

商务英语与英语语言学词汇密度特征研究

——一项基于语料库的研究

殷和萍

上海海事大学, 上海

收稿日期: 2022年8月7日; 录用日期: 2022年9月8日; 发布日期: 2022年9月15日

摘要

本文自建商务英语文本及英语语言学文本小型语料库, 利用语料库分析软件Antconc, 从词汇密度和词频两个角度, 在两类语料库中进行词汇对比。研究表明: 商务英语的词汇密度低于英语语言学词汇密度, 其语篇难度小于英语语言学; 由于商务英语和英语语言学文本中专业词汇较多, 商务英语和英语语言学语料库的高频词中共有的词汇主要是功能词, 实义词较少; 而只出现在商务英语语料库中的高频词均为实义词, 且与商务交易、商务活动等相关, 但只出现在英语语言学语料库中的高频词均与语义学、语用学等语言学层面有关。

关键词

商务英语, 英语语言学, 词汇密度, 高频词

Study on Characteristics of Lexical Density in Business English and English Linguistics

—A Corpus-Based Study

Heping Yin

Shanghai Maritime University, Shanghai

Received: Aug. 7th, 2022; accepted: Sep. 8th, 2022; published: Sep. 15th, 2022

Abstract

By means of analysis software Antconc, this paper conducted a comparative study in terms of lexical density and high-frequency words from self-built small corpus of business English text and linguistics text. The results show that the lexical density of business English is lower than that of

linguistics, and the difficulty of discourse in business English is lower than that of linguistics. Due to the large number of professional words in business English and linguistic texts, the common words of high frequency words in business English and linguistic corpora are mainly function words, while content words are few. The high-frequency words that only appear in the business English corpus are content words and related to business transactions and business activities, while the high-frequency words that only appear in linguistic corpus are related to semantics, pragmatics and other linguistic aspects.

Keywords

Business English, Linguistics, Lexical Density, High-Frequency Words

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

词汇研究有多种层次与视角,如从语言学内部对词的意义、内容进行研究,及从跨学科角度,如词汇认知、词汇的计算的研究等。近年来,随着计算机技术的发展,语料库在语言学研究 and 语言教学中发挥着越来越重要的作用,语料库语言学的发展为词汇本质及差异性特征的分析和学习带来了新的方法。Nelson (2001)利用自己创建的商务英语语料库从高频词、低频词等层面进行词汇研究,把高频词划分为5个类别,即商务人员(people in business)、商务活动(business activity)、商务行为(business action)、商务描述(business description)、商务事件与实体(business events and entities) [1]。

目前,商务英语词汇研究主要是定性研究,鲜有基于语料库的实证研究。并且对商务英语和语言学英语的词汇比较少有量化研究,比如,在词汇学习和使用中,英语语言学和商务英语二者关系如何、这两种文本词汇学习的难易度如何,英语语言学词汇是否可以覆盖商务英语词汇。本文在上述文献基础上,对商务英语和英语语言学词汇进行实证研究,探讨两者在词汇密度、平均词长及高频词等方面的差异及其原因,旨在为商务英语与英语语言学词汇学习和使用、商务英语及英语语言学大纲设计及教材编写等提供一定帮助。

2. 文献综述

词汇密度(lexical density)是衡量单位篇章信息,含量的尺度,因此决定语篇的难易程度。类符/形符比(types/tokens,简称TTR)通常用来测算语篇的词汇密度。Halliday (1985)认为,一个句子中的词语分为两类:一类是语法项,又称功能项,指在句中起限定作用的词,包括冠词、代词、大部分介词、连词、某些副词和限定性动词;另一类为词汇项,或称实词。词汇密度反映的是句中实词所占的比例[2]。由于在句中通常由实词来传递大部分信息,因此,如果一句话的词汇密度越大,则句中的实词越多,句子所载的信息量也越大。反之,信息量就越少。Ure (1971)是最早研究词汇密度的学者之一,其篇章词汇密度计算方法成为国内外学者运用最多的传统的计算方法。Ure (1971)提出词汇密度的计算公式为用实词数除以词汇总数得到的百分比,即:词汇密度 = 实词数/词汇总数 × 100% [3]。

Ure (同上)认为词汇密度是口语语体和书面语体最显著的区别之一。他发现口语的词汇密度要远低于书面语的词汇密度。一般说来,口语的词汇密度大概在40%以下,而书面语的词汇密度则在40%以上[3]。

在词汇密度的应用上,学者杨信彰(1995)在英语书面语体中的词汇密度特征这篇文献中从词汇密度的角度出发,探讨英语书面语体的正式程度。杨指出我们应该看到,英语中的词汇密度(lexical density)有助于我们认识英语书面语体的正式程度。学者通过具体的例子进行了话语分析,经过讨论,认为非正式书面语体的词汇密度低[4]。吴瑾和邹青(2009)基于中国学生英语口语语料库(SWECCL)和中国学习者英语语料库(CLEC),采用中介语对比分析法,利用语料库分析软件 WordSmith,从词汇密度和词频两个角度,研究了中国学生英语口语的语体特征[5]。吴国向(2020)采用两个实验探讨同音字密度对汉字识别的影响。实验一采用命名任务,结果发现高同音字密度的汉字的加工时间短于低同音字密度的汉字,说明在汉字识别中,同音字密度起促进作用;实验二运用眼动分析的方法,考察句子阅读中的同音字密度效应。词频是发现文体或语体特征的重要参考信息[6]。

3. 数据来源及方法

3.1. 研究方法

本研究从自建的商务英语文本及英语语言学文本小型语料库中,采用定量和定性分析方法相结合,比较商务英语和英语语言学的词汇分布特征。具体包括以下2个问题:

- 1) 基于两个小型的语料库,商务英语词汇以及英语语言学词汇密度各是多少?两者是否有显著差异?
- 2) 基于两个小型的语料库,商务英语词汇以及英语语言学词汇使用频率最高的50个词是哪些?能得出什么结论?

3.2. 数据计算

3.2.1. 词汇密度的计算

Ure 是最早研究词汇密度的学者之一,其篇章词汇密度计算方法成为国内外学者运用最多的传统的计算方法。按照 Ure (1971)采用公式 $TTR = \frac{\text{类符}}{\text{形符}} \times 100\%$ 的词汇密度公式,词汇密度计算前提是确定某一语料中实词和词汇总数的数量。词汇总数即所有单词的数量,可通过语料库检索软件 Antconc 计算[3]。

3.2.2. 词频的计算

Antconc 软件的单词列表功能可将语料中所有的单词按出现频率的高低依次列出,本研究选取的研究对象是词频位于前50位的单词。

4. 研究结果及讨论

4.1. 词汇密度及其差异情况

针对研究问题1,经过 Antconc 统计计算,基于两个自建小型语料库中商务英语词汇及英语语言学词汇的语料的功能词总数、词汇总数和词汇密度,采用 Ure 的词汇密度计算方法,通过类符形符比率研究商务英语和英语语言学词汇密度差异。结果如表1所示,见表1。

表1结果显示,在2个样本库形符总数相差不大的情况下,类符总数英语语言学比商务英语多3780。由此得出,英语语言学词汇密度(0.41)高于商务英语(0.37)。Antconc 检验数据表明,商务英语词汇量小于英语语言学,词汇密度也低。词汇密度常用作量度段落或文章难度。词汇密度越高,词汇量越大,语篇难度则越大,反之亦然。商务英语虽然属于专门用途英语,但从词汇密度层面来看,其语篇难度却低于英语语言学。由于商务英语中词汇重复率高,因此,商务交际谈判及交际中双方能够准确地表达,不会

产生歧义。商务英语用词难度较低,更容易传达书面及口语大意。而英语语言学词汇密度高,语篇分析难度较大,对学习者理解能力,纯语言功底及文学素养要求较高。

Table 1. Lexical density of English and linguistics textbooks

表 1. 商务英语和英语语言学词汇密度

| 样本库 | 商务英语 | 英语语言学 |
|------|------|-------|
| 形符总数 | 9960 | 18270 |
| 类符总数 | 3770 | 7550 |
| 词汇密度 | 0.37 | 0.41 |

4.2. 词频分析

针对研究问题 2,我们利用 Antconc 软件的单词列表功能,检出了两本教材语料词频最高的 50 个词。结果如表 2 所示,见表 2。

Table 2. 50 high-frequency words of English and linguistics textbooks

表 2. 商务英语和英语语言学教材中词频最高的 50 个词

| 序号 | 商务英语词汇 | 英语语言学词汇 | 序号 | 商务英语词汇 | 英语语言学词汇 |
|----|-------------|-------------|----|-------------|--------------------|
| 1 | the | the | 26 | book | have |
| 2 | and | of | 27 | exchange | its |
| 3 | to | and | 28 | production | series |
| 4 | of | in | 29 | trade | that |
| 5 | a | to | 30 | from | their |
| 6 | unit | a | 31 | making | work |
| 7 | for | English | 32 | material | written |
| 8 | in | by | 33 | module | register |
| 9 | is | for | 34 | practice | relations |
| 10 | business | this | 35 | market | studies |
| 11 | products | cohesion | 36 | socializing | text |
| 12 | you | from | 37 | service | were |
| 13 | communicate | as | 38 | your | semantic relations |
| 14 | production | at | 39 | are | most |
| 15 | as | linguistics | 40 | prices | or |
| 16 | their | on | 41 | develop | papers |
| 17 | can | semantics | 42 | business | speech roles |
| 18 | buyer | discourse | 43 | meeting | reference |
| 19 | meetings | has | 44 | cash | sentence |
| 20 | what | language | 45 | company | systems |
| 21 | with | other | 46 | structure | teaching |

Continued

| | | | | | |
|----|---------------|---------|----|--------------|---------|
| 22 | communicating | with | 47 | account | domain |
| 23 | interaction | which | 48 | cards | texture |
| 24 | on | are | 49 | negotiations | between |
| 25 | this | context | 50 | appointments | college |

通过 Antconc 对语料检索数据, 两本教材语料中出现频率最高的 50 个词都由功能词和实词构成, 但两者在实词的构成方面存在较大差异。语料库中词汇频数位列在前 50 的高频词(high frequency)占该语料库所有词目或类符的 50%左右, 并且能够体现语料的文体特点。本研究分别提取了 2 个样本库中所有词目并计算其频数, 按频数高低进行排序。商务英语与普通英语共有词汇频数最高的词为 the, to, and, a, as, at, by, this, from, are, but, you, all, English, she, more, year, some, do, structure, each, all, first, after, on, then, him, may, can, about, over, it, I, or so that, where, some, will。检索到的词主要为限定词(determiners), 如: the, a, that, this, his, one, their, its, her, some, only 等)、助动词(auxiliaries), 如: be, have, will, do, can, could, may, should 等)、介词(prepositions), 如: of, to, in, for, on, with, by, at, from, about, into, over 等)、连词(conjunctions, 如: and, as, but, or, if, when, so, that 等)、代词(pronouns), 如: it, I, he, you, which, they, we, all, she 等)。由于限定词、助动词、介词、连词和代词等通常属于功能词(function words), 因此, 商务英语与普通英语共有词汇频数最高的 50 个词中功能词约 35 个, 实义词(content words)约 15 个。这表明商务英语和英语语言学在功能词的使用方面区别不大, 功能词在商务英语和英语语言学中都属高频词。我们可以据此推测, 因为英语为形合语言, 注重形式化, 语句的构建、篇章的衔接与连贯, 语篇的分析都依赖功能词。

检索两个语料中词汇频数最高的各前十个词目中, 其中只出现在商务英语语料中的词目为 market, bank, business, meeting, company, customer, cost, goods, appointment, products, module, high interest, cultural exchange, contract, cash, service, project, negotiations。只出现在英语语言学语料中的词目为 language, cohesion, reference, semantics, linguistics, discourse, structure, sentence, context, semantics, register, text。以上词目表明, 只出现在商务英语语料中的 11 个词均为实义词, 大部分词除了具有作为普通英语的词义外, 在商务英语中还具有专门的意义。这表明商务英语除了使用普通英语的词汇外, 大量使用与商务交易相关的词汇。根据 Nelson 的理论, 上述商务英语词汇主要用于表达商务活动(business, trade, interest, projects, meetings)、商务交际(exchange, cultural communication, business interaction, negotiation)、商务事件与实体(products, services, production, goods, cost), 以及与商务人员相关(customer, buyer, seller)。只出现在商务英语语料中的词目也表明, 普通英语词汇通常引申为商务英语的高频词, 如 price, high, manage, market 等; 有些普通英语引申为商务英语词汇, 在商务环境中具有专门意义, 如 price (定价, 价格)、account (账户、账目)、exchange (交流)、interest (利息、股权、股份)等, 这些专业术语构成了商务英语的大部分词汇。相比之下, 只出现在英语语言学语料中的词目大多数为普通功能词与语言学专有词, 如 reference: 参考物, 参照物但在语言学当中意为“所指”, register: 登记, 在语言中意为“语域”。texture: 质地, 结构, 纹理, 语言中意为“语篇标识的特性”。

通过 Antconc 中的 concordance plot (索引定位), 检索商务词汇 business, contract, business negotiation 等词汇, 可以发现这种商务英语大量使用商务英语专业词汇, 这些专业词汇体现了商务英语的词汇特点。商务英语中的高频词, 如与商务活动、商务交际等相关的专业用语、产品名称等词往往较长, 但是在英语语言学中基本上没有这些商务英语词汇出现。并且随着信息技术的发展, 商务英语新词不断涌现, 而

且大多是复合词, 即由 2 个或 2 个以上的词结合起来构成的词, 例如, digital cash (电子现金)、cyber-marketing (网络化营销)、online shopping (网上购物)、pay on delivery (货到付款)等。而这些专业词汇不会出现在英语语言学词汇中。此外, 为了表述清晰, 体现语篇的正式性和庄重性, 商务英语合同或信函中通常使用专业的法律用语。例如, 友好解决: amicable settlement, 仲裁条款: arbitration clause, 转让条款: assignment clause 银行承兑: bank acceptance 等。这些专业术语在普通英语语言学中基本上不会使用。同样的, 在英语语言学教材中, 一些英语语言学专业词汇如 cohesion (衔接), linguistics (语言学), discourse (话语), text (篇章)在商务英语词汇中也不会出现。

5. 研究结果及讨论

本文从词汇密度、高频词以及索引关键词等层面从商务英语教材和英语语言学教材语料中比较了商务英语和英语语言学的词汇差异并分析其原因, 得出如下结论: 1) 在总词数相当的情况下, 商务英语词汇量小于英语语言学词汇量, 因而其词汇密度低于英语语言学。2) 在商务英语和英语语言学共有的高频词中, 大部分为功能词, 实义词较少; 2 个样本库各自 50 个高频词中, 两者在实词的构成方面存在较大差异, 而只出现在商务英语样本库中的词均为与商务交易等相关的词汇。造成以上词汇差异的原因是: 商务英语词汇重复率高, 其词汇多样性低于英语语言学; 与商务活动、商务描述等相关的专业用语、产品名称等词较长, 而且出现频次高; 由于语句的构建、篇章的衔接与连贯都依赖功能词, 商务英语和英语语言学在功能词使用方面区别不大, 两者高频词主要是共核词汇。商务英语词汇主要用于表达商务活动、商务描述和商务事件与实体等。英语语言学词汇主要用于研究语义学、语用学等语篇分析的方面。

本研究结果对商务英语与英语语言学词汇学习和使用、商务英语大纲设计及教材编写具有一定的意义。基于本文研究, 有两点建议: 第一, 商务英语和英语语言学中有很多的高频词, 这反映出二者的共核词汇在第二语言习得中应该始终作为学习的焦点, 扎实语言的基本功是学好英语语言学的基本词汇和熟练使用商务英语词汇的基础。第二, 商务英语和英语语言学 50 个高频词中的专有词汇反映出在整个商务英语和英语语言学学习中存在着大量的专有词汇, 因此商务英语和英语语言学各自的核心词汇有待于下一步量化研究。

参考文献

- [1] Nelson, M. (2000) A Corpus-Based Study of Business English and Business English Teaching Materials. University of Manchester, Manchester.
- [2] Halliday, M. (1985) Spoken and Written Language. Oxford University Press, Oxford.
- [3] Ure, J. (1971) Lexical Density and Register Differentiation Applications of Linguistics. Cambridge University Press, Cambridge.
- [4] 杨信彰. 英语书面语体中的词汇密度特征[J]. 解放军外语学院学报, 1995(3): 14-18.
- [5] 吴瑾, 邹青. 中国学生英语口语语体特征研究: 词汇密度与词频[J]. 山东外语教学, 2009, 30(1): 8-13.
- [6] 吴国向. 译文词汇密度及其波动研究——以多重《论语》英译为例[J]. 北京科技大学学报(社会科学版), 2020, 36(5): 25-32+42.