

# 中美企业年报CEO致辞的 词汇计量特征比较研究

何瑞云

浙江工商大学外国语学院, 浙江 杭州

收稿日期: 2023年8月14日; 录用日期: 2023年9月12日; 发布日期: 2023年9月26日

## 摘要

本研究以13家财富世界500强上榜的中美ICT企业年报CEO致辞作为研究对象, 自建语料库, 运用计量语言学方法, 从词汇变化度、词汇密度、词类分布三个方面, 对中美企业CEO致辞的词汇计量特征进行探讨。研究发现中美CEO致辞均属于正式的书面语体, 词类的概率分布均符合修正的右截尾齐普夫·阿列克谢耶夫模型, 具有规律性; 中美CEO致辞在三个指标上存在显著差异。该研究对CEO致辞的编撰、阅读和分析提供一定参考。

## 关键词

计量语言学, CEO致辞, 企业年报, 语料库, 词汇特征

# A Comparative Quantitative Study on Lexical Features of CEO Statements in Annual Reports of Chinese and American Enterprises

Ruiyun He

The School of Foreign Languages, Zhejiang Gongshang University, Hangzhou Zhejiang

Received: Aug. 14<sup>th</sup>, 2023; accepted: Sep. 12<sup>th</sup>, 2023; published: Sep. 26<sup>th</sup>, 2023

## Abstract

Using quantitative approach, this paper explores lexical features of CEO statements in annual re-

ports of Chinese and American enterprises in terms of lexical diversity, lexical density and part-of-speech distribution, taking the CEO statements of thirteen Chinese and American ICT companies listed among the Fortune Global 500 as the corpus data. The results show that both texts belong to a formal written genre, and part-of-speech distributions of both groups follow the modified right-truncated Zipf-Alekseev model, which suggests a certain regularity; there also exists significant differences in the three measures. This paper provides some references for the writing, reading and analysis of CEO statements.

## Keywords

Quantitative Linguistics, CEO Statements, Corporate Annual Reports, Corpus, Lexical Features

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

企业年度报告是商业领域的重要文件，股东、企业员工、政府机构、媒体等利益相关者通过年报了解企业上一年度的运营情况。CEO 致辞(又称总裁致辞、董事长致辞、致股东信)作为企业年报的重要组成部分，被视为最高管理层(如董事长和首席执行官)与股东、公众沟通的“窗口”。相较年报中财务报告这类“硬信息”，CEO 致辞更多呈现企业的“软信息”[1]。除了汇报财务业绩，最高管理层往往以信件的形式在 CEO 致辞中分享自己对企业治理、业务发展、社会责任和企业愿景的看法和计划。

CEO 致辞往往是企业经过深思熟虑的重要话语。第一，CEO 致辞位于年报开头部分，奠定年报的整体基调，是股东和公众广泛阅读的部分。这一部分语言大众化，是年报中最容易理解的部分[2]，也一直受到商务话语研究领域学者的关注。第二，CEO 致辞的信息披露是自愿性的，其披露的内容及形式均没有强制性规定，因而企业在披露时创作自由度高[1]。第三，CEO 致辞是企业高层与股东沟通、构建积极的企业身份并吸引投资的有效工具[2]。鉴于上述特点，CEO 致辞对企业信息披露、形象管理意义重大。

目前企业年报研究主要围绕体裁分析[3]、互文性[4] [5] [6]、立场标记语[7]、元话语[2] [8]、批评话语分析[9]等方面展开。上述研究丰富了企业年报的研究内容，但就研究对象而言，大多研究仅针对个别企业案例分析，研究结论的覆盖面有待扩大，或是选取来自各个行业的企业进行考察，尚未考虑行业变量对企业年报语言特征的影响；就研究方法而言，少有研究从计量语言学视角出发对年报 CEO 致辞的词汇特征进行精确的量化分析。计量语言学以定量方法对真实语言活动中的语言现象进行测量、观察、解释，正为语言特征的考察提供了准确可操作的方法。鉴于此，本研究自建的中美 ICT 企业 CEO 致辞文本语料库，运用计量语言学方法，从词汇层面比较中美企业年报 CEO 致辞的计量特征，以期为企业年报研究提供一个新的视角，对 CEO 致辞的编撰、阅读和分析提供一定参考。本文拟回答以下两个问题：1) 中美 CEO 致辞在词汇变化度、词汇密度、词类分布方面的词汇特征是什么？2) 上述中美 CEO 致辞的词汇特征有何不同？原因是什么？

## 2. 语料收集与研究方法

### 2.1. 语料收集

考虑到语料的权威性和代表性，以 2022 年财富世界 500 强企业排行榜为样本框，选取上榜的 ICT 企

业为研究样本,以避免行业差异对企业话语的影响。按照联合国 2008 年发布的第四版国际标准产业分类对 ICT 的定义,筛选得到 22 家中美两国 ICT 行业企业代表。随后在语料搜索过程中发现,1 家中国企业因企业性质的特殊性没有年报等信息披露,6 家美国企业官网只有不含有 CEO 致辞、提交至美国证券交易委员会的 10-K 文件,2 家企业缺少特定年份的 CEO 致辞。所以最终确定 13 家中美 ICT 企业作为本文的研究对象,统一于 2022 年 11 月从企业官网下载 2010 年到 2021 年的英文版年报。

截取上述 PDF 文件的 CEO 致辞文本部分(不包括表格和图片)转换成 WORD 文档,人工校对无误后批量转换为 TXT 文件。为了提高词性标注和统计分析的准确性,笔者进一步借助文本处理软件 PowerGREP5 及其 Text Cleaning Library 程序包对所有文本进行降噪清洗。文本的标点符号统一为英文输入,删除了无关符号(如项目编号和标注等)、网址、回车、换行等冗余信息和格式问题。由此建立两个语料库:中国 CEO 致辞文本(Chinese CEO Statement Text, CCST)、美国 CEO 致辞文本(American CEO Statement Text, ACST),语料库信息见表 1。本研究语料体裁一致,内容相似,总形符数相近,具有较高的可比性。

**Table 1.** Basic information of corpora

**表 1.** 语料库基本信息

语料库	形符数	类符数	时间	样本量
CCLT	171,934	9701	2010~2021 年	12
ACLT	165,558	9084	2010~2021 年	12

## 2.2. 研究方法

本研究采用定性与定量相结合的方法,利用 MATTR2.0、LancsBox6.0 软件提取中美 CEO 致辞的词汇特征数据,借助 IBM SPSS Statistics 23 进行统计学意义上的显著性差异分析,并尝试做出解释。为回答上文提出的两个问题,本文拟从词汇层面对 CCLT 和 ACLT 的词汇变化度、词汇密度、词类分布三个方面对中美 CEO 致辞的词汇计量特征进行考察。

词汇变化度指用词的丰富程度。该指标可以度量文本所使用不同单词的多少,反映了语言使用者的词汇量。型例比(Type Token Ratio, TTR)代表形符数与类符数之比,是计算词汇变化度使用最为广泛的方法,但这种方法很大程度受到文本长度的影响[10]。移动平均类符形符比(Moving Average Type Token Ratio, MATTR)是一种通过移动窗口计算 TTR 的算法,能有效减少文本长度的影响[11]。因此,本文使用 MATTR 衡量词汇变化度,沿用 Covington 和 Mcfall (2010)的方法将窗口大小定为 500 词。

词汇密度可以衡量口语或书面语的信息量大小,其计算方法为词汇词(或实词)与总词数的比值[12]。词汇词往往传达了文本的大部分信息,因此词汇词占比越高,文本传达的信息量就越大。词汇密度几乎不受语篇长度影响,一般用于区别语体[10]。具体操作中本文采取计算形符的方法来测算词汇密度[13]。

英语词类通常分为八类,名词、动词、形容词、副词为词汇词,代词、介词、连词、感叹词为语法词[14]。本研究基于 Huddleston (1984)和 Quirk 等 (1985) [15]的分类,并结合 CEO 致辞的用词特征,将词类分为十类,其中词汇词包括名词、动词、形容词、副词和数词,语法词为代词、介词、连词、限定词、情态动词。

## 3. 研究结果与讨论

### 3.1. 词汇变化度

移动平均类符形符比(MATTR, Moving Average Type Token Ratio)可以反映文本词汇是否多样化,是

否避免重复。Shapiro-Wilk 检验表明两个语料库每个样本的 MATTR 值均符合正态分布( $p > 0.05$ )。由此进行独立样本  $t$  检验, 结果表明两个库的 MATTR 值之间存在显著差异( $t = -4.317, p = 0.001$ )。具体而言, CCLT 的 MATTR 均值( $M = 0.505$ )显著低于 ACLT 的 MATTR 均值( $M = 0.523$ ), 这意味着 ACLT 具有更高的词汇变化度。

ACLT 的词汇变化度较高, 表明美国 CEO 致辞的词汇范围更广。比如, 在推介新型产品和服务时, 例(1)运用 introduce、launch、roll out 等含“推出”之义的动词或动词词组, 而例(2)反复使用 launch, 用词较为单调。究其原因, 主要受撰写者的英语词汇量和词汇应用能力限制的影响。作为英语学习者, 中国 CEO 致辞作者的词汇量和词汇应用能力相对有限。而词汇量丰富的语言使用者在写作中往往会借助近义词、反义词、上义词和下义词等词汇扩展方法进行大量的词汇变化。鉴于此, 中国企业可参照英语本族语者的企业年报扩大词汇量, 通过上述词汇扩展方法避免措辞单一和重复, 增加文本的可读性。

(1) X1 is available across our footprint, and we are already introducing the next generation of the X1 Platform... we launched our new Xfinity TV Go app in November 2013... We are also rolling out a new Xfinity TV app... (COMCAST)

(2) At the same time, the Company actively created an open and collaborative environment, launched more than 100 customized 3G handset models... The Company launched content application products... During the year, the Company launched 22 industry application products... At the end of the year, the Company launched “WO-Family” ... (CUCC)

### 3.2. 词汇密度

词汇密度大表明词汇词多, 文本所含信息量大。Shapiro-Wilk 检验表明两个语料库每个样本的词汇密度均符合正态分布( $p > 0.05$ )。由此进行独立样本  $t$  检验, 结果表明两个库的词汇密度之间存在显著差异( $t = 14.829, p = 0.000$ )。具体而言, CCLT 的词汇密度均值( $M = 0.650$ )显著高于 ACLT 的词汇密度均值( $M = 0.616$ ), 这意味着 CCLT 具有更高的词汇密度, 即中国企业 CEO 致辞的信息含量要高于美国企业 CEO 致辞。

另外, 词汇密度可以区分文本语体的正式程度。文本语体越正式, 词汇密度越高, 而非正式的、更接近自然口语的语体则具有较低的词汇密度[12]。本研究与中英媒体新闻社论的词汇密度较为接近(61.03%和 63.43%) [16], 且均采用了相同的词汇词界定, 使用了相同的方法计算词汇密度。因此, 就词汇密度而言, 中美企业年报 CEO 致辞均属于正式的书面语体。

### 3.3. 词类分布

词类是词的句法功能分类, 统计不同词类的使用频率可以反映文本词汇使用的特点。齐普夫·阿列克谢耶夫分布(Zipf-Alekseev distribution)能很好地捕捉词类的概率分布[17]。为了探讨中美 CEO 致辞词类的分布情况, 本研究借助计量语言学软件包 Altmann-Fitter3.1.0, 对 CCLT 和 ACLT 的词类分布进行统计, 使用修正的右截尾齐普夫·阿列克谢耶夫模型拟合数据。由表 3 和图 1 可知, CCLT 和 ACLT 的词类分布与修正的右截尾齐普夫·阿列克谢耶夫分布拟合结果很好(CCLT:  $R^2 = 0.9871 > 0.9$ ; ACLT:  $R^2 = 0.9774 > 0.9$ ), 均遵循幂律分布。中美 CEO 致辞的词类分布均存在规律, 说明该定律同样可以描写商务语体的词类分布, 具有一定普适性, 拓宽了其应用范围。观察表 2 可知, 在 CCLT 和 ACLT 的词类分布中, 两者的名词占比均为最高, 分别高达 34.61%和 30.15%, 这与前人关于自然语言中名词占比约为 33%的结果非常接近[18], 体现了名词的分布特点。另外, CCLT 和 ACLT 中限定词的频数并无显著差异( $p = 0.284$ ), 一定程度上说明中国 CEO 致辞撰写者运用限定词的频率倾向于本族语者的使用习惯。

**Table 2.** Frequencies of POS distributions in CCLT and ACLT and difference test  
**表 2.** CCLT 和 ACLT 中词类的频数及其差异性检验

CCLT			ACLT			卡方检验		
序号	词类	标准化频数	序号	词类	标准化频数	词类	$\chi^2$	Sig.
1	名词	3460.630	1	名词	3014.835	名词	764.852	0.000***
2	动词	1376.168	2	动词	1500.501	动词	-105.850	0.000***
3	介词	1052.555	3	介词	999.529	介词	25.687	0.000***
4	形容词	1034.176	4	形容词	925.114	形容词	113.299	0.000***
5	限定词	842.009	5	限定词	831.733	限定词	1.148	0.284
6	连词	654.321	6	代词	756.714	连词	-5.218	0.022*
7	代词	470.704	7	连词	673.963	代词	-1202.167	0.000***
8	副词	321.751	8	副词	397.142	副词	-138.385	0.000***
9	数词	309.712	9	数词	328.465	数词	-9.547	0.002**
10	情态动词	88.755	10	情态动词	82.267	情态动词	4.108	0.043*

中美 CEO 致辞在词类的具体分布上存在较大差异。卡方检验表明 CCLT 和 ACLT 之间名词、代词、介词、形容词、连词、代词、副词、数词和情态动词这九类词类的频数均存在统计学意义的显著差异。其中,在词汇词中,尽管 ACLT 的动词、副词、数词分布要多于 CCLT,但就占据分布前列名词(第 1 位)和形容词(第 4 位)而言,CCLT 的这两类词类频数显著高于 ACLT,从而解释了上文中国 CEO 致辞的词汇密度显著高于美国 CEO 致辞这一结论。

纵观主要词类分布,本研究限于篇幅仅具体讨论频数差异最为显著的代词、名词、形容词。首先,ACLT 使用的代词频数显著高于 CCLT ( $\chi^2 = -1202.167, p = 0.000$ )。LancsBox 自带的 TreeTagger 代词词性码包括四类,有人称代词、物主代词、wh 代词、所有格 wh 代词。考虑到 CEO 致辞的用词特征及其语篇的人际互动意义,笔者将聚焦人称代词和物主代词的使用上(其频数远远超出其他两个子类)。卡方检验结果表明,ACLT 中的人称代词和物主代词频数均显著高于 CCLT ( $p = 0.000$ )。细读文本发现,ACLT 往往使用第一人称代词指代自身,而 CCLT 倾向使用企业名称或 the company 这类自指语。经语料库检索统计,CCLT 出现的第一人称、第二人称代词及其物主代词均显著低于 ACLT ( $p = 0.000$ )。

(3) Let me take you through some of the key points of our financial outlook and capital allocation plan (AT&T).

(4) Gradually, the company will shift its focus from the annual business plan to the medium-to-long-term strategic plan (HUEWEI).

以上两个实例均为企业计划的陈述介绍,例(3)直接运用第一人称、第二人称代词及物主代词分别指代企业自身和目标读者股东,在文本中将作者和读者置于同一对话语境,拉近企业与投资者之间的距离,营造出双向、动态、平和的沟通氛围。而例(4)则使用第三人称 the company、its 表征自身,彰显了非个人化、较为正式的交际风格。中国企业倾向在年报等正式书面语体甚至包括书信形式的 CEO 致辞中将作者和读者抽离出来,尽可能避免第一人称、第二人称的使用,以保证语篇的客观公正性。这一现象可能源于中美文化差异,也为中国企业在对外沟通和对外宣传方面提供了参考价值。在编纂英文版 CEO 致辞的过程中,中国企业应树立跨文化意识,充分考虑目标国的写作习惯,适当调整人称视角等语言手段,构筑平等交流的互动氛围,吸引利益相关者加大投资力度。

其次,CCLT 使用的名词频数显著高于 ACLT ( $\chi^2 = 764.852, p = 0.000$ )。细读文本发现,这可能与中美 CEO 致辞相较美国 CEO 致辞更倾向使用名物化相关。名物化现象在商务话语中十分常见,可增加信息量、使话语客观化、衔接语篇[19]。以-ment 名物化为例,笔者借助 LancsBox 通配符检索此类名物化在 CCLT 和 ACLT 出现的频数,并手动排除后缀为-ment 但不属于名物化的单词,如 government 和 segment。

卡方检验结果表明两者存在显著差异( $\chi^2 = 668.643$ ,  $p = 0.000$ ), CCLT 中-ment 名物化的标准化频数是 ACLT 的三倍(分别为 125.979 和 43.852)。这在一定程度上导致中国 CEO 致辞的名词含量大大高于美国 CEO 致辞。

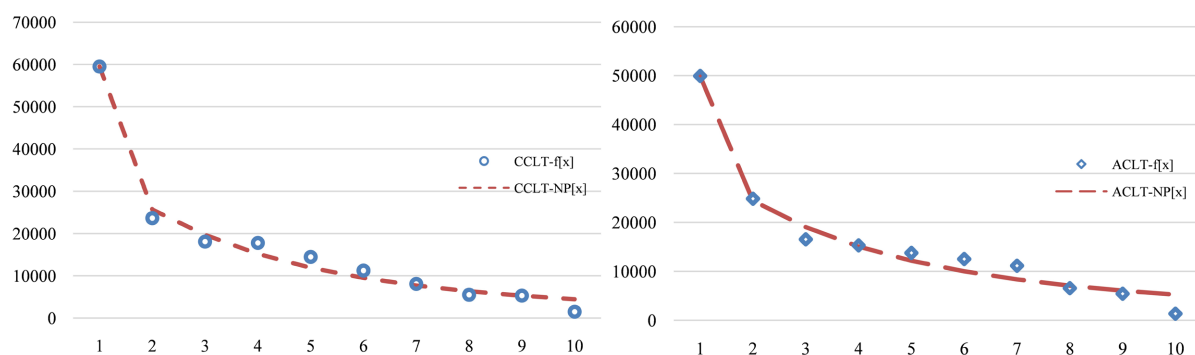
最后, CCLT 使用的形容词频数显著高于 ACLT ( $\chi^2 = 113.299$ ,  $p = 0.000$ )。细读文本发现, CCLT 中常见多个形容词修饰同一名词或名词词组。如例(5)所示, 中国致辞撰写者并未充分揣摩产品服务名字的具体内涵, 习惯直译品名, 逐字堆砌形容词定语, 增加了读者的阅读负担。在这方面, 中企应仔细推敲品名的翻译, 凸显英语用词的简洁和地道, 以提升中国企业海外传播的有效性。

(5) ... we aimed to provide intelligent-connected and vertically-integrated intelligent applications services for users (CTCC).

**Table 3.** Fitting results of the modified right-truncated Zipf-Alekseev distribution to POS distributions of CCLT and ACLT  
**表 3.** CCLT 和 ACLT 中词类分布与修正的右截尾齐普夫·阿列克谢耶夫分布拟合结果①

x	CCLT		ACLT	
	f[x]	NP[x]	f[x]	NP[x]
1	59500	59500.00	49913	49913.00
2	23661	25647.95	24842	24443.74
3	18097	19674.66	16548	19042.14
4	17781	15173.19	15316	15076.38
5	14477	11905.90	13770	12179.86
6	11250	9509.80	12528	10019.97
7	8093	7719.01	11158	8371.90
8	5532	6354.40	6575	7087.72
9	5325	5295.78	5438	6068.55
10	1526	4461.30	1362	5246.74
拟合结果	$a = 0.0098, b = 0.3595,$ $n = 10, \alpha = 0.3601, R^2 = 0.9871$		$a = 0.1096, b = 0.2826,$ $n = 10, \alpha = 0.3170, R^2 = 0.9774$	

① f[x]表示某词类的实际频数, NP[x]表示对应词类的公式拟合理论值。



②横坐标表示词类, 纵坐标表示频数; f[x]表示某词类的实际频数, NP[x]表示对应词类的公式拟合理论值。

**Figure 1.** Fitting results of the modified right-truncated Zipf-Alekseev distribution to POS distributions of CCLT and ACLT  
**图 1.** CCLT 和 ACLT 中词类分布与修正的右截尾齐普夫·阿列克谢耶夫分布拟合结果②

#### 4. 结语

本研究选取 13 家财富世界 500 强上榜的中美 ICT 企业年报 CEO 致辞自建语料库, 运用计量语言学方法, 从词汇变化度、词汇密度、词类分布三个方面探讨中美企业 CEO 致辞的词汇计量特征。研究发现,

中美 CEO 致辞的词汇特征存在共性：中美 CEO 致辞均属于正式的书面语体，词类的概率分布均遵循修正的右截尾齐普夫·阿列克谢耶夫模型，具有一定规律性。同时中美 CEO 致辞的词汇特征在三个指标上存在显著差异：中国 CEO 致辞的词汇变化度显著低于美国 CEO 致辞，词汇量较小，但其词汇密度较高，信息含量较大，使用更多名词和形容词。而美国 CEO 致辞使用更多第一人称、第二人称代词及其物主代词，注重企业与投资者平等交流的互动氛围。究其缘由，这可能与词汇量和词汇应用能力、名物化的使用倾向、社会文化等方面的差异有关。后续研究将扩大语料规模，进一步验证研究结果的有效性，并选取更多指标系统考察中美企业 CEO 致辞的词汇计量特征，以期为企业年报的编撰、阅读和分析提供一定参考，提升中国企业国际话语建构能力。

## 参考文献

- [1] Abrahamson, E. and Amir, E. (1996) The Information Content of the President's Letter to Shareholders. *Journal of Business Finance and Accounting*, **23**, 1157-1182. <https://doi.org/10.1111/j.1468-5957.1996.tb01163.x>
- [2] Hyland, K. (1998) Exploring Corporate Rhetoric: Metadiscourse in the CEO's Letter. *International Journal of Business Communication*, **35**, 224-244. <https://doi.org/10.1177/002194369803500203>
- [3] Nickerson, C. and De Groot, E. (2005) Dear Shareholder, Dear Stockholder, Dear Stakeholder: The Business Letter Genre in the Annual General Report. In: Gillaerts, P. and Gotti, M., Eds., *Genre Variation in Business Letters*, Peter Lang, Bern, 325-346.
- [4] Bhatia, V.K. (2010) Interdiscursivity in Professional Communication. *Discourse & Communication*, **4**, 32-50. <https://doi.org/10.1177/1750481309351208>
- [5] 王立非, 李焯坤. 中美商务语篇互文性多维对比研究[J]. 外语教学理论与实践, 2018(3): 56-62.
- [6] 乌楠, 张敬源. 主题互文视域下企业年报语篇的协变量结构与企业身份建构[J]. 解放军外国语学院学报, 2020, 43(4): 78-86.
- [7] Fuoli, M. (2018) Building a Trustworthy Corporate Identity: A Corpus-Based Analysis of Stance in Annual and Corporate Social Responsibility Reports. *Applied Linguistics*, **39**, 846-885. <https://doi.org/10.1093/applin/amw058>
- [8] 胡春雨, 李旭妍. 基于语料库的腾讯亚马逊致股东信元话语研究[J]. 外语学刊, 2018(1): 24-32.
- [9] 乌楠, 张敬源. 中美企业机构身份的话语建构策略[J]. 现代外语, 2019, 42(2): 220-230.
- [10] 张艳, 陈纪梁. 言语产出中词汇丰富性的定量测量方法[J]. 外语测试与教学, 2012(3): 34-40.
- [11] Covington, M.A. and McFall J.D. (2010) Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR). *Journal of Quantitative Linguistics*, **17**, 94-100. <https://doi.org/10.1080/09296171003643098>
- [12] Ure, J. (1971) Lexical Density and Register Differentiation. *Applications of Linguistics*, **27**, 443-452.
- [13] 鲍贵. 二语学习者作文词汇丰富性发展多纬度研究[J]. 外语电化教学, 2008(5): 38-44.
- [14] Huddleston, R. (1984) *Introduction to the Grammar of English*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9781139165785>
- [15] Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. Longman, London.
- [16] 王琳, 朱少勤. 中英媒体英语社论的词汇计量特征比较研究——以《中国日报》和《经济学人》为例[M]//黄伟, 等. 词汇与句法计量研究. 杭州: 浙江大学出版社, 2022: 56-68.
- [17] Hammerl, R. (1990) Untersuchungen zur Verteilung der Wortarten im Text. *Glottometrika*, **11**, 142-156.
- [18] Liang, J. and Liu, H. (2013) Noun Distribution in Natural Languages. *Poznań Studies in Contemporary Linguistics*, **49**, 509-529. <https://doi.org/10.1515/psicl-2013-0019>
- [19] 王立非. 商务话语名物化研究[M]. 北京: 对外经济贸易大学出版社, 2016.