



# A Corpus-Assisted Investigation into Joe Biden's Nuclear Policy Discourse: Evidence from Congressional Hearings

Sun Guang

National University of Defense Technology, Changsha, China

Email: [youngsong24@foxmail.com](mailto:youngsong24@foxmail.com)

**Abstract:** This paper takes Biden's nuclear-related speeches at U.S. Congressional hearings during his tenure as a senator as the corpus, and uses corpus tools such as AntConc and LancesBox to analyze the subject words. With the mutual information value as a reference, it obtains the high-frequency collocations of words, identifies the main topics of concern in Biden's speeches, and analyzes the characteristics of Biden's speeches. It is found that Biden pays more attention to five aspects: major nuclear-related countries, nuclear weapons development, nuclear arms control treaties, international organizations, especially the North Atlantic Treaty Organization, and diplomatic relations. This study also finds that Biden's speeches exhibited three distinct characteristics: a strong emphasis on the development of nuclear weapons capabilities, significant attention to nuclear arms control, and close focus on "Countries of Concern". Biden's rhetoric reflected a dual strategy of deterrence and containment. While reaffirming the United States' commitment to bolstering its nuclear deterrence capabilities, his statements simultaneously revealed efforts to undermine the nuclear arsenals of other nations and curb their nuclear development. This study can provide a foundation for comparative research with President Trump.

**Keywords:** Corpus; Joe Biden; Nuclear Policy Discourse; Hearing

## Introduction

On the international political stage, the discourse of the U.S. President has a pivotal influence. As political leaders, their words not only reflect personal political stances, but also often foreshadow the direction of their national policy and changes in international relations. Especially, the statements of the U.S. President on the sensitive and critical issue of nuclear weapons affect the nerves of the world. Therefore, in-depth research on the discourse of the U.S. President, especially the expression of nuclear-related issues, is of important academic value and practical significance for understanding its foreign policy and predicting the development of the international situation.

Joe Biden, an experienced politician, has a 36-year career in the Federal Senate, and his speaking records during his tenure as a member of Congress provide a rich corpus. This study aims to reveal his political ideas and strategic intentions by analyzing Biden's nuclear-related discourse at hearings, and then provide a new perspective for understanding the evolution of U.S. nuclear policy.

This study is based on the corpus-based analysis method, combined with technical means such as subject word analysis, collocation analysis and retrieval line analysis, to deeply explore Biden's discourse. By constructing a self-checking corpus, this study can systematically collect and organize Biden's speeches at hearings, ensuring the comprehensiveness and accuracy of the analysis. In addition, through the analysis of subject words and collocations, this study can reveal the core concepts and semantic network in Biden's discourse, so as to more accurately grasp his discourse strategies and intentions. Through retrieval line analysis, this study can track the frequency and context of specific words in Biden's discourse, and further explore the evolution and strategic adjustment of his discourse.

As noted in recent scholarship on regional and country research, political discourse analysis requires integration of corpus linguistics and cognitive linguistics to uncover hidden ideological structures. This study responds to this call by constructing a specialized corpus of hearing data, enabling identification of American nuclear policies. This methodological innovation provides empirical linguistic evidence for how nuclear policy is discursively constructed as necessary.

The Biden administration's nuclear policy is marked by contradictory signals—proposing "sole purpose" reforms during the campaign while later expanding low-yield nuclear weapons and lowering use thresholds. Analyzing hearing discourses can decode how the administration redefines nuclear use conditions without explicit policy revisions.

The significance of this study lies not only in providing a new perspective and method for the academic community to study the discourse of the U.S. President, but also in its contribution to understanding the current international nuclear policy and promoting international peace and stability. Through in-depth analysis of Biden's nuclear-related discourse, this study hopes to provide valuable reference and inspiration for policy makers, international relations scholars and the general public.



## Literature Review

The study of political leaders' discourse began to rise under the influence of postmodernism in the 1970s, and gradually became the focus of academic attention with the development of globalization and international politics. The research on Biden's discourse has also shown a surge after he took office.

Academic discourse currently centers on Biden's victory speech, inaugural address, and foreign policy statements, predominantly employing Critical Discourse Analysis (CDA) to uncover rhetorical strategies such as "unity" and "hope"<sup>[4]</sup>. Conversely, the institutional discourse from his 36-year congressional career, particularly testimony delivered during hearings, has received only sporadic attention<sup>[9]</sup>, with no studies to date establishing dedicated corpora for systematic examination. The most attention is naturally paid to Biden's campaign speeches and inaugural addresses. Siregar analyzes Biden's elected president speech, focusing on its discourse structure and underlying ideologies, which serve to legitimize his leadership and policy directions<sup>[6]</sup>. In the context of specific speeches, such as Biden's inauguration address, Pramadya et. al. explore how Biden employs political language to project hope and policy commitments, with each message reflecting underlying ideological values and power relations<sup>[5]</sup>. Nisreen N. Al-Khawaldeh et al. analyzed the language strategies in Biden's 2021 inaugural address, such as creativity, metaphor, contrast, indirectness, citation, and intertextuality<sup>[1]</sup>.

Corpus methods have been widely applied in analyzing political discourse, but their use in nuclear policy research remains limited. Existing corpus studies mostly focus on electoral discourse or international relations rhetoric, with few targeting the specialized field of nuclear policy. This gap is notable given that corpus tools can quantify key concept frequencies and reveal discursive dominance relationships. He<sup>[4]</sup> uses corpus methods to investigate the president Biden's inaugural speech and finds that the main theme is unity, and the main language skills used in the discourse include personal pronouns, modal verbs, passivity, nominalization, declarative sentences, implicit messages and rhetorical devices. Although nuclear arms control and non-proliferation remain core issues in international security research, existing discourse on Biden's policies continues to focus on topics such as the Middle East, Ukraine, and COVID-19. Discussions specifically addressing "nuclear policy" discourse remain virtually non-existent. El-Zaghal analyzes Biden's speech on the Middle East, reflecting American officials' reactions to the Israeli-Palestinian conflict and indicating tendencies toward conflict resolution and diplomatic positioning<sup>[3]</sup>. Tymoshchuk investigates the use of euphemisms in Biden's discourse during the Ukraine war, highlighting how linguistic choices serve to frame the conflict in a manner that influences public perception and political legitimacy<sup>[7]</sup>. However, few studies focus on nuclear policy discourse.

There are many studies on Biden's speeches as president, and not a few studies focus on him as a representative of U.S. government policy, but there is still a gap in the field of focusing on his speeches as a member of Congress. It can be said that the number of studies focusing solely on Biden's identity as a legislator is also not considerable. Zhao pointed out that related research mostly focuses on congressional decision-making and rarely delves into the perspective of image perception<sup>[10]</sup>. Therefore, this study focuses on Biden's speeches in Congress as a legislator.

## Data and Method Selection

All corpora in this study are derived from <https://www.govinfo.gov/>, an official legal and regulatory retrieval website directed by the U.S. Government Publishing Office (GPO). The core rationale for selecting this platform lies in the authority and originality of its data. As the official publication platform for the U.S. federal government, it hosts congressional hearing records, federal legal texts, and other materials that are first-hand and unaltered by third parties. This ensures the authenticity of the political and legal contexts within the corpus and mitigates data biases from non-official sources.

Under the "Browse by Document Type" section on govinfo.gov, "Congressional Hearings" were explicitly selected as the target document type. "Biden" (name) and "nuclear" (core topic) were used as content keywords, combined with the "AND" logical relationship for retrieval. This excluded irrelevant texts mentioning only a single keyword. The initial retrieval time window was set to cover Biden's full term as a federal senator (1973-2009). However, feedback from the platform's retrieval results revealed a systemic absence of congressional hearing data from 1982 to 1996. Consequently, the time frame was ultimately limited to Biden's 5th and 6th terms as a senator (1997-2008). The 5th term corresponds to 1997-2003 (105th-107th Congresses), and the 6th term to 2003-2008 (108th-110th Congresses).

A total of 570 congressional hearing records related to "Biden" and "nuclear" were obtained under the aforementioned retrieval criteria. Each document was manually verified to retain only those containing Biden's actual remarks. The screening criteria included: (1) Explicit labels in the main text such as "Statement by Senator Joe Biden" or "Remarks by Sen. Biden" indicating speech content; (2) Direct quotation paragraphs in the dialogue records starting with identifiers like "Sen. Biden:" or "Mr. BIDEN.". Ultimately, 144 documents meeting the criteria were selected, excluding texts that only mentioned Biden's name without including his remarks.

First, the 144 documents were converted from the original PDF format to editable text encoded in UTF-8 (txt format). Subsequently, based on the formatting conventions of hearing records, the start and end boundaries of Biden's speeches were marked. The start points were identified by exclusive identifiers such as "Sen. Biden:" or "Mr. BIDEN.", and the end points by identifiers of the next participant such as "Sen. Kerry:", or procedural markers like "[End of Statement]".

Finally, Biden's speeches were extracted separately from each text to create a self-built corpus, resulting in a corpus with a total of 592728 tokens. With the help of AntConc corpus software, online word cloud generator, LancsBox etc., themes and collocations are analyzed to find out the main topics that Biden focused on in his speeches at hearings.

Keyword Analysis

Keywords are not the absolute frequency of words, but rather words that appear with particularly high frequency when comparing the target corpus with a reference corpus. These words can reflect the thematic content of the target corpus. Summarizing and analyzing a certain number of thematic words cannot only lead to an overall understanding of the content of the discourse, but also reveal its underlying socio-cultural connotations<sup>[8]</sup>. The reference corpus must be representative and comprehensive. The British National Corpus (BNC) contains approximately 100 million words of modern British English from the late 20th century. Therefore, this study uses the British National Corpus as a reference corpus. After comparing the two corpora, the top 100 thematic words were selected, and it was found that many words were pronouns and function words such as "I," "we," "you," and "that." This may be related to the fact that the corpus text comes from spoken hearings. In addition, given that a word may not fully express its meaning (such as Korea, which does not indicate whether it is South Korea or North Korea), the text was pasted into the TerMine terminology extraction tool to obtain a list of terms, and the top 100 were selected. After merging singular and plural forms and duplicate words, a total of 82 thematic words (groups) were obtained, which can be roughly divided into four categories: country/region, weapon/project, treaty/cooperation, personnel/organization.

Country/Region



Figure 1 Country/Region Category with “Iraq” the most

Figure 1 shows that the most frequently occurring country/region word is "Iraq". This indicates that the United States Congress, especially Biden, is very concerned about the issue of the Iraq region. The word "Iraq" appeared 1203 times, averaging more than eight times per text. To this end, the frequency of the word "Iraq" was retrieved by year, resulting in Table 1.

Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Frequency	6	34	23	40	21	117	388	233	202	106	33	0

Table 1 Frequency of "Iraq" by Year

Table 1 shows that after 2001, the frequency of "Iraq" increased sharply, reaching a peak in 2003, and then gradually decreased. This is because, in 2001, the United States suffered the "9/11" terrorist attacks, and U.S. President Bush declared a war on terrorism, listing Iraq and other countries as part of the “Axis of Evil”. On March 20, 2003, a joint force led by British and American troops bypassed the UN Security Council and unilaterally launched a military strike against Iraq on the grounds that Iraq possessed weapons of mass destruction and secretly supported terrorists. Until August 2010, U.S. combat troops began to withdraw from Iraq. This may be a major reason why Biden has long been concerned about Iraq. In addition, since taking office, Biden has made several slips of the tongue in public, mistakenly referring to “Ukraine” as “Iraq”, which may be related to Biden's many years of work related to Iraq. Subsequently, the most frequently occurring countries are the United States (669 times), Russia (572 times), North Korea (291 times), China (276 times), Afghanistan (197 times), and the Middle East (113 times). The mention of the United States in the hearings of the United States Congress is self-evident. Russia is a traditional nuclear power and has the most dealings with the United States in terms of nuclear weapons. North Korea is also listed as an Axis of Evil and is also called a “rogue state”. The United States claims that they violate humanitarian spirit, violate internationally recognized human rights, and attempt to acquire nuclear weapons and long-range missiles to achieve more rogue terrorist purposes. China is also one of the “countries of concern” defined by the United States and is one of the few countries in the world with nuclear weapons, naturally becoming one of the focuses of discussion for Biden. In 2001, a coalition led by the United States launched the Afghanistan War against al-Qaeda and the Taliban in Afghanistan on October 7, which occurred during this time period. The United States always associates Afghanistan with terrorism, and a major core issue for the United States during this time period was counter-terrorism, so Afghanistan also became a country they often mentioned.

Weapon/Project





7	power	26	6.9	22	proliferation	14	7.2
8	capacity	24	7.2	23	tests	12	8.8
9	iran	30	6.3	24	material	13	7.2
10	north	28	6.3	25	programs	15	6.3
11	testing	19	8.3	26	Korea	16	6.0
12	test	19	7.4	27	reaction	11	8.4
13	nonproliferation	17	7.5	28	device	11	8.4
14	their	43	5.2	29	arms	13	6.4
15	treaty	27	5.6	30	materials	10	8.7

Table 2 Collocations with “Nuclear”

According to the collocation relationship shown by the Graph Coll of the LancsBox tool, the closer the distance to the node word, the stronger the collocation relationship between the two and the higher the statistical significance; the higher the frequency of collocation with the node word, the larger the point where the collocation word is located. The color of the collocation word indicates the frequency of the word in the entire corpus. The following figure is a collocation graph of the word “nuclear”.

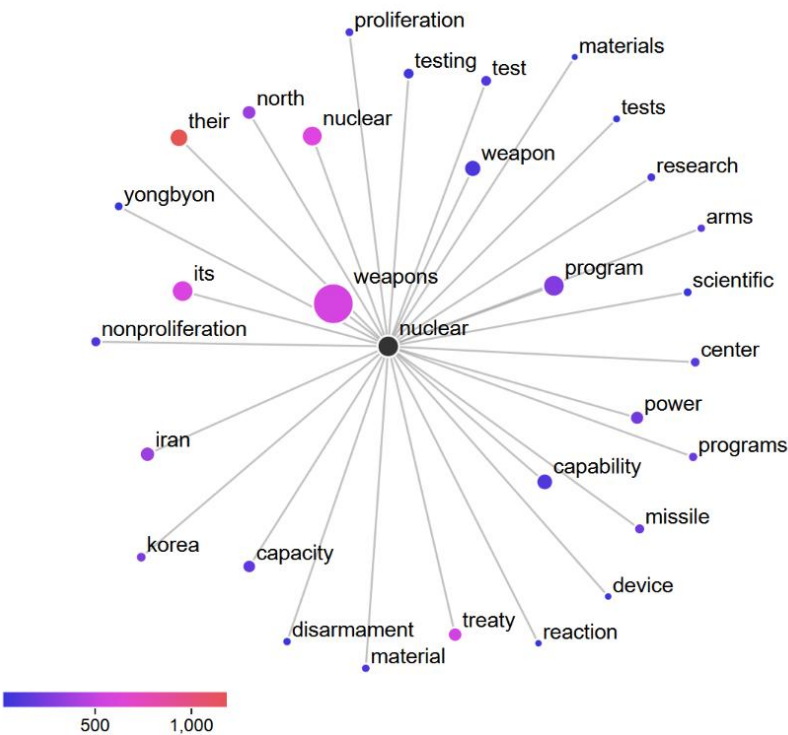


Figure 5 Collocation Graph of “Nuclear”

As shown in Figure 5, taking MI3 as the reference, the high-frequency collocation words of “nuclear” are obtained, showing the semantic field strongly related to “nuclear”. At the same time, we find that the top 20 high-frequency collocation words with “nuclear” reflect several characteristics of Biden’s nuclear policy discourse. First, it emphasizes the development of nuclear weapons capabilities. The top eight highly relevant collocations are related to nuclear weapons and nuclear forces. Words such as nuclear weapon, nuclear capability, nuclear power, and nuclear capacity appear frequently, which undoubtedly reflects the United States’ continued commitment to the modernization of its nuclear forces. The United States has always regarded nuclear weapons as an important means of maintaining its national interests and world peace. After the end of the Cold War, despite the reduced nuclear threat, the United States still attaches importance to maintaining and developing its nuclear deterrent to ensure its superpower status and the needs of its hegemonic politics. The United States focuses on the modernization of nuclear weapons and their delivery systems (nuclear device) to maintain the effectiveness and advancement of its nuclear arsenal. This includes investing in the research and development of new nuclear weapons and the maintenance of existing nuclear weapons, ensuring the safety and reliability of nuclear weapons. The United States also conducted some demonstrations and tests of its nuclear forces during this period (nuclear tests), including test-firing intercontinental ballistic missiles and conducting subcritical nuclear tests, to demonstrate its nuclear capabilities and collect important data on nuclear weapon materials (nuclear material). Furthermore, it emphasizes nuclear arms control. The figure 6 shows that a category of collocations directly related to the Nuclear Non-Proliferation Treaty and nuclear disarmament are also included. On the one hand, this is highly related to Biden’s own identity and role. Biden is an expert in nuclear arms control and has been engaged in nuclear arms control work for a long time. On the other hand, it is related to the political considerations of the United States. The United States

actively promotes international nuclear non-proliferation policies, striving to prevent other countries, especially those considered potential threats such as Iraq, Iran, and North Korea, from developing nuclear weapons. The United States uses diplomatic and military means to try to weaken the nuclear forces of other countries and maintain the ability to retaliate against these countries. The United States wants to maintain its own nuclear arsenal strength and prevent other countries from possessing nuclear arsenals that could threaten the United States. Therefore, the United States is committed to seeking a leading position in nuclear arms control, attempting to dominate the discourse, prevent non-nuclear countries from acquiring nuclear weapons, and reduce the nuclear arsenals of nuclear-weapon states.

Besides, it pays close attention to “Countries of Concern”. Countries of Concern are defined as any foreign government that has long engaged in activities that seriously endanger the national security or the safety of Americans, specifically China, Russia, Iran, North Korea, etc. Highly relevant collocations not only directly mention countries such as Iran and North Korea, but also highlight an institution: Yongbyon Nuclear Scientific Research Center. This center is North Korea’s main nuclear research center. It houses North Korea’s only operating 5-megawatt nuclear reactor. According to the “Geneva Agreed Framework”, North Korea closed the research center, and the United States provided North Korea with 500,000 tons of heavy oil annually. After the outbreak of the Second North Korean Nuclear Crisis at the end of 2002, the United States stopped providing heavy oil to North Korea. North Korea then restarted the frozen Yongbyon nuclear facilities. The United States regards Iran and North Korea as key factors in regional security. Iran's influence in the Middle East is increasing, while North Korea's nuclear weapons program directly threatens the security situation in Northeast Asia. As a maintainer of the international order, the United States is constantly concerned about any behavior that may undermine the international system in order to maintain its hegemonic position.

## Conclusion

Through the study of Biden’s nuclear-related speeches at hearings during his tenure as a member of Congress from 1997 to 2008, it is found that Biden paid more attention to five aspects: major nuclear countries, nuclear weapons development, nuclear arms control treaties, international organizations, especially the North Atlantic Treaty Organization, and diplomatic relations. Biden’s speech reflects three characteristics: first, emphasizing the development of nuclear weapons capabilities; second, attaching importance to nuclear arms control; and third, paying close attention to “Countries of Concern”. Biden's rhetoric reflects a dual strategy of deterrence and restraint. On the one hand, the United States attaches importance to the maintenance and development of nuclear deterrence, and on the other hand, it attempts to weaken the nuclear forces of other countries and prevent other countries from developing nuclear weapons, demonstrating the hegemonic mentality of the United States as a global superpower.

This discovery not only fills a gap in corpus-based research on Biden's nuclear discourse as a senator but also holds broader applications. This study provides a replicable research template for subsequent investigations. Researchers can apply the same framework to examine Biden's nuclear discourse across different time periods, tracing its historical evolution. Alternatively, the analysis could be extended to Republican senators or even President Trump, with comparative corpora suggested to validate partisan differences.

Finally, it should be pointed out that the corpus of this study all comes from the testimonies of the US Congressional hearings. Although the data is highly authentic and reliable, the selected corpus only involves Biden’s fifth and sixth terms as a member of Congress, corresponding to the time range of 1996-2002 and 2003-2008, and there are certain limitations in the time span, which is not enough to cover Biden’s entire career as a member of Congress, and the content of the corpus is not detailed enough. Therefore, the research conclusions on Biden’s discourse characteristics and positions are not very comprehensive.

## REFERENCES

- [1]. Al-Khawaldeh, N. N., et al. (2023). The art of rhetoric: persuasive strategies in Biden’s inauguration speech: a critical discourse analysis. *Humanities and Social Sciences Communications*, 10(1).
- [2]. Baker P. (2006). *Using Corpora in Discourse Analysis*. New York: Routledge, 95-96.
- [3]. El-Zaghal, Fatma. (2021). Fields of Text Studies and Discourse Analysis: A Study in light of applying Dell Hymes Speaking Model to Joe Biden's Speech on the Middle East (2021). *International Journal of Humanities and Language Research*, 4(1), 1-39.
- [4]. Jinyu He. (2022). A Corpus-Based Critical Discourse Analysis of President Biden’s Inaugural Speech. *Modern Linguistics*, 10(12), 2965-2972.
- [5]. Pramadya, Teguh & Rahmanhadi, Anna. (2021). A day of history and hope: A critical discourse analysis of Joe Biden’s Inauguration speech. *Rainbow: Journal of Literature, Linguistics and Culture Studies*, 10(2), 1-10.
- [6]. T. M. Siregar. (2021). The critical discourse analysis on Joe Biden’s elected president speech. *Journal Of Applied Studies in Language*, 5(1), 79-86.
- [7]. Tymoshchuk, Nataliia. (2023). EUPHEMISMS IN JOSEPH BIDEN’S MODERN POLITICAL DISCOURSE IN THE WAR IN UKRAINE. *Alfred Nobel University Journal of Philology*, 2(26/2), 210-224.
- [8]. Xu Jiajin. (2019). *Corpus and Discourse Research*. Beijing: Foreign Language Teaching and Research Press, 64.
- [9]. Zhang Yuanyuan, Zhang Xinlu. A Corpus-Based Critical Discourse Analysis on China in the RIA Novosti During the COVID-19 Pandemic [J]. *Modern Linguistics*, 2021, 9(5): 1199-1206.
- [10]. Zhao Guojun. (2007). *A Study on the Views of Chinese by American Congressmen (1989-2006)*. Fudan University.