



# A Corpus-Based Analysis of Discourse on Tangshan Beating Incident in Foreign Social Media

Jie Huang

School of Foreign Languages, East China University of Science and Technology, Shanghai, China  
Email: 807795828@qq.com

**How to cite this paper:** Huang, J. (2024) A Corpus-Based Analysis of Discourse on Tangshan Beating Incident in Foreign Social Media. *Open Access Library Journal*, 11: e1066.  
<https://doi.org/10.4236/oalib.1111066>

**Received:** November 30, 2023

**Accepted:** January 16, 2024

**Published:** January 19, 2024

Copyright © 2024 by author(s) and Open Access Library 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

## Abstract

Based on the corpus-based critical discourse analysis, this paper combines quantitative and qualitative method to analyze the tweets on Tangshan Beating Incident. By observing word list and keyword list, it is found that tweets prefer *incident* in which *police* is put in negative contexts to *case* in which *police* in positive contexts. Moreover, using passive *arrested* emphasizes the inaction of Chinese police and highlight the negative image. For the analysis of the use of *CCP*, tweets link the beating incident with China and its political parties and maliciously discredit the Communist Party of China by being straightforward; affirming status but actually irony; distorting facts and associate thought. Comparing *woman* and *man*, it turns out that the tweets stress the quantitative and physiological advantages of men and weaken the fact of their violence, which indicates discrimination against women. Faced with such situation, it is of great significance to enhance China's ability to build its foreign image and improve the external public opinion environment.

## Subject Areas

Linguistics

## Keywords

Corpus Linguistics, Discourse Analysis, Tangshan Beating Incident, Twitter

## 1. 研究背景和理论基础

2022年6月10日凌晨,河北省唐山市机场路的一家烧烤店发生了一起涉嫌寻衅滋事、暴力殴打他人的案件,涉案监控视频在网络上曝光后,引发群众热议,得到了全国人民的关注。6月10日下午5时,唐山打人事件登上热

搜，霸占了各大媒体和短视频的热搜头条，社会舆论高度关注，大量网友对打人者表达了强烈的愤怒。打人事件不断发酵，引发全社会对案件侦办工作的高度关注。截至 7 月 31 日，新浪微博上“#唐山打人#”的话题已获得 416.1 万讨论和 47.1 亿阅读。

话语是一种交际行为或交际事件，即包含语言形式、意义和行为的复杂系统(Van Dijk, 1988) [1]，是语言表述和社会事件过程的结合体。话语分析是从语言入手对于社会事件本身的关注，实质是探讨话语和社会之间复杂的理论联系(Van Dijk, 1997) [2]。批评话语分析(Critical Discourse Analysis, 以下简称 CDA)和语料库语言学(Corpus Linguistics, 以下简称 CL)都承认语言的社会属性，因此也都将自然发生的语篇作为研究对象，但 CDA 一直以来受到诸多批评。Widdowson (1995) [3]、Stubbs (1997) [4]、Toolan (1997) [5]等学者认为，CDA 所分析的文本数量过小，而且缺乏代表性，对话语的解释又缺少客观性和系统性，并非是在“认真的”、“严格的”、“系统的”语篇分析的基础之上进行社会批判，而是出于某种政治目的，在语篇中找到希望缺席或者在场的语言证据，具有循环论证之嫌。Chouliaraki 和 Fairclough (1999) [6]在对 CDA 进行反思时承认，CDA 一直以来都因为没有对大规模的、有代表性的语篇进行系统分析而受到批评，包括定量的和计算方法的使用，这实际上能够对 CDA 的关于话语的社会主张提供坚实的语言基础。

语料库的方法有助于弥补 CDA 的不足，一是在于 CL 基于大批真实、有代表性的语料；二是 CL 使用计算机辅助的研究手段，通过检索来提取批量的信息，再做进一步量化分析和定性研究；三在通过实证性研究，能总结和发现具有倾向性的话语模式。话语是意识形态的反映，将语料库的发现与社会情境相结合，在繁复大量的语言资料中挖掘共性与规律，从而提炼关键特征，理解语言背后的深层含义，归纳出“话语模式”(辛斌、高小丽, 2013) [7]，能为抽象的话语解释提供事实依据，使研究结果更客观、可靠，有代表性。

推特(Twitter)是美国主流社交媒体平台，受众广，拥有超过 10 亿的注册用户，1.92 亿日活跃用户和 3.53 亿的月活跃用户。作为社交平台，推特具有真实、非权威、碎片化的特点，建构出西方国家的民间舆论场。本研究基于语料库检索工具 AntConc3.5.8 对推特关于唐山打人事件的话语进行词频分析和主题词分析。

## 2. 文献综述

国内针对社交媒体平台中话语的语料库研究较少，目前查到的文献国内主要是基于微博，国外主要基于推特平台。截至 2022 年 8 月，中国知网中共检索到 5 篇基于语料库对微博话语的研究和 3 篇对推特话语的研究。焦新平，敖锋(2018) [8]基于语料库对军事微博进行话语研究。但研究中只用了一次语料库 Wordsmith 的检索功能，而花很大的篇幅对具体微博实例进行定性分析，且主观随意性强，没有交代选择这些博文的原因。杨凯和陈滢钧(2020) [9]以 2010 到 2019 年推特涉穗话语为语料，自建语料库，进行词频、搭配词、索引行和情感词分析，总结反映出的广州城市形象特征和传播特点，探究广州城市形象在推特平台上所呈现出的不足并提出未来改进方向。这两个研究的共

同问题是没有充分利用社交平台时效性强的优势，关注点没有放在社会热点事件上。张瑜,刘思雯(2021) [10]将研究重点聚焦新冠疫情这一时下热点话题，基于评价理论，将疫情分为六个阶段，通过主题词，和“China”相关的高频形容词，印象词云和检索行探究推特平台上美国民众对于中国关注点和中国形象的历时性变化。

近期，唐山打人事件获得了较强的社会关注度和公众参与度，本研究关注该事件在国外主流社交媒体上的反响，通过归纳总结推特中关于唐山打人事件的推文表层的话语模式，揭露背后西方社会群众眼中的中国形象和反映出的意识形态。

### 3. 研究方案

本研究以 2022 年 6 月 10 日至 7 月 31 日美国主流社交平台推特中关于唐山打人事件的推文为语料，自建小型语料库。以 Tangshan 和 beating 为检索项，由于本研究仅考察国外网友和媒体对唐山打人事件的评价，筛选出所有英文推文后，删去其中 15 篇来自中国官方媒体的推文。共有来自 174 个用户的 448 篇推文，平均每天发布 9 条关于唐山打人事件的推文。不将图片，视频及其他外部链接统计入内，总计 13990 形符，2361 类符，通过借助语料库软件 Antconc3.5.8w 中的高频词列表(Word Lists)、索引(Concordances)、搭配词(Collocates)、词簇(Clusters)、主题词(Keywords)等分析工具，对语料库中的推文进行分析，提取语言形式上的突出特征，分析国外社交媒体对唐山打人事件的话语策略其评价态度。

## 4. 结果和发现

### 4.1. 词频分析

话语分析要对文本的“语言运用单位进行清晰的、系统的描写”(Van Dijk, 1988) [11]。词频是语料库中最基本的统计方法。通过筛选高频实词，能够直观了解话语的社会情景与情感倾向。将高频词按照词根进行排序，选取了前 20 项，其中 8 项是实义词(见表 1)。

表 1. 高频词表

排序	频次	词根	词形
1	1091	the	the 1091
2	575	<b>tangshan</b>	tangshan 575
3	540	<b>beat</b>	beat 21 beaten 33 beating 484 beatings 2
4	497	of	of 497
5	441	in	in 441
6	439	be	am 4 be 67 being 23 is 177 was 96 were 72
7	306	a	a 306
8	293	to	to 293
9	269	and	and 269

续表

10	263	<b>incident</b>	incident 255 incidents 8
11	207	<b>woman</b>	woman 75 women 132
12	151	have	had 12 has 74 have 63 having 2
13	131	on	on 131
14	111	<b>china</b>	china 111
15	109	<b>people</b>	people 109
16	108	<b>case</b>	case 103 cases 5
17	97	s	s 97
18	92	that	that 91 thats 1
19	90	<b>man</b>	man 30 men 60
20	80	at	at 80

将这些实词进行分类, 主要涉及到唐山事件的事件、地点和人物(见表 2)。其中, 表示事件的 **incident** 和 **case** 是近义词, 中文都可翻译为“事件”。男性和女性, 施暴者和受害者作为事件中对立的双方, 也在表中同时出现, 本文将借助 Antconc 语料库软件对这两组概念进行索引行分析和进一步讨论。

表 2. 高频实词分类

事件	Beat, Incident, Case
地点	Tangshan, China
人物	Woman, People, Man

#### 4.1.1. Incident 和 Case 对比分析

表示事件的 **incident** 和 **case** 意义相近, 考察 **incident** 和 **case** 的常用搭配, 跨距设为-1/+1, 发现他们最常搭配的前五个词中都有 **beating**, **the** 和 **of**, 且 **beating** 同时是 **incident** 和 **case** 最常搭配的词, **incident** 和 **case** 主要有两个类联接 **N + incident/case**, **incident/case + of**。将 **incident** 和 **case** 进行词丛分析, 词丛长度选择 2 到 5, 结果如表 3。可见两个词的搭配实例和用法有着高度重合。不同的是, 在语料库中, **incident** 出现频率更高, 是 **case** 的两倍之多, 且 **incident** 与 **beating** 和 **of** 搭配出现的次数更多, T 值更高(见表 4、表 5, T 值表示与节点词的共现率, T 值高的搭配倾向于使用高频词汇)。本研究将继续探究此现象的原因。

表 3. Incident 和 case 的高频词丛

频数	词丛	排序	词丛	频数
191	beating incident	1	beating case	68
103	tangshan beating incident	2	tangshan beating case	53
53	the tangshan beating incident	3	the tangshan beating case	22
49	the beating incident	4	the case	18
25	the incident	5	the beating case	9

表 4. Incident 搭配词列表

排序	搭配词	T值	频数	频数(左)	频数(右)
1	beating	13.36986	196	196	0
2	the	3.63532	44	25	19
3	in	5.14776	41	0	41
4	of	3.82329	30	0	30
5	at	4.49178	23	0	23

表 5. Case 搭配词列表

排序	搭配词	T值	频数	频数(左)	频数(右)
1	beating	8.06533	72	72	0
2	the	2.97791	22	18	4
3	of	3.08522	16	1	15
4	is	2.75019	10	0	10
5	were	2.23308	6	0	6

牛津词典对 incident 的解释是 a serious or violent event, such as a crime, an accident or an attack, 而 case 的定义是 a matter that is being officially investigated, especially by the police。柯林斯词典中, incident 指 something that happens, often something that is unpleasant, 而 case 的定义是 a crime or mystery that the police are investigating。可见对 case 的解释两者都强调了警察和调查。根据定义里的关键词, 在 incident 和 case 的上下文中检索 police, 上下文跨距设为-5/+5, 发现 police 在 case/cases 的上下文中出现 6 次, 而在 incident/incidents 的上下文中出现 11 次。police 在 incident 的语境中出现更多, 与词典定义中强调的内容有所出入。

继续对 incident 和 case 上下文中包含 police 的索引行进行分析, 由于推文被限制在 140 个字符之内, 部分推文不完整, 限于篇幅, 只展示部分索引的部分内容(见表 6)。考察推特用户对于警察的语义韵倾向, 发现在 incident 的上下文中, 与 police 搭配的动作具有消极的倾向, 如 many police chiefs in Tangshan were reported (2 例), The police stood by and nobody came, 警察局长被告发, 警察袖手旁观, 无人问津。警察被推特网友描述成不负责任, 以身试法的负面形象, 具有明显的消极特点, 使整个语境弥漫一种消极的语义氛围。当然, report 也有中性的意思, 指汇报, 报告, 如 The police reported the Tangshan beating incident, police report is coming 这两处与 police 搭配的是中性词。

表 6. Incident 和 case 上下文中包含 police 的索引

Incident 上下文中 有 police 的部分索引行	Case 上下文中 有 police 的部分索引行
Whole network is angry! The <b>police</b> reported the Tangshan beating <b>incident</b> !	The <b>police</b> in the Tangshan beating <b>case</b> revealed that forensic tampering;

续表

In Tangshan, there was another <b>incident</b> of beating girls. The <b>police</b> stood by and nobody came to help.	Hebei police authorities said on 11.3 people from other provinces involved in the #Tangshan beating <b>case</b> were arrested, the <b>police</b> have arrested 8 people involved in the case.
#Tangshan beating <b>incident</b> <b>police</b> report is coming	Tangshan beating <b>case</b> was investigated by Langfang <b>police</b> !
One of the police officers of the Tangshan mass beating of women <b>incident</b> : when the <b>police</b> arrived at the scene, the attacker had fled	Visiting the scene of Tangshan beating <b>case</b> , the <b>police</b> said that the police car arrived at the on-site hospital more than 10 minutes later
After the vicious beating <b>incident</b> , many <b>police</b> chiefs in Tangshan were reported	Hebei Tangshan beating <b>case</b> investigated by Langfang <b>police</b> #in westigation #langfang #Tangshan

还有的与 **police** 共现的词语表面上看上去是中性的，而实际上以一定方式与消极词语相联系，因此受到消极语义传染，也具有消极涵义。如“Because no gun is not equal to safety, referring to the beating incident in Tangshan, China these days, the attitude of the police after the incident, I completely lost hope for the COMMUNIST Party of China.” 这条推文没有指明警察的态度，但从“我”的反应中体现了对中国共产党和警察的负面评价。总体来看，在 **incident** 的上下文语境中，**police** 具有中性涵义或消极涵义，但没有积极涵义，显示了一种消极语义韵。而在 **case** 的上下文语境中，**police** 则具有明显的积极涵义。如 Tangshan beating case investigated by Langfang police (2 例)，the police in the Tangshan beating case revealed that forensic tampering, the police have arrested 8 people involved in the case，警方调查案件，揭露违法行为，逮捕罪犯，主动出击，积极履行职责和义务，建构了一种积极正面的形象。

再根据定义中另一个关键词 **investigated**，在 **incident** 和 **case** 的上下文中检索 **investigate**, **investigates**, **investigated**, **investigating** 和 **investigation**，发现在 **case** 的上下文中出现 11 次，而在 **incident** 的上下文中仅出现一处，与词典定义中强调的内容相吻合。可见，**case** 和 **investigate** 语义紧密相关，强调警方处理问题，依法侦查办案，转危为安的过程，表现出积极的语义韵；而 **incident** 强调事件的严重性和恶劣程度，弱化了对事件发生后对人们的应对和解决措施的描述，具有消极的语义韵。

对比分析 **incident** 和 **case**，本研究发现，语料库中 **incident** 比 **case** 出现的频率更高，且与 **beating** 和 **of** 搭配的共现率更高。在 **incident** 的上下文中，与 **police** 搭配的动作具有消极的倾向；而在 **case** 的上下文语境中，**police** 则具有明显的积极涵义，推文通过多用 **incident** 表达对中国警方在打人事件中的消极评价。

#### 4.1.2. Woman 和 Man 对比分析

Woman 和 man 都出现在高频词表中，且都多以复数形式出现，

woman/women 的频次是 man/men 的两倍之多，可见外国网友的焦点聚焦于受害女性，对女性的描述更多。对 woman/women, man/men 进行词丛分析，词丛长度选择 2 到 5，选取前十项进行对比。

从表 7 可以看到，在修饰男性数量时，单数有 a man，复数有 group of/several men；而在修饰女性数量时，单数用 a/the woman，复数则直接用复数形式 women，不加任何修饰或限定词，凸显了男性在唐山事件中的数量优势。beating man 和 beating women 都出现在此表中，对此进一步分析发现，beating 既可以作形容词，后面加动作的施事者，即主语；也可以作动名词，后面加动作的受事者，也就是宾语。在这里，beating man 里的 beating 作形容词，而 beating women 里的 beating 作动名词。但在实际使用中，beating 做形容词的用法并不常见，大多数情况下，beating 用作动名词的情况更多，后面接的人是被打的人而不是打人的人。因此 beating man 可能会误导不了解该事件的网友看到认为被打的是男性，有帮施暴男性脱罪之嫌。此外，在词丛中，man 前面有两处用实意形容词修饰，分别是 beating 和 strong，而 woman 前则没有，凸显了唐山事件中男性的生理优势。

表 7. Woman 和 man 的高频词丛表

频数	woman/women的词丛	排序	man/men的词丛	频数
35	a woman	1	of men	22
28	of women	2	a group of men	15
23	beating women	3	group of men	15
16	beating a woman	4	a man	9
16	beating of women	5	saw a group of men	5
11	up women	6	several men	5
10	the woman	7	beating man	4
8	beating up women	8	strong man	4
7	and beating women	9	the beating man	4
7	incident of beating a woman	10	beating incident. men	3

唐山打人事件中犯罪男性骚扰、侵犯、侮辱女性，严重威胁挑战女性的社会地位。可以看得到，虽然国外网友更加关注女性，花更多的篇幅提及 women，但在行文中还是不经意间突出唐山事件中男性的数量优势和生理优势，自然地流露出对男性的看好和偏爱。文明社会男女平等说起来容易，但在实现真正的平等的过程中我们还有很多路要走，至少在语言上还没有完全做到。不仅女性要团结起来，维护自己的基本权益，语言文字工作的部门也要努力减少语言中的性别歧视，让女性获得精神和物质上的双重平等。

#### 4.2. 主题词分析

主题词(keywords)是与合适参照库相比，出现频率远超常态的词汇(Scott

& Tribble, 2006) [12]。实词主题词可以突显文本焦点内容并可以为其他研究提供出发点(李晓倩、胡开宝, 2017) [13]。本研究将 Brown Corpus 作为参照语料库, 获取本语料库的主题词(Brown Corpus 是在 1960 年代中期构建的一个具有代表性的平衡语料库, 合计约 100 万单词)。将主题词按照关键性排序, 计算方法选择卡方检验,  $p < 0.05$ , 选取前 20 项, 见表 8。

**表 8.** 主题词表

排序	关键性	主题词
1	14149.74	tangshan
2	11907.36	beating
3	5869.5	incident
4	2952.04	women
5	2650.36	china
6	1916.78	barbecue
7	1889.53	restaurant
8	1645.1	police
9	1629.58	woman
10	1362.76	people
11	1277.76	video
12	1275.67	chinese
13	1256.84	case
14	1081.16	hebei
15	981.5	arrested
16	810.84	beaten
17	809.73	media
18	784.1	girl
19	770.39	security
20	761.7	CCP

表 1 中的实词除了 man 都出现在表 8 中, 其他主题词也突显了语料的焦点内容。Arrested 在语料库中出现 41 次, 左侧出现 be 动词的情况有 27 次, 见表 9。Arrest 的动作发出者是 police, 打人事件发生后, 警察在次日将九名逃离现场的犯罪嫌疑人全部捕获, 足见中国警方齐心协力, 办案效率高。而在 arrested 的上下文语境(跨距设为-5/+5)中 police 仅出现 11 次。被动语态掩盖了动作执行者, 让国外网友感觉不到河北警方在案件侦查过程中的行动和作为。Beaten 和 beating 同时出现在主题词表, 但与动名词 beating 比起来, 被动态 beaten 在语料库中出现较少, 只有 33 次。这证实了邵惟讷和邵志洪(2015) [14]的观点, 英语多用名词以体现其静态特征。在 beating 的上下文语

境(跨距设为-5/+5)中检索 man/men, 结果只有 39 条, 不到总数的十分之一。此外, 被动态 beaten 的选择凸显出女性被动、挨打的受迫害局面, 但 beaten 右侧检索 by(范围设为 0/+5)只有 4 处, 可见推文用 beaten 和 beating 的同时, 意图隐瞒男性是施暴者的身份信息, 有意为施暴男性洗白辩护。

表 9. Arrested 左侧搭配词表

arrested左一单词	频数
been	13
were	9
being	3
was	2

主题词 media 可以体现唐山事件得到国内和国外官方媒体和社交媒体的关注; security 既用在短语 public security bureau 中, 指公安机关公安局, 又暗示公众的安全感受到威胁和挑战。还有一些主题词因出现频次不高, 且缺乏分析的价值, 这里不再详细讨论, 下文重点讨论主题词 CCP 的索引行。

CCP 是 Chinese Communist Party 的缩写, 指中国共产党。西方国家在提到中国共产党时, 常用这种表述。但该说法是不正式、不符合国际惯例的。中国共产党官方认可的英文表述是 the Communist Party of China, 官方简称是 CPC。建党初期中国共产党在力量上比较弱小, 当时选用了 CCP 作为中文译本, 但后来党不断发展, 力量不断壮大, 中国共产党把英文名称改成了 CPC。从语言角度看, Chinese 表示与中国有关的东西, 可以指中文或者中国人民, Chinese Communist Party 的重点是放在 Communist Party, 而 the Communist Party of China 的关键在 of China, 体现出这个政党是属于中国的政党、是为中国奋斗的政党, 而 Chinese Communist Party 相对而言弱化了这一从属关系。

在语料库中对 communist party 进行检索, 共有 6 条, 其中四条用的是 the Communist Party of China, 两条用的是 Chinese Communist Party。这表明, 外国网友实际上清楚地知道中国共产党的官方表述, 并在多数情况下能正确使用全称, 但在语料库中检索简称时, 只能找到 CCP, 并未检索到 CPC 的词条, 让人不禁疑惑, 国外网友是否有意而为之, 故意用 CCP 传达关于中共的负面内容。结果发现并不是凡用 CCP 的推文就一定充满了负面色彩, 用 CPC 就是正面的, 具体是正面还是负面还是由具体内容决定。下文将继续对 CCP 索引行的内容进行分析, 见表 10。

表 10. CCP 的部分索引行

序号	索引行
1	Using the “Beating Women at Tangshan BBQ Restaurant” Incident, <b>CCP is to Launch a Political Movement—Whistleblower Movement</b>
2	Shenzhen woman accuses her of being sexually assaulted, police ask her to apologize to rapist <b>The Tangshan beating case has drawn attention to the tragic situation of Chinese women, and the CCP has been accused of being the general backstage of the underworld.</b>

续表

3	<u>#China under #CCP is a joke, #ZeroCovid biohazard members are gangsters with criminal charges, involving #TangShan beating case.</u>
4	Another slave who kneels and licks the CCP.
5	Tangshan is a major force in China's underworld to expose the inside story of the <u>CCP's underworld</u>
6	The locals in #Tangshan send flowers to the next door of the crime scene of 6/10 beating. They witnessed other 2 girls beaten to death here, no videos available. <u>#CCP won't let people know it. #China, an evil country!</u>
7	Tangshan beating case is full of doubts: "No one can be safe under the rule of the CCP" <u>The Tangshan beating incident is still fermenting, but the authorities have begun a hot search to cover up the injury and current situation of the person involved.</u>
8	Huawei and ZTE are completely out of action; the <u>CCP has entered a period of collapse to support the scene with fake party member numbers; the police in the Tangshan beating case revealed that forensic tampering;</u>
9	All the chaos in the <u>CCP is unsolved. The "chain girl" incident in Fengxian County, Xuzhou; the "girl beating up" incident in Tangshan; Shanghai men slaughtering women in the street; Chongqing traffic police shooting and killing people.</u>
10	The <u>#CCP police department are storytelling to coverup this #TangShan beating case, media and the internet are both censored. Victims families are silent or threatened. Usual #CCP Poohland tactics, gangsters police are related.</u>
11	[Tangshan# Beating People Case Tens of Thousands Accounts Banned and Netizens Ridiculously: Thieves Shout to Catch Thieves]
12	The <u>CCP's new regulations strictly control thread comments, netizens ridicule: the Internet is also surnamed the party</u> When the Tangshan beating incident caused public grievances, the <u>CCP's Cyberspace Administration of China issued new regulations to strengthen the management.</u>

CCP 和 Communist Party 在语料库中共出现了 39 次，分布在 35 篇推文中。有两处以 CCP China 的形式出现，因此在主题词表中未显现。在这 35 篇推文中，有 18 篇同时出现了 China, Chinese, PRC, country 的字眼，有党才有国，政党和国家密不可分，可见外国网友对中国共产党和中国意识形态的关注度高，将党的形象与中国国家形象形成很高的关联度。探究推文中对于 CCP 的语义韵倾向，35 篇推文中 19 篇表现了明显的消极语义韵，对中共进行负面评价，控诉揭露中共的“恶行”。

从话语策略上看，推特网民主要使用了直白；肯定地位，实则反讽；歪曲事实，传播谣言；联系扩展思维四种话语策略抹黑中国共产党。

首先是直白，国外新闻媒体在中国问题上往往谨言慎行，喜欢采用迂回的话语平衡策略。相较之下，社交媒体对中国的评价更加简单直接。如 joke (3), evil (6), collapse (8) 等词充满敌意和攻击性，挑衅意味明显，火药味十足，可见国外社交媒体直接抹黑攻击中国和中共。究其原因，社交平台网络

言论的匿名性和不可追踪的特点带给网民更大的自由，网络舆论无人监管，网民不用对自己的发言负责，用户更能坦白直接地表达自己对政治外交等敏感问题的态度。

其次是肯定地位，实则反讽，如 4 中 *kneels and licks* 一方面肯定了中共在中国的地位和身份，中国共产党是中国的执政党，党领导国家；同时 *slave* 则从侧面表现出共产党在中国“奴役百姓”，“专制独裁”的形象。在西方社会眼中，中国共产党在中国“一家独大”，在事务决策上“垄断”，缺乏权力制衡和约束机制，十分“危险”。再如 1 中 *Whistleblower Movement* 指唐山事件发生后多人实名举报当地黑恶势力，充满讽刺意味，实则指责嘲讽共产党的“治理不力”和“无能”，是一种黑色幽默。

此外，这些推文中普遍存在歪曲事实，传播谣言的现象。6 和 7 中的谣言都围绕被打女性的伤势，事实上，被暴力殴打的四名妇女是受伤，目前伤情稳定，并没有面临生命危险或死亡。警察和当局在事后也进行了案件侦办进展情况的通报，犯罪嫌疑人于 6 月 11 日全部逮捕落网。10 中认为受害者家属被威胁，警察编造事实，掩盖真相，警匪相互勾结，这实属是无凭无据的污蔑诋毁，散布谣言。

最后是善用联系思维，将社会恶性事件互相关联，过度泛化。如 2 由唐山事件中被打的女性联想到深圳女子被强奸，警方让她给强奸犯道歉的事件。9 中列举了上半年来中国各地发生的社会恶性事件，如徐州丰县生育八孩女子事件，上海金山男子当街砍人事件，重庆北碚警察枪击事件。误导网友将个别社会不良事件当作是一般普遍现象，让他们认为中国在中国共产党的领导下社会恶性事件频繁发生，对中共产生质疑和否定。事实上这几年，随着扫黑除恶专项斗争的开展，在中国共产党的领导下，社会治安综合治理逐渐加强，中国已经成为世界上最有安全感的国家之一，城市的治安越来越好。只是比起积极的正面新闻，人们更容易被负面新闻或报道吸引视线，过度关注，举一反三，从而缺乏理性思考和判断。报道出来的不代表是普遍存在的，未被报道的不等于就不存在。作为信息的接收者，更要理性思考，尊重事实。

从内容上看，中共被“指控”是黑社会的后台(2 和 5)，编造事实的独裁者(4 和 10)，善于封锁消息，控制网络舆论(11 和 12)。在前面的分析中已经解释过，黑社会和掩盖事实是抹黑污蔑。至于最后一点，在唐山事件中，确实对网络舆论进行了严格的控制和管理，但这主要针对发表的不当言论和不实信息，如个别网友发布被打女子已死、某网友是打人者等不实信息，散布谣言、混淆视听、扰乱传播秩序。为了避免三人成虎，引起更多网友不必要的恐慌，所以处置了相关违规内容，清理了谣言，这是正当合理的，却被有心人加以利用，恶意曲解。

总的来说，推文常常将地方性社会不良事件与国家或政党挂钩，与中国的意识形态紧密相连，通过直白；肯定地位，实则反讽；歪曲事实，传播谣言；联系扩展思维等话语策略丑化中国共产党的形象。究其原因，中西方在政治制度上存在差异，且西方社会对中国长久以来有误解和偏见。面对来自国外社交媒体的诟病抹黑，中国首先要正己，对于社会黑恶势力，要从严从快及时依法惩处，让施暴者受到法律严惩。相关事件的信息发布要做到公开、

透明、及时。此外，要积极扩散对外正面言论，以对抗负面评价。在推特上关于唐山打人事件的 448 篇推文来自国外，只有 15 篇来自中国官媒，可见对国外社交平台话语的重视有待提高。归根到底，保障百姓的安全感，提升民众对党和政府的信任度和信赖感，回应社会关切，让人民感受到公平正义，才能以防此类暴力事件再度发生，落人口舌。

## 5. 结论

本研究从微观语言结构的角度入手，关注与现实接轨的社会问题，基于 Antconc 软件，自建语料库，对国外社交媒体推特中关于唐山打人事件的话语进行分析。

通过观察高频词表，对比了 incident 和 case, woman 和 man 这两组词。结果发现 incident 和 case 在搭配和用法上相似度高，但表达事件时推文倾向用 incident 而不是 case，因为 incident 的上下文中，police 多表现出消极语义韵，而在 case 的上下文中，police 多表现积极语义韵。此外，在表达逮捕时多用被动态的 arrested，但在上下文很少提到动作执行者——警方，强化警察的不作为，弱化其侦察行动，突出中国警方的负面形象。Woman 和 man 是打人事件中对立的双方，表面上国外网友更加关注女性，花更多的篇幅提到 women，但在用词上突出了唐山事件中男性的数量优势和生理优势，不经意表现出对男性的偏好。此外多用被动态 beaten，不说明施暴者是男性，有意为施暴男性脱罪，对女性充满歧视。

观察主题词表，与平衡语料库相比，推文凸显 CCP，将唐山打人事件与中国的国家和政党挂钩，通过直白；肯定地位，实则反讽；歪曲事实，传播谣言；联系扩展思维等话语策略恶意丑化中国共产党的形象。

总之，应对唐山类似事件，全面整治社会治安，积极开展“雷霆风暴”专项行动，进一步扫黑除恶很有必要。提高政府治理能力才能从民间自下而上地改变外部舆论环境，避免落人口舌，借题发挥。同时要提升我国对外国家形象构建能力，不断改善和消除中国和中共在西方国家民众心中的刻板印象，构建正面且负责任的大国形象。

## Conflicts of Interest

The author declares no conflicts of interest.

## References

- [1] Van Dijk, T. (1988) *News Analysis: Case Studies of International News in the Press*. Lawrence Erlbaum Associates Publishers, Hove and London.
- [2] Van Dijk, T. (1997) *Discourse as Social Interaction*. Sage Publication, London.
- [3] Widdowson, H.G. (1995) *Discourse Analysis: A Critical View*. *Language and Literature*, 4, 157-172. <https://doi.org/10.1177/096394709500400301>
- [4] Stubbs, M. (1997) Whorf's Children: Critical Comments on Critical Discourse Analysis. In: Ryan, A. and Wray, A., eds., *Evolving Models of Language*, Multilingual Matters/BAAL, Clevedon, 1-15.
- [5] Toolan, M. (1997) *What Is Critical Discourse Analysis and Why Are People Saying*

- such Terrible Things about It? *Language and Literature*, 6, 83-103.  
<https://doi.org/10.1177/096394709700600201>
- [6] Chouliaraki, L. and Fairclough, N. (1999) *Discourse in Late Modernity: Rethinking Critical Discourse Analysis*. Edinburgh University Press, Edinburgh.
- [7] 辛斌, 高小丽. 批评话语分析: 目标、方法与动态[J]. 外语与外语教学, 2013(4): 1-5+16.
- [8] 焦新平, 敖锋. 基于语料库的军事微博话语研究——以@军报记者为例[J]. 外语研究, 2018, 35(1): 8-12+74.
- [9] 杨凯, 陈滢钧. 海外社交平台的广州城市形象呈现——基于推特十年涉穗话语语料库分析[J]. 国际传播, 2020(3): 76-85.
- [10] 张瑜, 刘思雯. 评价理论视角下中国国家形象网络媒体话语建构——基于语料库的美国民众推特社交平台疫情话语分析[J]. 西部学刊, 2021(17): 146-151.
- [11] Van Dijk, T. (1988) *News as Discourse*. Erlbaum, Hillsdale, NJ.
- [12] Scott, M. and Tribble, C. (2006) *Textual Patterns: Key Words and Corpus Analysis in Language Education*. John Benjamins Publishing Company, Philadelphia.  
<https://doi.org/10.1075/scl.22>
- [13] 李晓倩, 胡开宝. 中国政府工作报告英译文中主题词及其搭配研究[J]. 中国外语, 2017, 14(6): 81-89.
- [14] 邵惟斌, 邵志洪. 静态与动态——传统和认知语法视角下的英汉语言表达状态对比[J]. 外国语文, 2015, 31(1): 98-105.

## Appendix (Abstract and Keywords in Chinese)

### 基于语料库的国外社交媒体中关于唐山打人事件的话语研究

**摘要:** 本文采用基于语料库的批判话语分析方法, 将定量与定性研究相结合, 对国外社交媒体——推特中关于唐山打人事件的推文进行话语分析。观察高频词表和主题词表, 发现推文在 incident 和 case 中多用 incident, 表现出 police 的消极语义韵; 此外, 用被动态的 arrested 强化警察的不作为, 弱化其侦察行动, 突出中国警方的负面形象。接着对 ccp 分析, 推文将唐山打人事件与中国的国家和政党挂钩, 通过直白; 肯定地位, 实则反讽; 歪曲事实, 传播谣言; 联系扩展思维等话语策略恶意抹黑中国共产党。对比 woman 和 man, 发现推文突出了男性的数量优势和生理优势, 弱化其施暴事实, 对女性充满歧视。对此, 提升我国对外国家形象构建能力, 改善外部舆论环境具有重要意义。

**关键词:** 语料库语言学, 话语分析, 唐山打人事件, 推特