总结	finally, at last, in conclusion, as I have shown, in other words, in brief, in short, in general, on the whole, as has been stated, anyway, in a word
并列	and, also, as well, as well as, or, too, not only...but also, both...and, either...or, neither...nor
陈述事实	in fact, actually, as a matter of fact, to tell you the truth

3. 研究方法

这部分主要阐述了本文所需的语料和所用数据的处理方法。

3.1. 语料库

本文主要选用两个语料库: LEC (Logistics English Corpus, 物流专业英语语料库)和 BNC (British National Corpus, 英国国家语料库)。

1) LEC。本研究收集了 Donald F. Wood 编著的 *International Logistics* 和 Nchandrasekara 主编的 *Fundamentals of Logistics* 两本书, 以国家物流标准化技术委员会和国家物流信息管理标准化委员会提出并统一的《中华人民共和国国家标准: 物流专业英语》(GB/T 18354-2006)为参照专业词汇, 计算出两本书中专业词汇占比共为 21.91%, 因此选取者两本书在相关在线专业英语提取网站, 提取了共 4008 条物流专业术语, 自建了一个小型的物流专业英语语料库。

2) BNC。从 BNC (英国国家语料库)语料库中的书面英语(库容 9000 万词)提取共 916,575 个词为参照语料库。

3.2. 数据处理

1) 确定研究对象。如前面所提到的, 本论文以 Quirk 等人的分类法为依据, 对 112 个过渡词进行了

修正。这些过渡词正是本论文所要探讨的问题。

2) 处理语料。在 LEC 中, 笔者删除了所有的图表, 不重复下载站内链接, 然后将 Word 文档转换为纯文本文档进行检索。BNC 是一个有代码的语料, 必须使用 FOX PRO 对代码进行拆分。

3) 搜索关键字。本文使用了两种不同的语料, 通过 Antconc3.30 对 112 个过渡词进行了搜索, 并得出了它们在不同的上下文中出现的情况。

4) 手动操作。目的是将那些不属于过渡词的词语剔除掉, 例如那些在句子中充当连接词或附加词的词语, 为了确保得到的资料的正确性, 作者将这些词语进行了两次人工操作。

4. 研究结果

这部分从四个角度对研究的结果进行了探讨: 词汇密度, 词汇频率, 空间分布以及语义分布。

4.1. 词汇密度

本文从以下两个方面研究过渡词在两个语料库中的词汇密度: 一是过渡词在两个语料库中数量的标准值, 即每 100,000 个单词中过渡词出现的数量; 二是 TTR, 即过渡词的类符形符之比。统计结果如表 2 所示。

Table 2. Statistical result

表 2. 统计结果

语料库	LEC	BNC
语料库形符	1,150,346	84,145
过渡词形符	9477	1024
标准值	824	1217

注: 标准值 = 过渡词形符数/语料库形符数*100,000。

如表 2 所示, LEC 中形符数和过渡词形符数都比 BNC 中的多, 但是过渡词在 LEC 中的标准值却比其在 BNC 语料库中的数值小。这表明, 与 BNC 相比, LEC 中的过渡词要少一些。过渡词的 TTR 结果如表 3 所示。

Table 3. The TTR results for the transition words

表 3. 过渡词的 TTR 结果

语料库	LEC	BNC
类符	60	96
形符	9477	1024
TTR	0.006	0.093

注: TTR = 类符/形符。

从 TTR 的角度来看, LEC 语料中的 TTR 值为 0.006, BNC 语料中的 TTR 值为 0.093, 因此, LEC 语料中的 TTR 值仍然远远低于 BNC。从这一点可以看出, LEC 中的过渡词的词汇密度要比 BNC 小得多。

4.2. 词频

高频词反映了语言的共性, 中频词汇的使用情况可以反映出实体内容和语言使用者的水平, 低频词可以看到文本的特点。词频统计是语料库研究的基本统计手段[5]。在词汇的频度上, 我们可以从如下角度进行考察: 1) 从两个语料中选取 10 个最常用的转化词; 2) 与 BNC 相比, LEC 中出现了大量的过渡词; 3) 与

BNC 相比, LEC 中的过渡词用得太多。本文将两个语料库中的过渡词转换成标准词, 以十万个词语为单位, 对两个语料库中的过渡词语进行统计。其中, 在两个语料库中使用频率最高的 10 个过渡词如表 4 所示。

Table 4. Top 10 transition words in LEC and BNC

表 4. LEC 和 BNC 中前 10 个过渡词

过渡词	LEC		过渡词	BNC	
	数量	比例/%		数量	比例/%
now	876	8.97	now	171	16.63
so	766	7.84	then	155	15.08
as well	630	6.45	so	101	9.82
thus	630	6.45	because	71	6.91
suddenly	575	5.89	too	40	3.89
after	537	5.50	and	36	3.50
finally	444	4.54	first(ly)	34	3.31
for example (e.g.)	376	3.85	also	33	3.21
first (ly)	367	3.76	as well	32	3.11
then	339	3.47	of course	28	2.72
合计	5540	56.72	合计	701	68.18

注: 比例是指本单词在本语料库所有过渡词中所占的比例。

Table 5. Overused transition words in the LEC

表 5. 在 LEC 中过多使用的过渡词

过渡词	LEC	BNC
	T/100,000	T/100,000
so	35.39	0
as a result	18.17	0
gradually	7.49	0
actually	6.14	0
in general	3.99	0
in effect	3.99	0
in the same way	3.51	0
in fact	3.35	0
for this purpose	2.39	0
certainly	2.31	0
above all	2.31	0
therefore	2.23	0
as a matter of fact	2.07	0
in conclusion	2.07	0
as a consequence	2.07	0

由表 4 可知: 1) 在两个语料库中, 前 10 个过渡词所占的比重均超过一半, 并且在 LEC 中所占比重低于 BNC; 2) 在所用次数最多的 10 个过渡词中, LEC 中的最大值与最小值之差为 5.50% ($8.97\% - 3.47\% = 5.50\%$), 而 BNC 中的最大值与最小指之差为 13.91% ($16.63\% - 2.72\% = 13.91\%$), 比 LEC 中的差异大得多, 说明 LEC 中对单个词语的依赖程度并不高。3) 从两组数据中还可以发现, “now, so, then” 是两组中共有而且使用频率比较高的 3 个词。可以得知, 两种体裁依赖的单词是不一样的。相比于 BNC 语料库, 在 LEC 中过多使用的过渡词统计结果如表 5 所示。

从表 5 可以发现, 作者所收集到的 15 个词汇都是在 LEC 中出现频率很高而在 BNC 中未出现的词汇。与之相比, 在物流英语的篇章中, 这 15 个词的用法被高频使用。这些句式可分为六大语义类, 其中以罗列补充和比较性让步最为突出。

此外, 笔者还统计了在 LEC 中过少使用的过渡词, 结果如表 6 所示。

Table 6. The underused transition words in the LEC

表 6. 在 LEC 中过少使用的过渡词

过渡词	LEC		BNC	
	T/100,000		T/100,000	
also	0		37.36	
besides	0		4.53	
well	0		23.77	
anyway	0.16		115.85	
somehow	0.24		9.06	
likewise	0.24		5.66	
next	0.64		12.45	
then	0.80		14.72	
apparently	0.96		11.32	

其中 3 个单词未在 LEC 中出现, 另外 6 个在 LEC 中出现的频率均远低于 BNC。经过对这 9 个词的义项的分析, 作者发现 9 个词的义项也都是以罗列的形式出现。从这两个表中我们可以看出, 这两个语料对特定过渡词语的运用有很大的差异。

4.3. 位置分布

笔者在本文中将过渡词的位置分为句子开头、句子中间和句子结尾 3 种。统计结果如表 7 所示。

Table 7. Statistical results of position distribution

表 7. 位置分布统计结果

位置	LEC		BNC	
	数量	比例/%	数量	比例/%
句前	2000	19.0	256	22.5
句中	7321	79.3	688	69.8
句尾	156	1.7	80	7.7
合计	9477	100	1024	100

注: 比例是指此位置的过渡词占全部总数的比例。

从表 7 中我们可以看出, 与 BNC 相比, LEC 在句中的过渡词地位显著高于 BNC, 而在句后的地位却低于 BNC。正如前面提到的, 在较高的学术论文中, 过渡词更多地被用在句首的位置。从这一点可以看出, 物流英语的语篇在学术上要优于一般的英语。

4.4. 语义分布

笔者将 Quirk 等人研究的 112 个过渡词分为 7 类, 并分别对其进行检索统计, 结果如表 8 所示。

Table 8. Semantic distribution of the transition words in the two corpora

表 8. 过渡词在两个语料库中的语义分布

语义类	LEC		BNC		Chi-square test	
	Tokens	T/100,000	Tokens	T/100,000	χ^2	P
增加	2590	216.64	354	374.71	871.268	0.00
时间顺序	180	15.78	7	7.92	333.837	0.00
空间顺序	623	51.17	21	26.04	102.776	0.00
比较	2306	183.80	208	235.47	640.275	0.00
结果和原因	2178	173.60	200	245.66	637.406	0.00
目的	1000	79.71	194	219.62	390.034	0.00
强调	600	58.03	40	54.34	170.044	0.00
合计	9477	778.73	1024	1163.76	3145.64	0.00

注: T/100,000 是指过渡词每 10 万字出现的数量。

从表 8 中的数据对比可以看出: 1) 在两种语料的语义分布上有相似的趋势, 均侧重于列举附加、结果推断、对比让步和转折 4 个义项, 而且在 4 个义项中, LEC 所含的过渡词数目显著低于 BNC; 2) 其余 3 种语义组所占比重较小, 但其中 LEC 中的过渡词的数目略大于 BNC。

5. 结语

本文的结论是: 与一般的英语相比, 物流英语中的过渡词数目较少, 且密集程度较低。尽管与一般英语一样, 都对前 10 位的过渡词依赖性很强, 但物流英语依赖性很小, 而且依赖于不同的词汇。就位置分布而言, 物流英语篇章中位于句中的过渡词比一般英语多得多, 而位于句尾的过渡词比一般英语少得多。从语义上看, 物流英语与一般英语中的过渡词主要分为 4 个类别: 罗列补充、比较让步、结果推断、转折; 另外, 在 4 个语义类中, 物流英语篇章中的过渡词比一般英语要少, 其它 3 个语义类中, 物流英语篇章中的过渡词比一般英语稍多一些。

虽然笔者已经做了周密的规划与准备, 但仍有一些不足之处。1) 语料的大小、代表性等因素对研究的信度、有效性有一定的影响, 因此, 语料的大小、样本大小等因素也会对研究的结果产生一定的影响。2) 由于本论文是在 Quirk 等的基础上进行的, 因此忽视了其它学者的工作。3) 本论文仅考察了篇章间的过渡关系, 对其它关系的探讨还需深入探讨。4) 手动过程中难免会出现主观判断失误。本研究表明, 过渡词在不同的英语体裁中有不同的使用特点, 因此, 在物流英语教学中, 教师和学生都要注意这些差别, 并且要多接触物流英语, 这样才能够熟练准确地应用过渡词。

参考文献

- [1] 王芳, 连天雪. 基于语料库的商务英语与普通英语的词汇比较研究[J]. 大连理工大学学报(社会科学版), 2013,

34(3): 130-133.

- [2] Quirk, R., Greenbaum, S., Leech, G., *et al.* (1985) *A Comprehensive Grammar of the English Language*. Longman, London.
- [3] Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English*. Longman Pub Group, London.
- [4] Biber, D., Johansson, S., Leech, G., *et al.* (2000) *Longman Grammar of Spoken and Written English*. 2nd Edition, Foreign Language Teaching and Research Press, Beijing.
- [5] 梁茂成, 李文中, 许家金. 语料库应用教程[M]. 北京: 外语教学与研究出版社, 2010.