

基于语料库和Gephi的“一带一路”话语分析

包乐莹

浙江工商大学外国语学院, 浙江 杭州

收稿日期: 2024年3月11日; 录用日期: 2024年4月19日; 发布日期: 2024年4月30日

摘要

十多年来,“一带一路”倡议从愿景到行动发展迅速,并且受到了国内外主流媒体的广泛关注。本研究选取美国主流媒体《纽约时报》自2015年至2023年12月间刊登的21篇关于“一带一路”的报道作为研究样本,运用语料库的方法和可视化软件Gephi,对收集到的文本进行索引行分析和共现分析。本研究发现,《纽约时报》对“一带一路”的话语建构积极、中立与消极并存。本文使用Python对收集的文本进行处理,使用nltk库对文本进行分词、词形还原,并识别词组,去掉其中的停用词。对预处理结果进行词频统计,从中筛选出前100个关键词。对关键词进行共现计算,得到词共现矩阵,再使用Gephi绘图。根据得到的关键词图谱进一步分析《纽约时报》围绕着“一带一路”的报道背后所呈现的国际关系与局势,研究发现这些报道背后隐含了错综复杂的国际关系,中国与非洲、美国、欧洲、印度之间的关系都是这些报道关注的热点。

关键词

一带一路, 纽约时报, 语料库, Gephi

Discourse Analysis of “The Belt and Road” Initiative Based on Corpus and Gephi

Leying Bao

School of Foreign Languages, Zhejiang Gongshang University, Hangzhou Zhejiang

Received: Mar. 11th, 2024; accepted: Apr. 19th, 2024; published: Apr. 30th, 2024

Abstract

For more than a decade, the “the Belt and Road” initiative has developed rapidly from vision to ac-

tion, and has received wide media attention at home and abroad. This study selects 21 reports on the “the Belt and Road” initiative published by *The New York Times* which is the US mainstream media from 2015 to 2023 as research samples. And then this study uses corpus based methods and visualization software Gephi to conduct concordance analysis and co-occurrence analysis of collected texts. The concordance lines of “the Belt and Road” initiative show that there are positive, neutral and negative meanings implied in these reports. This article uses Python to process the collected text, and uses the nltk library to segment the text, restore word forms, identify phrases, and remove stop words. And then this research conducts word frequency analysis on the pre-processing results and selects the top 100 keywords from them. Next, the research performs co-occurrence calculations on the top 100 keywords to obtain a co-occurrence matrix, and then use Gephi to draw the keyword graph. By further analyzing the keyword graph, it can be found that reports related to “the Belt and Road” initiative on the New York Times imply complicated international relations and situations, and the relations between China and Africa, the United States, Europe, and India are the focus of these reports.

Keywords

The Belt and Road, The New York Times, Corpus, Gephi

Copyright © 2024 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

“一带一路”倡议是习近平总书记于 2013 年提出的，包括“丝绸之路经济带”和“21 世纪海上丝绸之路”。“一带一路”倡议自提出以来就得到了国际社会的热烈关注，该倡议的目的在于与沿线多个国家建立经济伙伴关系，以经济为纽带，推动沿线国家的经济交流与政治互信。这一倡议在中外媒体平台都有着较大的关注热度，对于“一带一路”倡议的报道，不同国家媒体有着不同的立场，中国媒体大多从经济角度出发，强调“一带一路”倡议的实施可以促进沿线国家经济的发展，进而促进政治、文化等或各领域的发展；而外国媒体对“一带一路”倡议既有正面报道，也有负面报道，一些媒体从政治与安全视角出发，将“一带一路”倡议视为中国扩大势力范围、争夺地区主导权的战略。本研究将聚焦于 2015 年以来《纽约时报》围绕“一带一路”倡议的报道，基于语料库的方法和网络分析工具 Gephi，使得定性与定量分析相结合，关注报道中体现出的话语特征和话语背后的国际关系。

2. 理论基础

对话语的分析长期以来以定性研究为主，这种定性研究方法因偏重主观性和缺乏代表性而遭到部分学者的批评。使用语料库的方法可以扩大实证基础，减少研究者本身会带有的偏见。Baker & McEnery 提出了值得用语料库方法的几条理由：其一，它至少提供了一个更可信、更大规模的基础，结果昭然若揭。其二，语料库分析可以更巧妙的方式来得出一个明显的结果。诸如主题字、搭配词和检索行分析等技术有助于更详细地了解语言在使用中的功能。其三，语料库分析可以揭示真正令人惊讶的发现，这与我们的预期截然不同[1]。

数据可视化是快速认识、深入了解数据集的捷径，相对于列表，可以更加直观地展示数据内容，挖掘数据内涵。

3. 研究设计

《纽约时报》作为美国具有代表性的报纸，是美国国家精英报的代表，长期以来是国际新闻界的标杆，拥有良好的国际公信力。本研究通过 LexisNexis 数据库收集 2015 年至 2023 年 12 月期间《纽约时报》含有 One Belt One Road, Belt and Road, Silk Road 的报道，共 274 篇。初步分析发现报道大致分为两类：一种是主题报道，即“一带一路”作为文本的主要内容；另一种为素材报道，即“一带一路”仅作为文本素材被提及。在此基础上，为了保证话语分析的深入和聚焦程度，同时确保样本的代表性，笔者从这些文章中抽取 21 篇不同年份的，主题关联度较高的报道建立语料库。本研究将使用语料库分析软件 WordSmith Tools6 对收集的文本进行索引行分析；并通过词频统计，在词频统计表中筛选出前 100 个关键词，导入 Gephi 软件进行共现分析，得到关键词图谱。

研究问题主要涉及三项内容：

- 1) 《纽约时报》对“一带一路”倡议呈现出怎样的态度？
- 2) 《纽约时报》对“一带一路”倡议的报道主要关注哪些问题？
- 3) 《纽约时报》对“一带一路”倡议的报道体现了怎样的国际关系及局势？

4. 实证分析

4.1. 索引行分析

索引分析是语料库语言学的核心，因为通过它可以让我们看到文本中许多重要的语言型式[2]。索引行能够让我们观察到围绕“一带一路”的上下文信息，据此我们可以进一步观察围绕“一带一路”的语义韵，以帮助我们认清话语的本质。在语料库语言学领域，语义韵是指关键词(Key word, 即节点词 Node word)的典型搭配词在其语境中营造的语义氛围[3]。一般而言，语义韵可分为积极、中性与消极三大类。

The screenshot shows the Concordance software interface with a list of concordance lines. The columns include line number (N), the concordance text, and various statistical fields (Set, Tag, Word #, Senti, Para, Hear, Sect, File). The text in the concordance lines includes phrases like 'of what it calls the "One Belt, One Road" initiative', 'in infrastructure, One Belt, One Road also invests', 'package, because One Belt, One Road is nothing less than the', 'have also been wary of One Belt, One Road investment', 'the Chinese initiative, One Belt, One Road, would be the signing of a', 'memorandums, One Belt and One Road. It is a very fertile area for', 'movements. The "One Belt, One Road" initiative — Chinese President', 'leaders to attend the One Belt, One Road conference in Beijing. "Italy can', 'also forward-looking. One Belt, One Road is Beijing's full-throated answer', 'to participate in the "One Belt, One Road" project, the vast, strategic', 'and corporate interests. One Belt, One Road, with its many faces, is neither a', 'Italy's participation in One Belt, One Road, and under secretary of state at', 'and controversial "One Belt, One Road" infrastructure initiative. Most', 'initiative called "One Belt, One Road" would compromise a nation's', 'to Europe under its One Belt, One Road project. Chinese goods would', 'head" of its vast "One Belt, One Road" project, a new Silk Road into', 'president and attend One Belt, One Road forums to draw investment, he', 'and Russia China's "One Belt, One Road" initiative, an economic', 'By contrast, Xi's One Belt, One Road initiative is an enormous', 'this month plugging his "One Belt, One Road" initiative of construction from', 'support for China's "One Belt, One Road" initiative and participated in the'.

Figure 1. Concordance lines of “One Belt, One Road”

图 1. “一带一路”的索引行

媒体在报道外国新闻时依赖于政府的视角和“解读方案”，国际报道是新闻报道中政治影响最强烈的领域，同时，外交政策也在一定程度上影响着国际新闻报道。媒体平台已成为对国际关系和外交政策的不同叙述的战场，所有媒体都戴着他们的“本国眼镜”[4]。因此，《纽约时报》在架构“一带一路”相关新闻时往往采用“以美国为中心”的叙事框架，虽然对“一带一路”有积极和中性的话语表征，但也存在许多消极表征。“以美国为中心”是指以美国的国家利益为中心，根据美国的社会规范及价值观标准来解读。根据图 1 “一带一路”的索引行可知，《纽约时报》对“一带一路”的话语建构，不仅是一种语言上表征意义的实践，更是一种社会实践，通过意义符号的建构投射出作为“他者”的中国，塑造了西方视野下的中国的国家形象。

自“一带一路”倡议提出以来，在全世界范围内得到了“一带一路”沿线国家的高度支持和全方位参与。《纽约时报》中对于我国“一带一路”倡议的积极报道构建了中国互利合作的经济形象与开放包容的政治形象，采用较为客观的态度，报道了“一带一路”倡议下中国与各国的互利合作情况，塑造了中国互利共赢的合作伙伴形象和值得发展中国家信赖的大国形象。比如关于意大利和法国对“一带一路”倡议的态度的报道。2019年3月29日，意大利不顾美国、德国和法国等传统西方盟友的阻挠与质疑，与中国签署了共同推进“一带一路”建设的谅解备忘录，成为G7国家中第一个与中国签署这一合作文件的国家。随后，双方在“一带一路”倡议框架下的合作有条不紊地展开。这一倡议在意大利的顺利推进离不开意大利国内对于该倡议的接受与认可。法国在即使许多西方国家持谨慎态度的情况下，对“一带一路”倡议表示了较为正面的态度，欢迎中国的投资。欧洲作为“一带一路”的终点，如果能够参与到其中来，将受益于这一倡议所带来的交通基础设施建设改善、运输物流升级等，并可以促进欧亚之间的互联互通。中立的报道采取较为客观谨慎的态度对“一带一路”进行评价。如一篇题为：中国的“一带一路”倡议是殖民主义吗？的报道。首先，该报道否认了中国是殖民主义。中国的“一带一路”倡议不仅是基础设施投资项目，也塑造了中国在国际上的良好形象，体现了中国倡导和平，互利繁荣和多文化的包容。同时，该报道也提出要从多方面独立评估“一带一路”的各个项目。其中那些消极的报道则对中国在他国进行的投资活动进行无理诘责，比如将“一带一路”倡议看成中国版的“马歇尔计划”，认为此计划是中国借助于发展基础设施建设，争夺世界更大的经济和政治影响力。在这些媒体人看来，中国为一些欠发达国家提供发展资金贷款就是为他国制造“债务陷阱”，甚至鼓吹“中国威胁论”，将我国的经济活动过度政治化，使得中国的国家形象遭到污名化。

对“一带一路”的索引行分析表明美国主流报刊对中国“一带一路”倡议的报道积极、中立与消极并存，同时流露出一种复杂的心态与不安的心情，这是一种对中国经济、政治日益强大后威胁其自身经济利益的不安。随着中国综合国力的提升和海外影响力的持续扩大，西方社会对中国的态度有所转变，一些报道能够正确看待近年来中国取得的发展成就，肯定中国对世界做出的贡献，但由于固有偏见和刻板印象的存在，仍有不少媒体会进行不实的报道，污名化中国的国家形象。媒体是塑造国家形象的主要载体，这种对于某个国家的国家形象的解读是由多方面因素共同作用的。了解一些西方主流媒体在进行与中国有关内容报道时的传播方式以及态度理念有利于我国根据国际形势调整对外政策，进而塑造良好的国际形象。

4.2. 共现分析

使用 Python 对收集的文本进行处理，使用 nltk 库对文本进行分词、词形还原，并识别词组，去掉其中的停用词，图 2 为上述预处理过程。对预处理结果进行词频统计，从中筛选出前 100 个关键词，对关键词进行共现计算，得到词共现矩阵，具体过程见图 3 和图 4。最后使用 Gephi 绘图，得到的结果见图 5，图中点不同的颜色是模块化后不同类的结果，点的大小是词频，线的粗细是词共现的次数。

```

#预处理
def deal_line(line):
    line = line.lower() #全部转为小写
    line = remove_punctuation(line)
    line = ' '.join(line.split())
    #词组识别
    line = get_phrases(line)
    #分词
    line = word_tokenize(line)
    #词性标注
    line = pos_tag(line)
    #词形还原 (归并)
    wnl = WordNetLemmatizer()
    line = [wnl.lemmatize(ws[0],get_wordnet_pos(ws[1])) for ws in line]
    #短语还原
    line = [replace_phrasesword(w) for w in line]
    #去除停用词
    line = [w for w in line if not w in stopwords and len(w) > 2 and w.isdigit()==False]
    return line
    
```

Figure 2. Pre-process

图 2. 预处理过程

```

#####共现计算#####
comword = [line.strip().lower() for line in open('共现词.txt', encoding='utf-8')]
comword = list(set(comword))
#计算词共现
def get_wordcom(data_cut,comword):
    data_count = pd.DataFrame()
    for word in comword:
        data_count[word] = data_cut.map(lambda x: x.count(word)>0).astype(int)
    comdata = []
    for worda in comword:
        for wordb in comword:
            if worda!= wordb:
                num = data_count[[worda,wordb]].min(axis=1).sum()/2
            else:
                num = data_count[worda].sum()/3
            comdata.append([worda,wordb,num])
    comdata = pd.DataFrame(comdata,columns=['worda','wordb','num'])
    comdata = comdata.sort_values('num',ascending=False)
    return comdata
    
```

Figure 3. Co-occurrence calculation

图 3. 共现计算过程

#	A	B	C	D	E	F	G
1	word	count					
2	china	256					
3	chinese	123					
4	country	60					
5	road	52					
6	governme	52					
7	beijing	50					
8	hong kong	47					
9	italy	39					
10	time	38					
11	india	38					
12	united	38					
13	official	35					
14	investmen	35					
15	leader	34					
16	belt	33					
17	security	33					
18	president	33					
19	company	31					
20	people	30					
21	initiative	29					
22	foreign	26					
23	project	26					
24	modi	26					
25	european	25					
26	economic	24					
27	national	24					
28	port	23					
29	europe	23					
30	american	23					

Figure 4. Result of word frequency (top 30)

图 4. 词频统计结果(前 30 位)

局势的风险。

印度对于“一带一路”倡议的态度也值得关注。在2023年召开的上合理事会上，除了印度之外，其他各成员国都明确表示了对中国“一带一路”项目的支持，并表示将继续推动该项目的发展与完善。印度总理莫迪强调虽然“一带一路”对于促进成员国之间贸易往来具有一定的积极促进作用，但同时也要注意尊重他国主权。其话外之意就是暗指“一带一路”波及到了印度的利益。事实上，在“一带一路”项目刚开始推进的时候，美国和印度是反对态度最强烈的两个国家。美国之所以反对，是因为“一带一路”将极大地加强中国与沿线国家的联系，增强中国的国际影响力，而印度暗中与中国较劲，希望通过美国和西方各国的帮扶，成为所谓的国际大国。

安全是“一带一路”倡议中的一个关键问题，体现在中国人民在巴基斯坦的安全问题。中国与巴基斯坦的合作帮助了巴基斯坦，在这一合作过程中，有不少中国人民常驻巴基斯坦，在“一带一路”建设的最前线奉献自己，为共建“一带一路”和巴基斯坦经济社会发展贡献自己的力量。但是，巴基斯坦国内宗教部族势力、地方分裂势力和犯罪团伙等交织，国际恐怖势力的渗透，使得巴基斯坦恐怖袭击较为活跃，加上一些恐怖分子认为中国在巴基斯坦建设的项目触及其利益，尤其是“中巴经济走廊项目”，对恐怖分子而言，袭击中国项目与公民既能向政府施压，造成广泛的影响，又能对项目进程造成厌恶与破坏。因而尽管巴基斯坦政府对中巴经济合作持积极态度，但是中国人民在巴基斯坦的安全仍是个令人担忧的问题。

5. 结论

本文将定量分析与定性分析相结合，基于语料库分析软件 WordSmith Tools 和可视化软件 Gephi，关注《纽约时报》对“一带一路”的报道，从索引行分析和共现分析两方面进行研究。结果显示，《纽约时报》对中国“一带一路”倡议的报道积极、中立与消极并存，这些报道背后体现了错综复杂的国际关系。

致 谢

本论文在经过修改完善之后终于完成，这份论文将作为我过去半年学业的一份总结。感谢所有的授课老师，在老师们的耐心教导下，我才能在日积月累中学到皮毛，完成这篇论文。同时也要感谢在语料库语言学领域做出宝贵学术研究的学者们，专业的见解、深入的研究结果给我提供了参考资料，是我学习和提升的不竭源泉。

本论文的完成并非终点，文中的不足之处是我未来学习过程中新的起点。

参考文献

- [1] Baker, P. and McEnery, T. (2015) *Corpora and Discourse Studies: Integrating Discourse and Corpora*. Palgrave Macmillan, London. <https://doi.org/10.1057/9781137431738>
- [2] 钱毓芳. 语料库与批判话语分析[J]. 外语教学与研究, 2010, 42(3): 198-202, 241.
- [3] 要文静, 范京晶. 不同语域视角下程度副词 too 语义韵研究[J]. 长江大学学报(社科版), 2016, 39(1): 64-69.
- [4] Hillel, N. (2004) Our News and Their News: The Role of National Identity in the Coverage of Foreign News. *Journalism*, 3, 343-368. <https://doi.org/10.1177/1464884904044941>
- [5] 央视网. 中非合作论坛北京峰会成果丰硕: 37 国签署共建“一带一路”备忘录[EB/OL]. <http://news.cctv.com/2018/09/08/ARTIfObtrHDnBXR8tvESeFq4180908.shtml>, 2018-09-08.