

基于语料库的中西方致股东信文本特征对比研究

陈倩倩

天津财经大学, 天津

Email: 2832849417@qq.com

收稿日期: 2020年9月8日; 录用日期: 2020年9月23日; 发布日期: 2020年9月30日

摘 要

致股东信是为了吸引更多投资者, 实现企业利益的重要手段。研究选取中西方致股东信为研究语料, 自建语料库, 对比分析了中西方致股东信在词汇、句法、语篇层面的文本特征及差异。研究发现, 1) 中西方致股东信的平均词长、词汇密度、情态系统存在差异, 中国致股东信平均词长略低于西方, 但词汇密度更加丰富。与西方致股东信相比, 中国致股东信更多使用中值情态, 过少使用高值和低值情态词; 2) 中西方致股东信的平均句长和定语从句存在差异, 中国致股东信的平均句长大于西方, 但定语从句的使用率低于西方; 3) 中国致股东信的语篇可读性低于西方, 更难阅读。本研究对中国企业致股东信的书写有启示作用, 能提高中国企业吸引海外投资的竞争力和吸引力。

关键词

语料库, 致股东信, 文本特征板

A Corpus-Based Comparative Study of Textual Characteristics between Chinese and Western CEO's Letters

Qianqian Chen

Tianjin University of Finance and Economics, Tianjin

Email: 2832849417@qq.com

Received: Sep. 8th, 2020; accepted: Sep. 23rd, 2020; published: Sep. 30th, 2020

Abstract

The CEO's letter is an important means to attract more investors and maximize the benefits. This study selected Chinese and Western CEO's letters as data to analyze the lexical, syntactic and discursive characteristics and compare the differences between the two self-built corpora. The study found that 1) there were differences between Chinese and Western CEO's letters in AWL, lexical density and modality. The AWL of Corpus-CN is shorter than that of Corpus-SK, but the lexical density is larger. Compared with the Western, the Chinese CEO's letters displayed an overuse of medium modality and an underuse of high and low modality. 2) There were differences in ASL and attributive clauses. The ASL of Corpus-CN was longer than that of Corpus-SK, but the number of attributive clauses in Corpus-CN was less than that of Corpus-SK. 3) The text readability of Corpus-CN was lower than that of Corpus-SK. The study has implications for writing Chinese CEO's letters, and improving the competition and attraction of Chinese corporate.

Keywords

Corpus, CEO's Letters, Textual Characteristics

Copyright © 2020 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

商务话语已经成为语言学、管理学、社会学等多学科、多领域共同关注的热点研究问题。由于企业报告文本包含了众多信息，具有很高的研究价值，近年来，我国企业外宣资料(主要是企业年报、社会责任报告和企业简介等)引起了众多研究者的兴趣。致股东信通常位于企业报告的开头，是报告中最明显且最为广泛阅读的部分。由于中西方语言风格的不同，企业致股东信的语言表达存在一些差异，这对双方的利益沟通，信息传递，形象构建带来负面影响。两者文本的不同甚至会导致交流障碍和文化冲突，最终失去投资者。这一系列问题指引研究者将目光转向中西方致股东信文本特征的研究。一份成功的致股东信不仅能真实地陈述企业经营状况，高效地传递企业信息、构建积极的企业形象，更能缩小不同文化读者之间的文化差异，稳定住投资者的信心，扩大投资来源规模，建立良好的企业与股东关系。本研究拟自建小型语料库，以中西方企业社会责任报告的致股东信为语料，对比分析两类报告在词汇、句法、语篇三个层面的文本特征，并探究其背后的语言文化机制。

2. 相关文献回顾

西方的股票市场发展较早，所以致股东信在国外的历史较长，其重要性已得到普遍认可。众多国外学者都对致股东信的商务话语进行了不同方面的研究。有学者探讨了公司运营情况好坏与致股东信可读性(readability)之间的关系，研究者[1] [2] [3]发现在经营状况良好的年份中致股东信的可读性更高；Still和Smith [4]测量了英国企业致股东信的弗莱士易读度，指出80.7%母语成年人发现，将近77%的致股东信报告超出了他们的阅读能力；有的学者研究致股东信的沟通策略，如Hooghiemstra [5]和Poole [6]发现企业通过控制致股东信中的语言，尽可能地强调好消息，将其归功于管理者的能力，而在传达坏消息时模糊作者的临在(authorial presence)，将其归咎于不可控的外力因素以减轻管理者承担的责任；对致股东

信的语步分析(move analysis), 如 Mobasher 和 Mohamad [7] 都在体裁分析的基础上归纳总结出致股东信的语步模型; 还有对致股东信的元话语探究, 例如 Hyland [8]。

随着经济全球化的发展, 中国企业“走出去”步伐加快, 中国企业对致股东信的关注度也越来越高, 期望借助英文致股东信构建积极的企业形象并提高投资竞争力和吸引力。国内关于致股东信的研究近几年也发展起来。高天[9]和郭芸[10]对致股东信的体裁分析, 黄莹[11], 胡春雨和李旭妍[12]对致股东信的元话语研究, 胡春雨和谭金琳[13]对致股东信隐喻及翻译研究, 以及钱嘉慧, 王立非[14]对致股东信中出现的态度评价的分析。

综上所述, 现有研究存在以下不足: 缺乏对致股东信文本特征的语料库研究, 缺乏针对中西方企业致股东信的对比研究。基于此, 本研究将利用语料库分析工具对中西方企业致股东信的文本特征进行对比分析。

3. 研究设计

3.1. 研究问题

本研究主要回答以下三个问题: 1. 从词汇层面看, 中西方企业致股东信的词汇复杂度和情态动词呈现出何种特征, 是否存在差异? 2. 从句法层面看, 中西方企业致股东信的平均句长以及句法结构呈现出何种特征, 是否存在差异? 3. 从语篇层面看, 中西方企业致股东信在语篇可读性和主题词方面呈现出何种特征, 是否存在差异?

3.2. 语料库和研究工具

为了深入了解中西方能源企业英文致股东信的文本特征, 本研究依据标普全球普氏(S & P Global Platts)公布的 2019 年度全球 250 强能源公司排行榜, 选取了中国石化、中海油、中国神华能源和中国石油四家企业。为保证语料的可比性, 研究选取的西方能源企业有 Royal Dutch Shell plc、Exxon Mobil Corp、Cheron Corp、Glencore Plc、Next Era Energy Inc 五家知名企业。语料来自各个企业官网上的英文报告。本文均选取其 2016~2019 年的企业社会责任报告。获取 PDF 版报告后, 截取其中的致股东信部分转换成 word 格式, 手工校对无误后再转化为可供语料库工具检索的 txt 格式。

本研究采用 Claws 4 对语料进行词性赋码, 采用 Ant Conc 3.5.8 作为检索、统计软件, 对词频、主题词、词性分布等词汇特征进行研究。文本可读性分析采用许家金和贾云龙研发的 Readability Analyzer 1.0, 该软件能统计文本的词次、词形、词目、平均词长、平均句长, 并提供文本易度、难度、水平等数据。本研究还使用到的可读性软件是 QT Readability, 该软件可提供弗莱士易读度(Flesch Reading Ease), 金凯德难度级数(Flesch-Kincaid Grade level), 迷雾指数(Gunning Fog Index)等数据。Flesch 可读性公式为: $READEASE = 206.835 - (1.015 \times \text{total words}/\text{total sentences}) - (84.6 \times \text{total syllables}/\text{total words})$ 。

3.3. 测量指标选取和数据分析

文本特征指标会影响语篇的沟通有效性, 复杂度则会影响语篇可理解性和交际有效性。Alderson [15] 曾指出词汇复杂度、句法复杂度和可读性等变量会影响语篇复杂度和阅读难度, 其中词汇复杂度和句法复杂度分别指影响文本可理解性的词汇和句法特征, 如词长、词汇密度、句长、句法结构等。可读性代表文本更容易阅读的文本特征, 通常由弗莱士易读度, 金凯德年级水平和迷雾指数衡量。Flesch [16] 提出的弗莱士易读度可衡量文章的阅读难度, 分值越高表示文本越容易阅读。Kincaid [17] 提出了金凯德年级水平, 以词长和句长作为变量测算读者需要接受多少年的正规教育才能够顺畅读懂某篇文章, 因此分值越高, 代表语篇复杂度越高, 越难读懂。而迷雾指数从词频、词汇难度、平均句长和逻辑等方面考察文

本难易度,也以学习的时长来量化文本可读性。综上,本研究从词汇、句法、语篇三个层面选取的文本特征测量指标包括词长、词汇密度、情态动词、句长、句法结构、弗莱士易读度、金凯德年级水平和迷雾指数。

3.4. 研究程序

首先将研究语料分别保存于两个文件夹中,构成两个语料库:Corpus CN (中国企业致股东信语料库)和Corpus SK (西方企业致股东信语料库)。然后对语料库进行格式清理,并采用上述软件进行处理和统计分析。Corpus CN 的字符数为 14,748,Corpus SK 含 15,114 个字符。

4. 结果与讨论

4.1. 词汇特征分析

4.1.1. 词汇复杂度

本研究对于词汇复杂度的分析从平均词长(AWL)和词汇密度(Lexical Density)两个方面进行。词汇长度可以反映出文本中所使用的词汇的难易程度,长度越长,则词汇越难,同时会使得文本较为晦涩,增加文本的阅读难度。关于词汇密度的测量方法,学界有两种说法。其中一个说法是 Scott [18]提出的标准类符形符比(STTR)。另一种测量方法是 Stubbs [19]提出的,用语料库中所有实义词比上总词数来计算词汇密度。一般认为,英语的实义词指的是名词、动词、形容词和副词四个词类。本研究基于以上两种测量方法,对自建的语料库进行词汇复杂度的分析。首先,两个语料库经过可读性分析软件 Readability Analyzer 1.0,得出两个语料库的平均词长、类符、形符和标准类符形符比。然后,将两个语料库通过线上赋码工具 Claws 4 进行词性赋码,最后利用 Ant Conc 检索功能,对各个语料库中的名词(NN)、动词(VV)、形容词(JJ)及副词(RR)进行检索统计,计算出实词与总词的百分比。具体见表 1。

Table 1. The result of lexical complexity

表 1. 词汇复杂度结果

	平均词长	形符	类符	标准类符形符比	实词/总词
Corpus-CN	5.50	14,748	2,391	17.06%	46.69%
Corpus-SK	5.60	15,114	2,284	16.08%	40.92%

根据表 1 结果,中西方企业致股东信的词汇复杂度存在差异。具体来看,中国致股东信的平均词长是 5.50,相较于西方致股东信的平均词长,低了 0.10。这就说明中国致股东信使用的词汇更容易,词汇更易理解。对比中西方语料库的标准类符形符比发现,中国致股东信用词变化较大,文本词汇的多样性和丰富性高于西方致股东信。实词与总词的百分比同样看出,中国致股东信的信息密度大于西方。由此可以推断中国企业社会责任报告的致股东信披露的信息多于西方企业。

4.1.2. 情态动词

从情态系统看,说话者可以对一个命题的真伪表达不同程度的判断,也可以对听话者进行不同程度的提议或指令。根据 Halliday [20]对情态强度高、中、低三个基本量值的分类和描述,本研究关注的情态动词依次为 must/will, should, would/can, could, may, might。利用 Ant Conc 的检索搭配功能,对两个语料库中需要考察的情态动词进行检索,统计其频次。结果见表 2。

根据表 2 情态动词的词频统计,中西方企业致股东信的情态系统表达存在差异。中国致股东信使用最多的是中值情态词,而其中的高值情态词使用为零,显著低于西方致股东信中高值情态词的使用频率,

低值情态动词中表达能力的词(*can*)使用频率也显著低于西方。由此推断,西方企业主席在发表致辞时态度更强硬,对于承担的社会责任更加明确,对企业能力的认识也更清晰。中国企业较多使用中值情态词,想和股东们建立一个和谐友好的关系,以保证企业利益,并希望获取更多投资者的关注,争取更大的利益。这也看出中西方企业追求不同的企业形象。西方企业更倾向于展现自己富有责任,充满自信的形象,中国企业则倾向于表现合作共赢的态度。

Table 2. The statistics of modal verbs

表 2. 情态动词统计

	高值		中值			低值			
	must	will	should	would	can	could	may	might	
Corpus-CN	0	55	1	5	8	1	1	0	
Corpus-SK	20	63	5	8	22	4	1	2	

4.2. 句法特征分析

研究将从中西方致股东信的平均句长和句法结构(主要关注定语从句)进行句法层面的分析,讨论中西方致股东信的句法特征及差异。平均句长指的是语料库中每一个独立句子中平均包含的单词数量。一般而言,长句包含较多的分句及复合句,信息量更多,逻辑关系更复杂,阅读难度更大。依据英语国家商务题材写作标准,平均句长在 15 到 20 的语篇传达信息的效率最高。本研究利用 Readability Analyzer 分析软件分别统计出两个语料库的平均句长(ASL)。结果显示,Corpus-CN 的平均句长为 26.9,即平均每个句子包含 26.9 个单词。相对比,Corpus-SK 平均句长为 20.0,低于 Corpus-CN,其文本可读性更易理解,语篇传达信息的效率更高。

定语从句的统计主要是检索关系代词和关系副词的出现频率。常见引导词有 *that*, *who*, *which*, *whom*, *whose*, *when*, *where*, *why*, 特别指出的是,通过人工检验,不是引导定语从句或者不是作为关系词功能出现的次数已经被剔除。将两个语料库分别通过 Ant Conc 进行检索统计,得出表 3 结果。

Table 3. The statistics of attributive clause

表 3. 定语从句统计

	that	who	which	whom	whose	when	where	why	总
Corpus-CN	7	2	19	0	0	0	7	0	35
Corpus-SK	85	15	52	0	0	1	15	0	168

结果发现,中国致股东信中定语从句使用率明显低于西方,说明西方致股东信的句法更复杂,逻辑关系更复杂,阅读难度更大。中西方致股东信中都很少或不使用 *whom*, *whose*, *when*, *why* 引导的定语从句,双方更多使用 *that*, *who*, *which* 来引导句子。研究下一步从两个语料库中提取“*that*”,“*who*”,“*which*”引导定从的先行词搭配(见表 4),以期进一步发现两者在复杂句式使用上的具体差异。首先,在子库 Corpus-CN 的形符少于 Corpus-SK 的情况下,定语从句的使用频率也低于后者,先行词的种类也远低于后者。例如,在“*who*”引导的定从中,中国企业致股东信仅限于“*those*”,而西方企业有更多的词由“*who*”引导,如表 4 显示的有 *people*, *employees*, *parties* 等。这使得西方企业致股东信句法结构比中国致股东信更加多样化。

Table 4. The collocation of That, Who, Which in two corpuses
表 4. 两者定语从句中 that、who、which 的先行词搭配

	that	who	which
Corpus-CN	Perspectives, station, volunteering, system, problems, report	those	Agreement, development, mechanism, model, achievements, gas, topics, belief, practices, fleet, alleviation, et al.
Corpus-SK	View, assets, elements, emissions, commodities, risks, priorities, cycle, communities, chains, technologies, et al.	People, employees, those, parties, miners, report review panel	Workplace, programme, review, resolution, team, report, policy, business, values, way, health, world, committee, company, et al.

4.3. 语篇特征分析

本研究对致股东信语篇特征的分析聚焦于语篇可读性和主题词词表两方面。本研究的语篇可读性由弗莱士易读度, 金凯德年级水平和迷雾指数衡量。这三个指标的测量借助了 QT Readability 工具(见表 5)。在语料库语言学中, 主题词指与参照语料库相比, 某一观察文本或语料库中出现频率显著偏高或偏低的词。主题词可反映文本的词汇使用规律和倾向, 流露出语篇的核心主题和意义。关键度(keyness)是其最重要的衡量指标。为便于准确的对比分析, 本研究仅研究出现频率偏高的主题词, 利用 Ant Conc 语料库工具, 先后以 Corpus-CN 和 Corpus-SK 作为对方的参照语料库, 得出主题词词表(见表 6)。

根据表 5 的结果, 中国企业致股东信的弗莱士易读度是 11.61, 低于西方致股东信的 27.84, 看出中国致股东信更难理解和阅读。同样, 从金凯德年级水平和迷雾指数来看, 中国致股东信需要的年级水平高于西方, 迷雾指数也高于西方。整体推断出中国企业致股东信的语篇可读性更不易理解和阅读。

Table 5. The result of readability
表 5. 语篇可读性结果

	弗莱士易读度	金凯德年级水平	迷雾指数
Corpus-CN	11.61	18.32	22.62
Corpus-SK	27.84	14.55	18.88

Table 6. Keyword List (only the top 20)
表 6. 主题词词表(仅列出前 20 位)

次序	Corpus-CN	Keyness	Corpus-SK	Keyness
1	development	+160.65	our	+146.97
2	xb	+155.35	to	+71.73
3	China	+103.29	that	+70.44
4	and	+91.07	sustainability	+69.3
5	green	+72.27	are	+60.85
6	oil	+54.61	this	+59.78
7	cooperation	+50.02	glencore	+55.43
8	gas	+47.59	shell	+49.88
9	cnpc	+45.85	chevron	+44.33
10	shenhua	+44.45	report	+42.5
11	alleviation	+41.67	we	+41.68

Continued

12	reform	+41.67	emissions	+40.64
13	poverty	+39.13	is	+36.76
14	social	+38.72	do	+34.63
15	rmb	+34.72	hesc	+33.24
16	sinopec	+34.72	operations	+32.08
17	high	+34.55	how	+29.09
18	th	+34.22	change	+28.89
19	win	+31.95	climate	+28.03
20	company	+31.32	working	+28.02

根据表 6 的结果, 中西方致股东信的主题词存在较大差异。西方致股东信中第一人称代词 *our, we* 分别位居主题词词表的第 1 和 11 位, 与中国致股东信形成显著差异。可见西方致股东信的语气更加亲切自然, 从而拉近企业与股东的关系, 再次体现了西方企业更倾向于使用亲切又不失正式的语体行文, 字里行间体现着一种“顾客至上”的企业文化精神。*Cooperation* 在中国致股东信中的高频出现, 体现了中国企业以合作为导向的企业文化。作为社会责任报告中的致股东信, 两者都表现出各自的社会责任关注点。中国作为一个发展中国家, 发展放在首位, 企业也为消除贫困履行社会责任, 紧跟社会改革需要, 为社会发展贡献力量。西方企业更关心可持续发展, 排放量, 气候变化等方面的社会责任。以上的不同也受到中西方文化的影响。

5. 总结

本研究以中西方企业致股东信为研究语料, 借助语料库分析工具, 探究两者在词汇、句法、语篇层面的特征及差异。具体发现: (1) 中西方致股东信在词汇层面表现出不同的文本特征, 中国企业致股东信的平均词长略小于西方, 使用的词汇较简单, 但中国企业的词汇密度大于西方, 展现了更为丰富的词汇, 也加大了阅读难度; 对于两者语料库中情态动词的考察发现中国致股东信多使用中值情态, 很少或不使用高值情态和低值情态动词。(2) 中西方致股东信在句法层面表现出不同的文本特征, 中国致股东信的平均句长大于西方, 更难理解, 但根据定语从句的统计分析, 中国致股东信的定从数量低于西方, 句法逻辑性简单。(3) 中西方致股东信在语篇层面表现出不同的特征, 中国致股东信的语篇可读性低于西方, 文本更难阅读; 两个语料库的主题词词表呈现了中西方在社会责任方面不同的关注点。本研究建议, 企业致股东信的词汇和句法复杂度要调整到合理水平, 降低词汇和句子复杂度, 增强报告的语篇可读性。企业要提高致股东信的沟通有效性, 以此向股东传达高信度和高效度的企业信息, 稳定投资者的信心, 扩大投资来源规模, 提振股票市场。中国企业应考虑海外潜在及现有投资者的文化背景, 使英文致股东信符合海外投资者的语言习惯, 以便建立积极的企业形象并增添其对企业的感和信心, 从而实现吸引并留住投资者的目的。

参考文献

- [1] Kohut, G.F. and Segars, A.H. (1992) The President's Letter to Stockholder: An Examination of Corporate Communication Strategy. *Journal of Business Communication*, 29, 7-21. <https://doi.org/10.1177/002194369202900101>
- [2] Subramanian, R., Insley, R.G. and Blackwell, R.D. (1993) Performance and Readability: A Comparison of Annual Reports of Profitable and Unprofitable Corporations. *Journal of Business Communication*, 30, 49-61. <https://doi.org/10.1177/002194369303000103>

- [3] Jones, M.J. (1988) A Longitudinal Study of the Readability of the Chairman's Narratives in the Corporate Reports of a UK Company. *Accounting and Business Research*, **18**, 297-305. <https://doi.org/10.1080/00014788.1988.9729377>
- [4] Still, J.E. and Smith, N.P. (1971) Readability: A Measure of the Performance of the Communication or Corporate Function of Financial Reporting. *The Accounting Review*, **46**, 552-561.
- [5] Hooghiemstra, R. (2010) Letters to the Shareholders: A Content Analysis Comparison of Letters Written by CEOs in the United States and Japan. *The International Journal of Accounting*, **45**, 275-300. <https://doi.org/10.1016/j.intacc.2010.06.006>
- [6] Poole, R. (2016) Good Times, Bad Times: A Keyword Analysis of Letters to Shareholders of Two Fortune 500 Banking Institution. *International Journal of Business Communication*, **53**, 55-73. <https://doi.org/10.1177/2329488414525449>
- [7] Mobasher, A. and Mohamad, A. (2015) Exploring the Macrostructure of Management Forewords of Corporate Annual Reports. *International Journal of Applied Linguistics & English Literature*, **4**, 14-25.
- [8] Hyland, K. (1988) Exploring Corporate Rhetoric: Metadiscourse in the CEO's Letter. *The Journal of Business Communication*, **35**, 224-245. <https://doi.org/10.1177/002194369803500203>
- [9] 高天. 体裁分析理论在商务英语信函研究中的应用——以《致股东信》的体裁分析为例[J]. 湖北第二师范学院学报, 2013, 30(6): 17-21.
- [10] 郭芸, 任再新. 基于语料库的首席执行官报告的体裁分析[J]. 厦门理工学院学报, 2015(4): 83-88.
- [11] 黄莹. 元话语标记语的分布特征及聚类模式对比分析——以银行英文年报总裁信为例[J]. 外国语文, 2012(4): 84-90.
- [12] 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.
- [13] 胡春雨, 谭金琳. 汉语致股东信中的隐喻及英译研究[J]. 天津外国语学院学报, 2017, 24(1): 25-31.
- [14] 钱嘉慧, 王立非. 英文企业年报语篇中态度评价的语料库考察及分析[J]. 商务外语研究, 2015(1): 1-7.
- [15] Alderson, J.C. (2000) *Assessing Reading*. Cambridge UP, Cambridge. <https://doi.org/10.1017/CBO9780511732935>
- [16] Flesch, R. (1948) A New Readability Yardstick. *Journal of Applied Psychology*, **32**, 221-233. <https://doi.org/10.1037/h0057532>
- [17] Kincaid, J.P., et al. (1975) Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Institute for Simulation and Training, University of Central Florida, Millington.
- [18] Scott, M. (2004) *Wordsmith Tools Version 4*. OUP, Oxford.
- [19] Stubbs, M. (1986) Lexical Density: A Technique and Some Findings. In: Coulthard, M., Ed., *Talking about Text, Discourse Analysis, Monograph No. 13, English Language Research*, University of Birmingham, Birmingham, 27-42.
- [20] Halliday, M.A.K. (2000) *An Introduction to Functional Grammar*. Foreign Language Teaching and Research Press, Beijing.