

基于语料库的物流专业英语词汇特征研究

王善吉

上海海事大学外国语学院, 上海

收稿日期: 2022年2月21日; 录用日期: 2022年3月18日; 发布日期: 2022年3月25日

摘要

基于语料库, 通过使用AntConc3.5.8和WordSmith4.0作为语料分析以及处理工具, 从词汇密度、平均词长、高频词、搭配强度四个层面分析物流专业英语词汇特征。研究表明: 物流专业英语的词汇密度较高, 语篇难度大于普通书面英语; 名词和名词化短语居多且搭配强度高。希望本研究对物流专业英语的学习及使用、物流大纲设计及教材编写具有一定的意义。

关键词

物流专业英语, 词汇密度, 平均词长, 高频词, 搭配强度

A Corpus-Based Study on Lexical Features in the Logistics English

Shanji Wang

College of Foreign Languages, Shanghai Maritime University, Shanghai

Received: Feb. 21st, 2022; accepted: Mar. 18th, 2022; published: Mar. 25th, 2022

Abstract

Based on the corpus, AntConc3.5.8 and WordSmith4.0 were used as corpus analysis and processing tools to analyse the lexical features of English for logistics professionals in terms of lexical density, average word length, high-frequency words and collocation intensity. The study shows that the lexical density of professional English for logistics is high and the discourse difficulty is greater than that of ordinary written English; nouns and nominalized phrases are predominant and collocation intensity is high. It is hoped that this study will have implications for the learning and use of professional English in logistics, the design of logistics syllabuses and the development of teaching materials.

Keywords

English for Logistics Professionals, Vocabulary Density, Average Word Length, High-Frequency Words, Collocation Strength

Copyright © 2022 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

词汇是语言的基石。近年来,我国物流业发展迅速,正逐步与国际接轨,实现国际化;因此,掌握物流行业的国际通用语言——英语变得越来越重要。物流英语经过多年的发展逐渐形成了自己的特色词汇,但因物流英语词汇专业性强,覆盖面广,给学习者带来了很大的困难[1]。目前我国物流专业英语研究主要是定性分析,例如,研究物流英语的词汇特征及其翻译策略,还有有关物流专业英语教学策略,鲜有基于语料库的实证研究,例如,物流英语的词汇密度是多少?高频词有哪些?是否高频词的搭配强度很高?本文将通过语料库探讨物流专业英语在词汇密度、平均词长、高频词和词汇搭配等信息上的特征,旨在为物流专业英语学习和物流英语大纲设计及材料编写上有一定的帮助。

2. 数据来源及研究方法

Chung 和 Nation 将词汇分为高频词、学术词、专业词和低频词四大类[2],其中专业词是指在特定语境下出现,且体现特定学术领域含义的词汇。一般来说,在一篇给定的文本中专业词汇占总词汇 5%~30%。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 3016 条物流专业术语,自建了一个小型的物流专业英语语料库,然后从 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;以 AntConc3.5.8、WordSmith4.0 和在线专业英语提取网站为语料分析以及处理工具,计算出样本库中的单词总数、词汇量、高频词、平均词长、Log-log 等信息。本研究采取“定量分析为主,定性分析为辅”的方法,可以帮助我们更加全面、客观、真实地了解物流专业英语。

3. 研究结果与讨论

3.1. 词汇密度

词汇密度(lexical density)是用来衡量单位文本信息内容的尺度,因此它决定了文本的难度[3]。类符形符比(Type/Token Ratio, 简称 TTR)通常用来计算文本的词汇密度[1];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[4]。URE 是最早研究词汇密度的人之一,采用 $TTR = \text{类符/形符} \times 100\%$ 的公式[5]。但因为英文文本中的功能词居多,因此向文本中每添加一个词,形符就会增加,但类符不一定会随之增加,当文本越长,功能词越多,类符形符比随之降低。因此,使用类符/形符来计算长度不等的文本的词汇密度是不合理的,一般情况下,对于超过 1000 个形符的语料库,应使用标准化类符/形符

(standardized type-token ratio, 简称 STTR), 即按一定的长度(通常是 1000 个形符)分段计算的类符形符比, 也被认为是比较可靠的词汇密度的计算工具。本文采用标准化类符形符比来计算词汇密度, 观察物流专业英语的难易程度, 这里将物流专业英语与 BNC 书面英语进行比较, 从而更好地观察。结果如表 1 所示。

Table 1. Comparison of lexical density between the Logistics English and BNC written English

表 1. 物流专业英语与 BNC 书面英语词汇密度比较

样本库	物流专业英语	BNC 书面英语
类符	1479	33602
形符	5929	916575
类符/形符	25%	3.67%
标准化类符/形符	52.20%	46%

表 1 结果显示, 物流专业英语的类符形符比远远的高于 BNC 书面英语, 因为笔者在对物流英语进行专业英语提取时, 严格遵循专业英语提取语法, 专业英语语法是一组用 CQL 编写的规则, 英语的专业英语语法将专业英语定义为名词和形容词的序列(名词 + 名词 + 名词, 形容词 + 名词, 形容词 + 形容词 + 名词等), 从中剔除了大量的虚词及低频词, 所以得出的类符形符比较高。标准类符形符比也远远大于普通书面英语, 该差异表明物英的词汇变化程度远远高于书英。由此可知, 在物流英语中, 词汇密度高表明物流专业英语中实词所占比例大, 因此信息量也大, 难度也相应增加。

3.2. 平均词长

平均词长(mean word length)是反映文本文体特点的一个重要因素, 也是决定文本难度和复杂性的重要因素。通常有两种方法计算: 一种是根据类符字母总数与类符总数的比值, 即词汇量的平均长度, 计算样本库的词目; 另一种方法是根据语料库字母总数与形符总数(即单词总数)比值, 计算样本库中的形符 [6]; 词长标准偏差是指单词长度在平均词长周围浮动的程度。本文采用第二种方法来计算平均词长, 也就是字母总数与形符总数的比。通常情况下, 正式文本的文体常用长词来突出文本的文体特征, 以区别非正式或口语化文体的文本。以 Brown 语料库中三种不同文体的文本为例, 得出表 2 的数据。

Table 2. Mean word length and standard deviation of word length for three different text corpora in the Brown corpus

表 2. Brown 语料库三种不同文本文体平均词长和词长标准偏差

文本类型	平均词长	词长标准偏差
报刊类(Categories A, B, C)	4.73	2.59
学术类(Categories J)	4.93	2.87
各类小说(Categories K, L, M, N, P, R)	4.31	2.18

从表 2 中, 可以看出: 在这三类文本中, 学术文本是最正式的, 其次是报纸和期刊, 然后是小说。可以看出, 平均字长与文本的正式程度密切相关。通过对物流专业英语语料库的计算, 类符为 1419, 字母总数为 40793, 平均词长为 6.88, 词长标准差为 2.8。与表 2 相比, 可以看出, 物流专业英语的词长远大于三个文本, 这说明物流专业英语用词的复杂程度远远高于其他文本, 长词较多。大量的使用长词, 这反映了物流英语词汇的特点。

我们知道名词化是科技英语的一个显著特征, 这在物流英语中也很突出[7]。理论上, 英语是以动词为中心的, 但事实上, 名词占主导地位; 从理论上讲, 汉语着重于名词, 但实际上, 动词有一定的优势。

这两种违反民族心理的语言事实是由各自语言的特点决定的[8]。物流专业英语中有许多复合名词，也就是名词化短语，名词连词是合成名词的一种重要形式，它是由语言的经济性决定的。名词连词的形式是“名词 + 名词”结构，也就是说在名词头之前可以有多种形式不变的名词。它们是头部的介词形容词修饰语，被称为“扩展的名词前置修饰语”[9]。

Table 3. Structure and frequency of the Logistics English
表 3. 物流专业英语结构及频率

Patterns	Frequency (%)
Common_NounCommon_Noun	977 (32%)
Common_Noun	833 (28%)
Adjective Common_Noun	758 (25%)
Common_Noun Preposition Common_Noun	208 (7%)
Adjective Common_NounCommon_Noun	78 (3%)
Common_NounCommon_NounCommon_Noun	69 (2%)
Common_Noun Preposition Common_NounCommon_Noun	17 (1%)
Common_Noun Preposition Adjective Common_Noun	16 (1%)
Adjective Common_Noun Preposition Common_Noun	14 (0.5%)
Adjective AdjectiveCommon_Noun	13 (0.4%)
Adjective Coord_Conjunction Adjective Common_Noun	7 (0.2%)
Common_NounCommon_Noun Preposition Common_Noun	5 (0.2%)
Adjective Common_NounCommon_NounCommon_Noun	5 (0.2%)
Common_Noun Adjective Common_Noun	4 (0.1%)
Common_NounCommon_NounCommon_NounCommon_Noun	2 (0.1%)

笔者对专业英语又进行了结构分析，通过表 3 可以发现物流专业英语中名词化短语和名词居多，分别占比 32% (977)和 28% (833)。形容词性短语居第三，占比 25% (758)。名词连词的广泛使用可以有效地简化语言结构，避免过多的小句结构或介词短语结构，从而避免语言的复杂化；在物流专业英语中，这种结构层出不穷，这也就是物流专业英语的平均词长远高于其他文本的原因。

3.3. 高频词

梁茂成[10]认为，高频词(high-frequency word)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。Sinclair [11]指出，语料库中出现频率最高的单词的分布是稳定的，因此，如果顺序发生任何重大变化，可能会有意义的。通常，语料库词汇中的高频词大多是功能词，实词的出现表明实词所指的人、物或行为是语料库的核心，因此它会被反复提及。

通过对物流专业英语的检索得出了前 20 个高频词，分别为：logistic, cost, service, cargo, product, customer, vessel, time, firm, inventory, good, carrier, container, system, supply, chain, market, company, port, transportation。

通过前二十个高频词中，我们发现 firm 和 company 这两个单词，通常情况下，我们会将这两个词混为一谈。在传统的语言学研究中，人们常常孤立地去理解语言，将其分解为如语音、语法、词汇等。这种研究方法的缺陷在于它没能将语言中的各个方面进行内在的相互联系；在研究语义学时，使用的分析

方法主要集中在单词本省,比如确定词的一些语义属性,例如形容词、动词、名词等。然而,研究表明,在特定的语境中使用正确的单词,其意义才会明确;换句话说,一个词与其上下文的词连接成一个多词单元,意义才会明确[10]。通过语料库索引,观察发现:firm与之搭配词频率从高到低排序分别为:the, a, his, another, his, small, any, new, global, one, outside等;company与之搭配词频率从低到高排列顺序分别为:the, a, insurance, trading, leasing, peach, steamship, management, exporting, importing, manufacturing, trust, containership等。由此我们可以看出,company这个单词在物流专业英语中更加的专业,同时,笔者将这两个单词放入到自建的物流专业英语的语料库中进行检索,得出firm与global和outside搭配使用,而company与insurance, record, shipping, express, transportation, trucking, trading等连用,company的使用范围远远大于firm且更加的精准。

语料库数据驱动的优点之一是可以为学习者提供真实的学习环境。在这些索引行中,每行中的搜索词都在中间,并且在它们之前和之后有几个词,即搜索词的上下文[12]。在词汇学习中,学习者可以通过观察和分析节点词前后的语境,更准确地掌握物流专业英语节点词的具体含义和固定用法,并围绕节点词进行探索性学习;与此同时,还可以在上下文中比较一些意义相似、难以区分的单词,总结语言现象,加强对单词的理解和记忆,提高词汇使用的准确性。

3.4. 搭配强度

孤立地对待语言中的词存在着明显的缺陷,这在语言学领域被广泛接受。英国著名语言学家Firth在20世纪50年代提出了搭配(collocation)的概念,指出理解一个词需要通过它的伴词。换句话说,一个词的意义和它的搭配是相互影响的,而且搭配不是一个词的单方面行为。因此,在搭配强度的研究中,仅仅通过分析节点词的索引线来获得节点词的搭配是不够的。我们至少应该从它们各自的出现频率(occurrence)和共现频率(co-occurrences)两个方面来考察搭配的程度[13]。因此,为了避免搭配计算中偏重高频词(大多数高频词都是虚词)的问题,英国学者Adam Kilgarriff提出了Log-log搭配计算公式,主要在公式中加入了节点词和搭配词共现频数的对数值,从而降低高频词的权重,提高了其他实词搭配词的权重。如果我们使用搭配词进行内容分析的话,那么Log-log似乎更合适。

在平均词长中分析得到物流专业英语中名词和名词短语居多,因此推测是否在物流专业英语中名词和名词化短语的搭配强度比较高,因此在自建物流专业英语语料库中,运用Log-log计算其搭配强度,得出其中前十个搭配强度比较高的词如表4:

Table 4. Top 10 words with the highest collocation in the Logistics English
表 4. 物流专业英语搭配强度最高的前 10 个词

Candidate (grouping variant)	Log-log	Pattern
logistic	4935.2	Common_Noun
cargo	3397.11	Common_Noun
inventory	2910.17	Common_Noun
vessel	2522.09	Common_Noun
container	2477.71	Common_Noun
carrier	2472.17	Common_Noun
customer	2312.94	Common_Noun
supply chain	314	Common_NounCommon_Noun
shipper	2067.64	Common_Noun
good	1920.28	Common_Noun

通过分析,可以看出这十个高强度的搭配词都为名词和名词化短语。不仅仅是高强度的搭配词汇,其搭配的短语都需要物流专业人员重点学习。笔者将部分短语以表格形式呈现出来,如表5:

Table 5. Phrases of top 10 words with the highest collocation in the Logistics English
表 5. 物流专业英语搭配强度最高的前 10 个词的搭配短语

Candidate (grouping variant)	Terms
logistic	integrated logistics, advanced logistics, international logistics, global logistics, modern logistics, level of logistics, practice of logistics, strategic logistics, third-party logistics, function of logistics, area of logistics
cargo	general cargo, type of cargo, dry cargo, break-bulk cargo, cargo preference, bulk cargo, air cargo, cargo shipper, cargo vessel, cargo type, cargo airport, relief cargo, cargo containers, cargo rate, movement of cargo, cargo carriers, bulk cargo, cargo move, cargo movement, project cargo, cargo insurance
inventory	inventory level, inventory cost, inventory control, level of inventory, inventory management, reduction of inventory, total inventory, amount of inventory, availability of inventory, inventory investment, inventory deployment, inventory manager, inventory holding
vessel	ocean vessel, vessel owner, vessel operator, cargo vessel, specialized vessel, entire vessel, lash vessels, vessel type, vessel line, vessel load
container	container ship, loaded container, container line, container shipping, use of containers, container crane, container ports, cargo containers, number of containers, full container, container leasing, empty container, container load, intermodal container, container trade, container traffic

4. 结语

本文从词汇密度、平均词长、高频词和搭配强度四个层面分析物流专业英语词汇特征,得出以下结论:1) 物流专业英语的词汇变化程度远远高于普通书面英语。2) 物流专业英语中合成名词居多,因为大量的名词化短语可以有效简化语言结构。3) 物流专业英语中名词和名词化的短语搭配强度更高。4) 就物流作业中相同意义的词其使用的范围不同,具体需要结合语境。本研究对物流英语的学习和使用、物流教学大纲的设计和课本教材的编写具有一定的指导意义。基于这项研究,有两项建议:首先物流专业英语中出现的高频词、高强度的搭配词及短语是学习者学习的重点;其次是在研究物流专业英语语义时,一定要将词放入到语境中,以便区别其用法。

参考文献

- [1] 李晓旭. 物流英语的词汇特点及其翻译方法探讨[J]. 英语广场, 2017(7): 25-26.
- [2] Chung, T. and Nation, P. (2003) Technical Vocabulary in Specialized Texts. *Reading in a Foreign Language*, 15, 103-116.
- [3] Baayen, H. (2001) Word Frequency Distribution. Kluwer Academic Publishers, Dordrecht.
- [4] Nation, I. (2004) Teaching and Learning Vocabulary. Beijing Foreign Language Teaching and Research Press, Beijing.
- [5] Ure, J. (1971) Lexical Density and Register Differentiation. In: Perren, G.E. and Trim, I.L.M., Eds., *Applications of Linguistics*, Cambridge University Press, Cambridge, 443-452.
- [6] 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.
- [7] 谢志辉, 熊力游. 物流英语的词汇特征及其翻译[J]. 中国科技翻译, 2012, 25(1): 30-32+46.
- [8] 潘文国. 《汉英语对比纲要》[M]. 北京: 北京语言文化大学出版社, 2010.

-
- [9] 范勇. 电子商务英语的词汇特征及翻译[J]. 中国科技翻译, 2007(3): 36-38+32.
- [10] 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.
- [11] Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford University Press, New York.
- [12] 黄映秋. 国际物流英语的特点及翻译[J]. 中国科技翻译, 2008, 21(3): 30-32+61.
- [13] 孔德亮. 物流英语的特点及翻译[J]. 中国科技翻译, 2011, 24(1): 17-19+52.