

# *Reinterpreting the Concept of “Empire” in 19th-Century British Social Contexts Based on Corpus Evidence*

**Junjie Zhao**

*Sichuan University, Chengdu, 610200, Sichuan, China*

**Keywords:** 19th-century Britain; empire; corpus linguistics; discourse construction; semantic prosody

**Abstract:** This paper develops a corpus consisting of around 5 million words from the British imperial discourses at the end of the 19th century, which covers three main areas: official politics, royal correspondence and public intellectual debate. Using the GraphColl and KWIC facilities of the Lancsbox software, based on keyword collocation analysis and semantic prosody calculation, this study explores how the semantic development of the term “empire” has been shaped by the 19<sup>th</sup>-century British society as a whole. It is revealed that there were some distinctive features associated with different stages of the connotation of the British Empire depending on the changes of socio-historical background.

## **1. Introduction**

The 19<sup>th</sup>-century is an important chapter of British history. During the course of that century, Britain evolved from being one of Europe’s great powers to becoming a world colonial empire, making “empire” an important vocabulary item in its political/social lexicon. Prior work about this idea mostly relies on qualitative interpretation of historical texts, leaving systematic and temporal studies of word meanings in large corpora still underserved. Following a tradition from corpus linguistics, this research constructs a special database in order to qualitatively investigate the application status, meaning development, and sentiment tendency of the term “empire” during the Victorian Britain period, thus recovering a version of itself as it was perceived at that particular period.

## **2. Research Design**

### **2.1 Corpus Construction**

This study constructs the Nineteenth-Century British Discourse on Empire Corpus (NBDEC), containing a total of approximately 5 million words. To ensure the multidimensionality and representativeness of the data, the corpus is divided into three sub-corpora that comprehensively cover the core producers of 19th-century British imperial discourse.

(1) Harvard Hansard Parliamentary Debates Corpus (HHPDC). Based on the complete records of 19th-century Hansard parliamentary debates, this sub-corpus represents the discourse of British political circles. The data was gathered using stratified chronological sampling, extracting 10,000

sentences per decade to create a balanced sample.

(2) Queen Victoria’s Letters Corpus (QVLC). This sub-corpus includes the official and private correspondence of Queen Victoria from 1837 to 1861. As the central monarch of 19th-century Britain, her letters reflect how the royal apex perceived the empire, providing primary documents for understanding the expression of imperialism at the highest echelons of British power.

(3) Project Gutenberg Nineteenth-Century Essays Corpus (PGEC). Relying on the text resources of Project Gutenberg, this sub-corpus systematically collects essay collections from various fields—such as politics, education, and science—in 19th-century Britain. These texts represent the public discussions of contemporary British scholars, thinkers, and social elites, allowing for a comprehensive examination of the term’s evolution within the realm of social thought.

## 2.2 Research Methods and Procedures

This study adopts a mixed-methods approach, utilizing the Lancsbox software as the primary analytical tool, and is conducted in two core steps.

The first step is keyword collocation analysis. Using the GraphColl function in Lancsbox, we queried the word “empire” and its inflected forms across all time periods within the entire NBDEC corpus. The filtering criteria for collocates were set to a collocation frequency of  $\geq 1$  and a Log Dice value of  $\geq 10$ . By sorting the results in descending order based on the Log Dice values, we identified keywords significantly associated with “empire” for each period. The top three unique keywords in each period were recorded, and any overlapping terms were replaced by subsequent ranking words.

The second step involves the calculation and analysis of semantic prosody. Employing the KWIC (Key Word In Context) function in Lancsbox, the analysis was restricted to the HHPDC sub-corpus. With “empire” and its variants as node words, the span was set to 10 words on both the left and right sides, ultimately yielding 624 valid concordance lines. Manual sentiment polarity annotation was then performed on all concordance lines. The annotations were categorized into three types: positive (+1), negative (-1), and neutral (0), as detailed in Table 1:

Table 1. Polarity Judgment Criteria

<b>Polarity Label</b>	<b>Judgment Criteria</b>
Positive (+1)	The speaker expresses an attitude of praise, support, defense, or affirmation towards the empire, articulating its positive values, developmental advantages, or legitimacy.
Negative (-1)	The speaker expresses an attitude of criticism, concern, denial, or skepticism towards the empire, articulating its risks, crises, illegitimacy, or negative impacts.
Neutral (0)	Objective factual statements about the empire or policy explanations without clear emotional tendencies; no positive or negative attitudes are expressed.

Following the annotation, a standardized procedure for calculating semantic prosody was established. This ensures the rigor and reproducibility of the analytical results by eliminating the impact of uneven sample sizes across different periods. The procedure consists of four specific steps:

(1) Basic data compilation. Grouping the 624 annotated concordance lines into decades, we counted the number of positive, negative, and neutral lines within each period. These are recorded as positive frequency ( $F_p$ ), negative frequency ( $F_n$ ), and neutral frequency ( $F_z$ ), respectively. The total frequency for a single period is calculated as  $F_t = F_p + F_n + F_z$ .

(2) Relative proportion calculation. The relative proportions of the three polarities within a single period were calculated as follows: positive proportion ( $R_p$ ) =  $F_p/F_t$ , negative proportion ( $R_n$ ) =  $F_n/F_t$ , and neutral proportion ( $R_z$ ) =  $F_z/F_t$ . Utilizing relative proportions mitigates the interference caused by varying sample sizes across periods, thereby accurately reflecting the distribution of emotional tendencies in the discourse of a specific timeframe.

(3) Semantic prosody intensity (SI) calculation. It is calculated using the formula  $SI = R_p - R_n$ , ranging from [-1, 1]. A positive value indicates overall positive semantic prosody, while a negative value indicates a negative one; the larger the absolute value, the stronger the semantic prosody intensity. This study established a graded judgment standard:  $SI > 0.3$  indicates strong positive semantic prosody;  $0 < SI \leq 0.3$  weak positive;  $-0.3 \leq SI < 0$  weak negative;  $SI < -0.3$  strong negative; and SI approaching 0 indicates neutral semantic prosody<sup>[1][2]</sup>.

(4) Chronological trend analysis. By arranging the semantic prosody intensity values of each period in chronological order, we outlined the changing trends in the semantic prosody of the word “empire” within 19th-century British parliamentary discourse. Combined with major historical events of the corresponding periods, we further analyzed the core driving factors behind these prosodic fluctuations.

### 3. Empirical Analysis

#### 3.1 The Diachronic Evolution of Keywords in the Concept of “Empire”

A periodized keyword analysis of the NBDEC corpus shows that the collocates of “empire” underwent marked changes over the course of the nineteenth century. Table 2 lists the three collocates with the highest Log Dice scores for each period.

Table 2. Keywords Collocating with “Empire” from the 1800s to the 1900s

Decade	Collocate	Distribution	Freq. (collocation)	Freq. (subcorpus)	Log Dice	MI
1800s	british	-1	9	39	11.2	9.7
	united	-0.67	6	35	10.7	9.3
	part	-1	14	371	10	7.1
1810s	navies	-1	2	3	10.5	12.2
	recruited	1	2	4	10.5	11.8
	resource	-1	2	5	10.5	11.5
1820s	commoner	-1	1	1	11.2	14.5
	prohibits	1	1	1	11.2	14.5
	jarring	-1	1	1	11.2	14.5
1830s	integrity	-1	6	32	11.4	10.5
	ottoman	-1	3	14	10.8	10.7
	turkish	-1	2	2	10.6	12.9
1840s	indian	-1	3	8	11.4	12
	entangling	-1	1	1	10.1	13.4
	alliances	1	1	1	10.1	13.4
1850s	colonial	-1	10	157	10.6	8.4
	unity	-1	2	2	10.1	12.4
	stability	-1	2	3	10.1	11.8
1860s	nice	-1	5	44	10.9	9.9
	french	-1	9	141	10.7	9.1
	vast	-0.33	3	30	10.5	9.7
1870s	dismemberment	-1	6	6	11.6	12.4
	dismember	-1	3	3	10.7	12.4
	fabric	-1	3	6	10.6	11.4
1880s	disintegration	-1	7	10	10.9	10.8
	our	-0.64	22	422	10.4	7
	safety	-1	4	13	10	9.6
1890s	portion	-1	3	52	10.2	8.9
	might	-1	1	1	10	13.1
	enjoying	1	1	1	10	13.1

Decade	Collocate	Distribution	Freq. (collocation)	Freq. (subcorpus)	Log Dice	MI
1900s	expansion	-1	6	16	10.5	9.5
	extension	-1	6	17	10.4	9.4
	danger	-0.6	5	33	10	8.2

Using the significant collocates extracted through GraphColl and analyzing them in thirty-year intervals, this study finds that the connotations of the concept of “empire” in 19<sup>th</sup>-century Britain evolved in clear stages that were closely synchronized with broader historical developments. As the social context shifted, the core referent of the term also changed in distinct ways.

From 1800 to 1830, the core collocates uniquely associated with “empire” were “united”, “recruited”, and “navies”. In this period, the concept of empire was centered on survival under wartime conditions. Between 1800 and 1815, Britain remained under the military pressure of the Napoleonic Wars, while the threat of French invasion across the Channel challenged the connection between the British mainland and its dependent territories. The war lasted from 1793 to 1815, interrupted only briefly in 1802, and its total cost reached £1.657 billion, nearly three times the combined cost of all major wars Britain had fought since the Glorious Revolution<sup>[3]</sup>. In parliamentary discourse, “empire” was consistently tied to appeals for united resistance against foreign enemies. The Act of Union of 1801 formally established the United Kingdom of Great Britain and Ireland, marking the first fusion of Britain’s national identity with its imperial identity. At this stage, the concept of “empire” had not yet acquired a stable meaning centered on overseas colonial expansion.

From 1831 to 1860, the core collocates associated with “empire” were “Indian”, “colonial”, and “integrity”. The concept of “empire” thus shifted from defense toward the construction of an overseas colonial system. After the Napoleonic Wars, Britain established global maritime supremacy, and colonial expansion became the central trend of national development, with India gradually emerging as the core of the British imperial system. Parliamentary debates in 1840 contain extensive discussion of the governance, trade, and security of British India, making India the most important geographical embodiment of the concept of “empire.” After the Indian Rebellion of 1857, expressions concerning the integrity of the empire became especially frequent in parliamentary discourse, reflecting a major concern with maintaining stable rule over colonial possessions<sup>[4]</sup>. At the same time, debates on the colonial empire extended to multiple settlements, including Canada and Australia. Both the geographical scope and the governance implications of the concept of “empire” were substantially broadened, as it evolved from a political concept rooted in union at home into one denoting a global colonial system.

From 1861 to 1890, the core collocates uniquely associated with “empire” were “French”, “dismemberment”, and “safety”. During this period, the concept of “empire” came to focus on consolidating the colonial system and strengthening internal connections within it. Between the 1860s and 1870s, the European order underwent restructuring. The Franco-Italian War of 1859–1861 disrupted the balance of power, while the Franco-Prussian War of 1870 greatly weakened France. Russia then seized the opportunity to renounce the restrictions imposed by the 1856 Treaty of Paris on its role in relation to the Ottoman Empire, intensifying the crisis of Ottoman dismemberment. British imperial strategy accordingly shifted toward preserving the Ottoman Empire in order to safeguard the security of Near Eastern routes<sup>[5]</sup>. From 1880 to 1900, the European powers entered the phase of “New Imperialism”, and the Berlin Conference of 1884–1885 triggered the scramble for Africa. By working within the conference’s framework of free trade, regulated competition, and allied support, Britain consolidated its existing interests in Africa, avoided direct military confrontation with rival powers, and secured market access for its industrial goods and capital. In this way, it maintained its commercial and strategic advantages in Africa at relatively low cost<sup>[6]</sup>. Britain subsequently occupied Egypt in 1882<sup>[7]</sup>, and invaded Sudan in 1886<sup>[8]</sup>,

and its colonial expansion reached a peak. Yet the acceleration of expansion also brought mounting resistance in the colonies. Parliamentary discourse increasingly referred to the risk of imperial “dismemberment.” At its core, this phase of British colonial expansion was driven by the monopolistic needs of industrial capital. Military intervention in Egypt and Sudan sharply raised the costs of imperial rule: “implies militarism now and ruinous wars in the near future... The annexation of tropical and sub-tropical territories, and the attempt to govern ‘lower races,’ that is driving us down the steep road to militarism<sup>[9]</sup>.”

From 1891 to 1900, the core collocates of “empire” became sharply divided. Words such as “enjoying” and “expansion” appeared with high frequency alongside terms like “danger”. In 1900, the Boer War had reached a stalemate. Britain had initially deployed only 75,000 troops, but was ultimately forced to mobilize 450,000 men, including reinforcements from the dominions and colonies, far exceeding prewar expectations. This exposed serious deficiencies in military preparedness, while Britain’s international isolation pushed its expansionist project into crisis<sup>[10]</sup>. The outbreak of the Boer War triggered a comprehensive domestic debate over imperial policy, and the meaning of “empire” became increasingly polarized. On the one hand, there were discourses defending the maintenance of Britain’s global imperial supremacy; on the other, there also were sharp criticisms of reckless expansion and growing anxieties about the empire’s future. This division reflects the fractured nature of British perceptions of empire at the end of the century.

### 3.2 The Changing Semantic Prosody of “Empire” Discourse

#### 3.2.1 Statistical Results

Using the KWIC function in LancsBox and following the statistical and computational procedures outlined above, the following results were obtained. The statistical results of semantic prosody are shown in Table 3, and the statistical chart of the diachronic changes of semantic prosody is presented in Figure 1.

Table 3. Semantic Prosody Data, 1800s–1900s

Decade	Total number of index entries	Fp	Fz	Fn	Rp	Rz	Rn	SI
1800s	84	24	38	22	0.2857	0.4524	0.2619	0.024
1810s	41	15	16	10	0.3659	0.3902	0.2439	0.122
1820s	13	3	7	3	0.2308	0.5385	0.2308	0
1830s	40	11	20	9	0.275	0.5	0.225	0.05
1840s	28	8	12	8	0.2857	0.4286	0.2857	0
1850s	56	13	27	16	0.2321	0.4821	0.2857	-0.054
1860s	40	10	22	8	0.25	0.55	0.2	0.05
1870s	57	19	25	13	0.3333	0.4386	0.2281	0.105
1880s	111	23	60	28	0.2072	0.5405	0.2523	-0.045
1890s	30	5	16	9	0.1667	0.5333	0.3	-0.133
1900s	124	21	57	46	0.1694	0.4597	0.371	-0.202

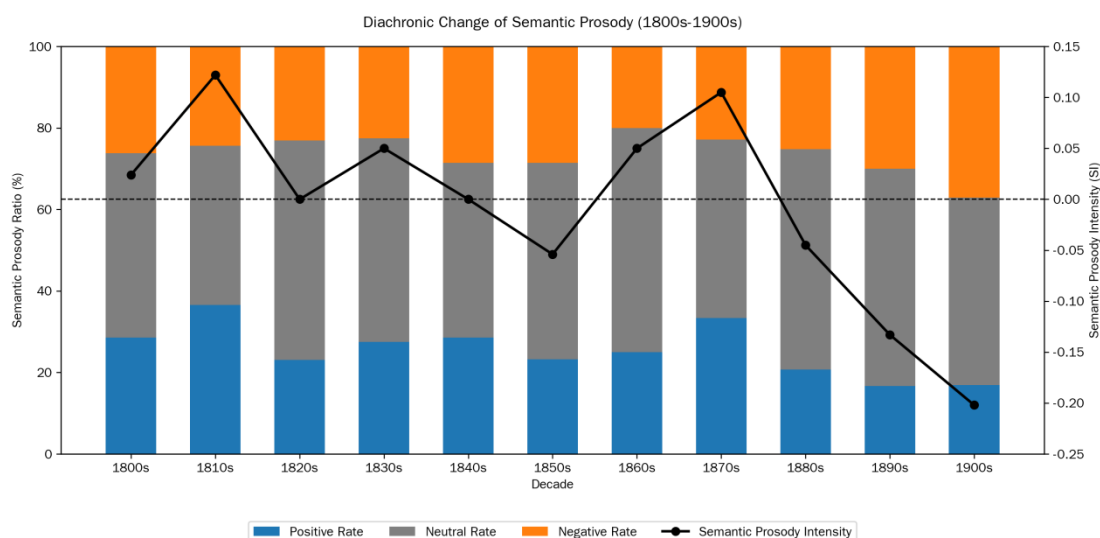


Figure 1. Semantic Prosody by Decade, 1800s–1900s

### 3.2.2 Temporal Changes in Semantic Prosody and Their Causes

Viewed diachronically, the semantic prosody of the target word in 19th-century British parliamentary debates presents three distinct phases, with its evolution closely corresponding to Britain’s contemporary political and social development.

From the 1800s to the 1840s, it maintained a weak positive balance, with  $SI \geq 0$  and no negative semantic prosody. The 1810s recorded the full-period positive peak, while the 1820s and 1840s had an SI of 0, though their small sample sizes limit the representativeness of findings. From the 1850s to the 1880s, it broke the long-term non-negative balance and entered a transitional phase of alternating positive and negative fluctuations. The 1850s saw the first negative SI (-0.054) in the observation period, the 1870s rebounded to the second-highest positive SI (0.105) of the whole period, and the 1880s returned to negative territory. Consistently high neutral prosody indicated the discourse core remained unchanged, with fluctuations driven by shifts in the issue of parliamentary debates. From the 1880s to the 1900s, negative semantic prosody intensified continuously, with SI falling to a full-period low of -0.202 in the 1900s, and the negative instance share surpassing the positive share by more than twofold for the first time.

This trajectory aligns with early-century parliamentary reform, mid-century industrial conflict and colonial disputes, and late-century global hegemony competition and domestic political tensions.

## 4. Conclusion

Drawing on a self-compiled corpus of nineteenth-century British imperial discourse and using methods from corpus linguistics, this study examines the changing connotations of the concept of “empire”. The findings show that the meaning of “empire” developed in clearly identifiable stages in response to shifts in historical context. In the early nineteenth century, the concept centered on military defense under wartime conditions. By the mid-nineteenth century, its focus had shifted to the construction of an overseas colonial system and the preservation of territorial integrity. In the late nineteenth century, it came to center on the control of colonial frontiers. By the end of the century, however, the concept had become divided, as discourses of maintaining hegemony came into conflict with critiques of expansion.

The changing semantic prosody of the word “empire” in 19<sup>th</sup>-century British parliamentary discourse closely corresponds to historical environment, too. Its semantic prosody moved from strongly positive in the early nineteenth century to weakly positive and fluctuating in the middle decades, before turning weakly negative by the end of the century. This shift directly reflects broader changes in British society’s understanding of empire.

The quantitative methods of corpus linguistics provide a verifiable approach to reinterpreting historical concepts. Quantitative analysis based on large-scale authentic texts can help overcome the subjective limitations of purely qualitative interpretation in traditional historical research, and can offer a more faithful account of the meaning and evolution of historical concepts in their specific contexts.

## References

- [1] Partington, Alan. “Utterly content in each other’s company: Semantic prosody and semantic preference.” *International journal of corpus linguistics* 9.1 (2004): 131-156.
- [2] Xiao, Richard, and Tony McEnery. “Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective.” *Applied linguistics* 27.1 (2006): 103-129.
- [3] Colley, Linda. *Britons: forging the nation, 1707-1837*. Yale University Press, 2003, p150.
- [4] Srivastava, Pramod Kumar. “Nationalism Imagined? hidden Impacts of the Uprising of 1857.” *South Asia Research* 38.3 (2018): 229-246.
- [5] Şakul, Kahraman. “From Mare Clausum to Mare Liberum: Black Sea Diplomacy in the Era of Russo-Ottoman Duopoly.” *Kritika: Explorations in Russian and Eurasian History* 21.4 (2020): 701-732.
- [6] Munene, G. Macharia. “The United States and the Berlin conference on the partition of Africa, 1884–1885.” *Transafrican Journal of History* (1990): 73-79.
- [7] Hopkins, Anthony G. “The Victorians and Africa: a reconsideration of the occupation of Egypt, 1882.” *The Journal of African History* 27.2 (1986): 363-391.
- [8] Sanderson, George N. “The European Powers and the Sudan in the Later Nineteenth Century.” *Sudan Notes and Records* 40 (1959): 79-100.
- [9] Hobson, John Atkinson. *Imperialism: A study*. Routledge, 2018, p137.
- [10] Morgan, Kenneth O. “The Boer War and the media (1899–1902).” *Twentieth Century British History* 13.1 (2002): 1-16.