Words That Will Shake the World:

Investigating the Interaction between US and Chinese Foreign Policy Discourse with Sentiment Analysis and Granger Causality Tests

Alexander Lin

Advisor: Professor Lisa Singh

A Thesis Submitted in Partial Fulfillment of the Requirements for the Award of Honors in International Politics, Edmund A. Walsh School of Foreign Service, Georgetown University, Spring 2023.

Table of Contents

Abstract **3**
Chilly Summit in Alaska: US-China Relations at a Perilous Crossroads **4**
Literature Review **5**
Research Question **10**
1. Finding the Provocateur: Drilling Down on the Core Hypothesis 10
2. Clarifying Key Concepts 11
3. Dissecting the Research Question 12
4. So What? The Far-Reaching Implications of Discourse Analysis 13
A) A Cheap Yet Effective Relationship Monitor? 13
B) Lifting the Veil of Perception 14
Data Collection **16**
1. In Search for the Best Sources of Data 16
2. Data Cleaning 19
A) US Data 19
B) Chinese Data 20
Transforming Tones into Numbers: Sentiment Analysis **23**
1. Sentiment Analysis: A Rich Tradition 23
2. Which Method Works Better? 25
3. Sentiment Analysis in Action 27
4. Visualizing Sentiments 33
Analyzing Sentiments with an Economist's Toolkit: Granger Causality **36**
1. The Mechanics of the Granger Test 36
2. Hot Topic Among Econometricians 38
3. Granger Causality in Action 40
A) Examining Test Assumptions 40
B) Fine-Tuning the Test 41
C) Applying the Test on Different Cuts of Data 42
What Story Does the Data Tell?: Key Findings **46**
1. Granger Causality with Specific Time Windows 46
2. Granger Causality by Topics 49
A) How Often is Each Topic Mentioned? 49
B) What Are the Sentiments of Each Topic? 53
C) Granger Causality 60
3. Granger Causality by Valence 63
4. Overview of Meaningful Findings 66
Discussion of Key Findings **68**
1. Mapping the Topography of Sentiment Time Series 68
2. What Are Ways to Explain Foreign Policy Discourse? 72
A) Discourse as a Reflection of Reality 72
B) Discourse as a Constitutive and Productive Player 74
C) Discourse as Performance and Theater 76
The Long (and Promising?) Road Ahead **78**
Works Cited **81**

“When Premier Zhou Enlai and I agreed on the communiqué that announced the secret visit, he said: ‘This will shake the world.’ What a culmination if, forty years later, the United States and China could merge their efforts not to shake the world, but to build it.”

—Henry Kissinger, *On China*

Abstract

As tensions continue to rise between China and the US, much of the constructive dialogue has been crowded out by virulent insult-hurling from both sides. This paper seeks to explore the patterns of interaction between US and Chinese foreign policy discourse in the context of this trend. Specifically, it applies data-driven methodologies to investigate whether Chinese political discourse vis-à-vis the US is reactive to US rhetoric about China—and vice versa—between 2002 and 2022.

Chinese Ministry of Foreign Affairs press conference transcripts compiled by Mochtak and Turcsanyi, as well as *People's Daily* publications collected by Fisher et. al., are paired with an original dataset with over 35,000 US documents mentioning China from the Presidential Library and Congressional databases. These textual data are then quantified based on sentiments by both lexicon dictionaries and machine learning models. The computed sentiments, transformed into a time series dyad representing the discursive tones of the US and China, are subsequently fitted with a Granger causality model.

This paper finds strong evidence for unidirectional Granger causality with US data as a response, indicating that China seems to initiate rhetorical reconciliation and hostilities more often, while the US plays a more reactive role in bilateral discourse. The Granger models also suggest that the correlation between negative sentiment discourse is much more significant than that between positive sentiment discourse.

The broader implications of analyzing discursive tones as a function of time are two-fold: 1) it offers a cost-efficient way to construct a barometer for the state of bilateral ties whose measurements are accurate in granular time frames over decades, and 2) it is a “proof of concept” that perceptions (and misperceptions) between great powers can be elucidated by incorporating quantitative, natural language processing methods into discourse analysis.

Chilly Summit in Alaska: US-China Relations at a Perilous Crossroads

Senior Chinese diplomats Yang Jiechi and Wang Yi sized up their high-profile counterparts Tony Blinken and Lloyd Austin, the American Secretaries of State and Defense, as they sat down on opposite sides of the long negotiating table. Spectating journalists filling the room tensed up and readied their cameras in anticipation of the conversation that could set the tone of US-China relations for many years to come.

This high-level, bilateral dialogue, held in Anchorage, Alaska in late March of 2021, was in many ways a potential “reset” for US-China relations following the hiatus of cordial diplomatic conversations during the Trump years. Many China watchers in the US wondered if the two countries could once again constructively engage with one another as they attempted under the successive rounds of Bush- and Obama-era Strategic and Economic Dialogues.

But the opening statements spoiled this wishful thinking. Yang, Wang, Blinken, and Austin made it crystal clear that the state of bilateral relations was as chilly as the Anchorage climate.

The US diplomats, seizing their home advantage, opened the talks by pressing China about their repressive domestic policy. The Chinese in turn delivered a 16-minute tirade blasting their counterparts for their “holier-than-thou,” interventionist diplomacy, which far exceeded the agreed-upon two-minute opening statement limit. The indignant Blinken and Austin then halted the reporters who were filing out to get one more riposte on the record.[1]

This seemingly untamed, off-script, and even-worse-than-expected shouting match dominated alarmist newspaper headlines the next day, and importantly, it made two foreign

[1] Thomas Wright, “The US and China Finally Get Real with Each Other,” *Brookings* (blog), March 22, 2021; “Secretary Antony J. Blinken, National Security Advisor Jake Sullivan, Director Yang And State Councilor Wang At the Top of Their Meeting,” *United States Department of State* (blog), accessed April 16, 2023.

policy realities very clear. First, government discourse possesses the capability to make or break relations between states. Setting the tone of how two states communicate with or perceive each other could strengthen a foundation for constructive engagement, or it could send these interlocutors tumbling into a spiral of aggressive posturing and competing nationalisms. Second, the Alaska summit is just one of many examples that foreign policy discourse between two states does not occur in a vacuum; rather, it is an interplay that is constantly being shaped and reshaped through processes of social interaction. As a case in point, Wang Yi and Yang Jiechi delivered their blistering remarks after the stern statements from their counterparts took them by surprise, and this Chinese response in turn elicited an improvised reprimand from the US. How China talks about and perceives the US as well as how the US talks about China are thus very much mutually constitutive.

Given these two lessons from the chilly Anchorage summit, we must recognize that understanding Chinese foreign policy discourse is and will remain a central prerogative for scholars of US-China relations. Being able to effectively parse, understand, and even forecast the words and attitudes emanating from Beijing, moreover, is dependent upon how this discourse is contextualized within a decades-long conversation between the US and China—perhaps the most consequential extended conversation in the 21st century. The goal of this paper is to better probe how this discursive interaction fluctuates through time, as well as the implication of these fluctuations.

Literature Review

There is no shortage of literature analyzing discourse within the realm of foreign policy. Chinese diplomatic rhetoric in particular is very extensively researched. Many China experts situates this subject within the nexus of historical identity, Chinese culture, and a unique Sino-centric worldview. Qing Cao, for example, contends that Chinese foreign policy discourse is inundated with Confucian and moralist sensitivities, but this philosophical underpinning—as well as the foreign policy vision it spawns—has not yet gained a lot of traction externally.[2] Similarly, William Callahan focuses on how the motif of Tianxia (天下, or "all under heaven") and all of its historical baggage has been deployed by China's foreign policy apparatus to promulgate a paradoxical mix of nationalism and cosmopolitanism.[3] A wealth of scholarship also probes how deep-seated historical grievance animates the spaces of Chinese political discourse and foreign policy. Wang Zheng's *Never Forget National Humiliation* is a classic when it comes to this topic.[4]

Other studies choose to focus on the discursive frameworks and motifs of particular generations of leadership. Zheng and Tok, as well as Su Hao, delve into the implications of the Hu Jintao-era mantra "harmonious world," and whether it signals an increased Chinese presence in a changing and complex global order.[5] Semanov and Tsvyk, on the other hand, parse the discursive centerpiece of Xi's foreign policy framework—the "community of common destiny." By dissecting this phrase with very precise semantic and linguistic analysis, the authors contend that China's rhetorical strategies are often not well-explained to—or clearly understood by—external states.[6] Camilla Sørensen provides yet another overview of the dramatic discursive

---

[2] Qing Cao, "Confucian Vision of a New World Order?: Culturalist Discourse, Foreign Policy and the Press in Contemporary China," *International Communication Gazette* 69, no. 5 (October 1, 2007): 431–50.
[3] William A Callahan, "Tianxia, Empire and the World: Soft Power and China's Foreign Policy Discourse in the 21st Century," *British Inter-University China Centre Working Paper Series* 1, no. 2 (2007): 1-24.
[4] Zheng Wang, *Never Forget National Humiliation: Historical Memory in Chinese Politics and Foreign Relations* (Columbia University Press, 2014).
[5] Yongnian Zheng and Sow Keat Tok, "Harmonious Society and Harmonious World: China's Policy Discourse Under Hu Jintao," *Briefing Series* 26 (2007): 1-12; Su Hao, "Harmonious World: the Conceived International Order in Framework of China's Foreign Affairs," *Foreign Affairs* 87, no. 1 (2008): 29-55.
[6] Alexander Semenov and Anatoly Tsvyk, "The Approach to the Chinese Diplomatic Discourse," *Fudan Journal of the Humanities and Social Sciences* 14, no. 4 (December 1, 2021): 565–86.

shift under Xi from "hiding and biding" to "striving for achievement," and offers the tentative suggestion that the drummed-up discourse is occurring in sync with more substantive transformations of diplomatic strategy.[7] Poh and Li, on the other hand, weighed the rhetorical posturing against how China conducts its foreign policy on the ground, and argued that Xi's expanded global footprint and more assertive presence is moderated by a lot of policy continuity.

Some other works that are particularly helpful towards understanding Chinese discourse are Biwu Zhang's *Chinese Perceptions of America*, along with David Shambaugh's *Beautiful Imperialist*. Both works wade into the variegated and complex currents of elite foreign policy discourse for the purpose of understanding the *image* that China holds of the US. The former examines whether probing Chinese perceptions can better help the US better understand its counterpart's geopolitical ambitions and justify, or dispute, the "China threat" theory.[8] The latter develops distinct typologies for Chinese elite discourse vis-à-vis the US, ranging from Marxists, to "ad-hoc" post-Marxists, to hegemonists.[9]

Despite the wealth of highly informative works produced by China watchers, only a sliver of the literature harnesses quantitative methods to conduct discourse analysis. A relevant work at this intersection is Sabine Mokry's "China's Foreign Policy Rhetoric Between Orchestration and Cacophony."[10] Specifically, the author examined the tensions between bureaucratic centralization and atomization under Xi by looking at how the relative numerical salience of specific foreign policy topics changed over time, and how these trends compare across the discursive practices of various political stakeholders.

---

[7] Angela Poh and Mingjiang Li, "A China in Transition: The Rhetoric and Substance of Chinese Foreign Policy under Xi Jinping," *Asian Security* 13, no. 2 (May 4, 2017): 84–97.
[8] Biwu Zhang, *Chinese Perceptions of the U.S.: An Exploration of China's Foreign Policy Motivations* (Lexington Books, 2011).
[9] David Shambaugh, *Beautiful Imperialist: China Perceives America, 1972-1990* (Princeton University Press, 1991), 277-280.
[10] Sabine Mokry, "China's Foreign Policy Rhetoric between Orchestration and Cacophony," *The Pacific Review* 0, no. 0 (February 20, 2023): 1–28.

Compared to literature that primarily focuses on Chinese discourse, there are also relatively few scholarly articles dealing with rhetoric emanating from Beijing in conjunction with discourse from Washington. Some of the more interesting works pertaining to this topic originate from the Chinese academy. Tian Guiqiu, for example, employs Appraisal Theory in psychology to analyze twelve authoritative samples of American and Chinese diplomatic speeches to assess the relative weights of "engagement" (i.e. expressions of judgment) "attitude" (i.e. the degree of force in the expression) and "graduation" (i.e. the extent of the speaker's espousal for the what is being said) embedded within their rhetoric.[11] Luo Ruifeng, on the other hand, identifies critical discourse analysis (CDA) as an investigative starting point to parse nineteen speeches pertinent to foreign policy from Donald Trump and Xi Jinping. After identifying more than two dozen political metaphors (i.e. "nation as a person," "journey,"or "war" metaphors) in these texts, Luo went on to contrast the more positive appeasement qualities of Chinese political discourse with Trump's more divisive, self vs. other patterns of speaking.[12] Zhu and Wang similarly situated Foreign Minister Wang Yi and President Trump's speeches at the 2017 UN General Assembly within the CDA framework. By studying the "linguistic practice," "discursive practice," and "social practice" embedded in these texts, the authors found that both leaders predominantly resorted to the affect of "judgment" in their speeches (i.e. attitudes towards people and how they behave).[13]

The Australian scholar Aditi Bhatia also stands out with an exploration of Jiang Zemin and Bush's press conference rhetoric, once again using CDA. The author finds that press conference language is often meticulously choreographed to hide vulnerabilities and discord

[11] 田桂秋, "基于评价理论的中美外交演讲对比研究" (硕士, 中国海洋大学, 2014).
[12] 罗瑞丰, "中美国家元首外交演讲话语的批评话语分析" (博士, 上海外国语大学, 2020).
[13] Lei Zhu and Wei Wang, "A Critical Discourse Analysis of the US and China Political Speeches—Based on the Two Speeches Respectively by Trump and Wang Yi in the General Debate of the 72nd Session of UN Assembly," *Journal of Language Teaching and Research* 11, no. 3 (May 1, 2020): 435.

while presenting a "united front."[14] This means that both statesmen often use a combination of positivity, evasiveness, and the flexing of influence and power to get their message across.

While much of the literature overviewed above can extract highly nuanced insights about the text (and intertext) of foreign policy discourse, they struggle with a few key limitations. First, the scopes of their methodologies are limited to very minimal samples of texts. Tian, Luo, Zhu and Zhang, as well as Bhatia all survey no more than 20 pieces of documents in their studies, and the discursive landscapes they survey are often confined to a very particular time frame. This means that it is often difficult to glean how discourse fluctuates with time through qualitative analyses. By contrast, using quantitative methods can drastically expand the frontiers of possibility when it comes to the scope or time frames of textual data. Indeed, algorithms derived from natural language processing (NLP) techniques can sift through tens of thousands of documents spanning decades with great efficiency, which gives us a much better idea of how discursive landscapes transform as a function of time period.

What's more, while some papers in this section make comparisons across US and Chinese foreign policy rhetoric, almost none of them focus on the patterns of *interaction* between the discursive practices of these two states. As such, my paper aims to delve into this gap in literature by discovering how US and Chinese political discourse vis-à-vis one another correlate and interact over time. To ensure that I survey adequately large corpora of documents spanning the past two decades from both the US and China, this paper also opts to tread the quantitative path, which is relatively under-explored by scholars of Chinese foreign policy.

[14] Aditi Bhatia, "Critical Discourse Analysis of Political Press Conferences," *Discourse & Society* 17, no. 2 (March 1, 2006): 195.

Research Question

1. Finding the Provocateur: Drilling Down on the Core Hypothesis

This paper seeks to use data-driven methods in order to derive insights about the patterns of interaction between US and Chinese political discourse. More specifically, it postulates the following core research question: *is it possible to pinpoint a "trend-setter" and a "trend-reactor" in the sentiments of US and Chinese government (and media) discourses vis-à-vis each other? If so, who more often plays the role of the sentiment "trend-setter," and who plays the role of the "trend-reactor?"*

In other words, if one were to analyze the turbulent ebb and flow in the US government's sentiments towards China, or vice versa, which side would emerge as the interlocutor that initiates rhetorical hostilities or reconciliations? Who is the provocateur that charts the troughs and peaks of this increasingly precarious, and increasingly consequential, great-power stalemate?

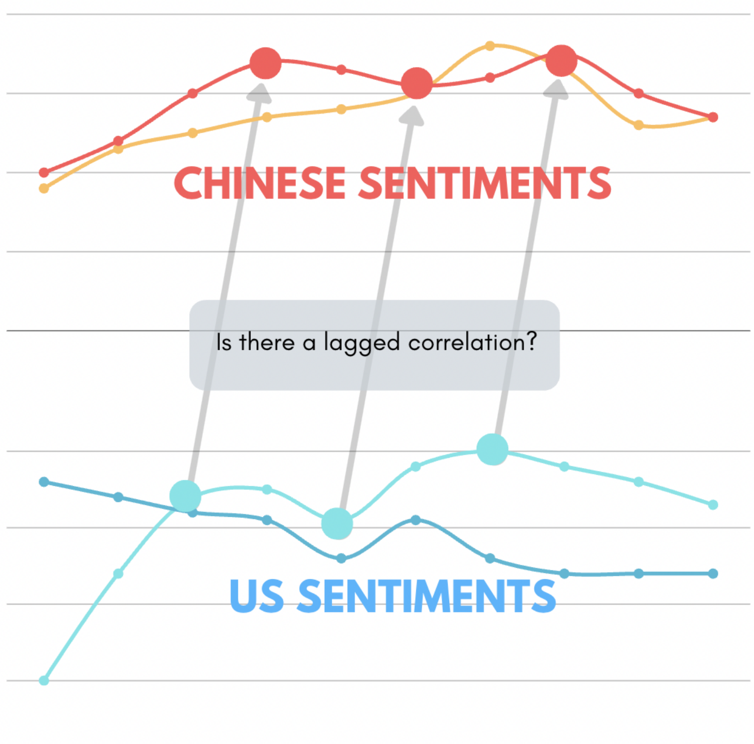

Figure 1: Hypothesizing Patterns of Interaction in US and Chinese Discourse

Figure 1 restates the core research question in color. If vast sets of government rhetoric from both countries were compiled, cleaned, and quantified by sentiments, then perhaps it is possible for one country's sentiment trends to mimic or forecast those of the other. For both countries, the rhetorical gestures of friendship (which we can imagine as peaks in the time series above), along with virulent insults (which are the troughs) might roughly sync up with some degree of "lagged correlation." In Figure 1, for instance, China's curves at times trace the US's sentiment footprint. Its peaks and troughs, as we can see, follow on the heels of those of its counterpart.

2. Clarifying Key Concepts

At this point, it is crucial to resolve some of the lingering ambiguities within the research question. Perhaps the most important of these is the necessity to be more precise about what exactly it means for one set of sentiments to "react" to another. In this paper, I use "reactivity" as a stand-in for lagged correlation. Correlation, of course, is not equivalent to causation, and as such this paper takes additional precaution in identifying causality between US and Chinese foreign policy discourse as a separate concept. Rather, political sentiments from a "trend-setter" merely possess some degree of predictive power for the sentiments of a "trend-reactor." This means that the contours of the US's sentiments in foreign policy contain similar features compared to the Chinese sentiment trends, and these dyads of sentiment features occur in close proximity to one another.

The concept of "sentiment," moreover, is simply a linear spectrum of tones ranging from negative to positive. A negative sentiment, for example, could be expressed as follows by a Chinese Ministry of Foreign Affairs spokesperson at a routine press meeting:

> A handful of American politicians, entrenched in their Cold War mentality and ideological bias, are bent on hyping up the so-called China threat theory. China is firmly against this. I want to stress that containment and suppression cannot stop China from growing stronger, but will only damage mutual trust and cooperation and escalate frictions.[15]

Positivity, on the other hand, could take the following form:

> Both sides [i.e. the US and China] should take more proactive measures to further facilitate people-to-people exchanges and create more enabling conditions for across-the-board exchanges and cooperation between our two countries.

In addition, "government discourse" in this context simply points towards any type of speeches, reports, or any other relevant texts emanating from an organ of the US or Chinese government that has the capacity to influence—or accurately broadcast—the foreign policy agenda. Discourse should also not be erroneously mixed up with the actual practice of foreign policy. As we shall see in the subsequent section, how the scope of "government discourse" is outlined largely depends on data availability.

3. Dissecting the Research Question

The US-China relationship is an ever-changing geopolitical saga with a high degree of dynamism and countless moving parts. Given the sheer complexity of bilateral ties, it would be hard to condense their variegated foreign policy discourse about each other into just two monolithic sentiment trends. As such, this paper also seeks to break the central research question down into more manageable chunks and investigate these components up-close from different perspectives.

For instance, we might ask how the patterns of interaction between US and Chinese discourse vary by foreign policy issue areas. Are there particular topics that prompt the US or China to take the rhetorical reins and set sentiment trends, whereas in other topics, the former trend-setter slips into a more reactive role? Moreover, do the discursive behavior of the US and

[15] All the relevant documents and excerpts of documents in my datasets can be found in my Honors Thesis Github repository: https://github.com/alinlzx/honors-thesis.

China in any of these topics correlate with how these topics are interpreted or acted upon in both countries' foreign policy agenda?

Another way to take a more fine-grained look at this relatively broad research question is to partition all the discourse that I am studying along a positive-negative divide. This creates an opportunity to decipher whether a "lagged correlation" in the overall datasets is driven by cycles of vicious, rhetorical escalations, or by reciprocity of more generous and friendly rhetorical gestures. Looking at patterns of interaction within just positive- or negative-sentiment discourse also elucidates possible tendencies for one country to solely exhibit "trend-setting" or "trend-following" behavior in one type of rhetoric, but not the other.

By zeroing in on smaller windows of time within broader sentiment trends, this paper gains yet another perspective into the research question. Indeed, in a highly dynamic and uncertain geopolitical landscape, the discursive realities of 2010 may no longer be applicable today. This means that salient patterns of interaction between US and Chinese sentiments vis-à-vis one another might only materialize when we zoom into particular time frames.

4. So What? The Far-Reaching Implications of Discourse Analysis

A) A Cheap Yet Effective Relationship Monitor?

The utility of parsing the patterns of interaction between US and Chinese political discourse are two-fold. The most obvious reason for tracking the ebb and flow of the two great powers' prevailing sentiments towards each other is that these statistics serve as powerful, relatively cost-efficient, and highly interpretable proxy for the warmth of bilateral relations. In this sense, by gathering and quantifying the emotionally-charged language across the two countries' foreign policy discursive landscape, it is possible to create a bilateral relations "thermometer" whose measurements are precise down to very granular increments of time, and whose historical data points spanning decades can be conducive to understanding great power politics in the long-term, and on a macro-scale.

To be sure, more and more scholars and institutions are trying to publish their own relationship monitors. Morning Consult, for example, curates and consistently updates a US-China Relations Barometer, which derives the positivity of US perceptions towards China (or vice versa) based on iterative polling of the general populace.[16] Highly influential research and educational institutions, from the Carter Center to Georgetown, have set up projects dedicated to monitoring the turbulent U.S.-China relationship and in turn produced truly invaluable contributions that have inspired this paper.

However, whether it is state-of-the-art surveys or top-of-the-line research, gauging the temperature of the multifaceted ties between the US and China is daunting, time-consuming, and potentially expensive work. The survey methodology runs into further obstacles when we try to move from studying relatively more accessible public opinion to piercing the veil of secrecy that often shrouds policymakers' opinion. These problems are compounded when we try to expand our inquiry about the evolution of bilateral relations over two or more decades.

The innovation of this paper is that it looks for a fast, efficient, and potentially lesser-explored way of constructing such a long-term yet high-resolution relationship monitor. Specifically, I would gather enormous mounds of textual data (i.e. discourse) pertinent to the US-China relationship, and then input these extensive datasets into a sentiment analysis pipeline, which in turn would generate quantifiable measures of the warmth of bilateral ties.

B) Lifting the Veil of Perception

As implied in the previous section, taking measurements of the temperature of US-China relations naturally treads upon a perennial puzzle that has dominated the field of international

[16] Morning Consult, "U.S.-China Relations Barometer," accessed April 16, 2023.

relations—how does one accurately identify the perceptions of country A's leaders towards country B and vice versa?

Once again, surveys and in-depth qualitative research can be costly, while both run up against the challenge that true elite perceptions are often kept behind closed gates. The near-impossibility of piercing the veil of perceptions is what animates the "tragedy of great power politics," as per Mearsheimer, because uncertainties vis-à-vis the intentions of a great power rival would continuously sow discord and distrust into an already uneasy relationship.[17] In *Perception and Misperception*, moreover, the great international relations (IR) scholar Bob Jervis points out that a security dilemma, at its essence, can be traced back to the daunting challenge of understanding mutual perceptions between two states.[18] In this sense, the "original sin" spoiling US-China rapport is a psychology of distrust and suspicion that thrives amidst an environment of uncertainty.

Andrew Scobell, applying Jervis to US-China relations, remarks that both states are domestically pre-occupied, struggle with "chronic empathy deficiency," and thus are growing alarmed by elevated threat perceptions as their power differentials shrink. "Images of the other as implacable adversaries appear firmly embedded in the cognitive maps of Washington and Beijing," writes Scobell while predicting that this psychological drama will drag the two great powers deeper into the muddy waters of competition and decoupling.[19]

This paper by no means aspires to fully crack open the perennial puzzle of misperception. Rather, it postulates whether perceptual uncertainties can be methodically reduced by fielding immense sets of textual data containing what the US and Chinese governments say about each

---

[17] John J. Mearsheimer, *The Tragedy of Great Power Politics* (W. W. Norton & Company, 2001), 31.
[18] Robert Jervis, *Perception and Misperception in International Politics: New Edition* (Princeton University Press, 2017), xv.
[19] Andrew Scobell, "Perception and Misperception in U.S.‑China Relations," *Political Science Quarterly (Wiley-Blackwell)* 135, no. 4 (December 2020): 664.

other. In other words, my research attempts to marry a theoretical conundrum in IR with the core, underlying philosophy in quantitative analysis—namely, by parsing and synthesizing a large number of accessible data points that very imperfectly represents bilateral perceptions, it is possible to *estimate* the true parameter of perceptual realities undergirding the US-China stand-off.

In fact, variations of this method have already been carried out successfully and cited widely by the scholarly community. Biwu Zhang's *Chinese Perceptions of the US*, as well as David Shambaugh's *Beautiful Imperialist*—two of the most authoritative studies delving into the Chinese elites' image of the US—both make inferences on elite perceptual realities by employing wide swaths of Chinese documents ranging from articles published in foreign policy institute-affiliated journals to restricted-circulation (内部) periodicals to newspapers.[20]

This paper hopes to help lift the veil of perception by taking Zhang and Shambaugh's research premises further into quantitative territory, harnessing even more data points, while also expanding their methods into a two-way inquiry (i.e. I hope to study not just how Chinese elites perceive the US, but also how the US perceives China, and how perceptions interact). The next few sections will delve into the mechanics of how my original datasets were curated, and how it was interpreted through an amalgam of natural language processing (NLP) and time series analysis.

Data Collection

1. In Search for the Best Sources of Data

This paper aims to collect and consolidate a corpus of documents—including press conference transcripts, speeches from policymakers, white papers, policy briefs, or any other

[20] Zhang, *Chinese Perceptions of the U.S.*, 18-24; Shambaugh, *Beautiful Imperialist*, 37-39.

entries within government archives—that adequately represents China's political discourse vis-à-vis the US, as well as the US's rhetoric towards China. Preferably, these documents should permeate all key issue areas pertaining to US-China relations, and because the research question calls for robust quantitative analysis that drills down to granular time frames along the long-term evolution of bilateral relations, the datasets should also achieve a reasonable volume so that the document count maintains an adequate density over an extended period of time.

Fortunately, data on Chinese foreign policy discourse came ready-made. Mochtak and Turcsanyi curated the impressive Chinese Ministry of Foreign Affairs Press Conferences Corpus. This dataset contains roughly 23,000 question-answer dyads, ranging from October 2002 to December 2021, that were extracted from press conference transcripts published on the MFA's English website.[21] Since the MFA acts as China's preeminent diplomatic interlocutor, statements from its spokespersons are likely the most authoritative representation of the government's official narrative vis-à-vis the US, and they would also be the primary channel through which the CCP disseminates its newly-minted foreign policy slogans and lingo to an international audience. As such, Mochtak and Turcsanyi's MFA corpus constitutes a core component of my own research.

A vast collection of English *People's Daily* publications, generously provided by Fisher and Klein through the FOCUSdata project, supplemented the MFA corpus to round out my Chinese foreign policy dataset. The *People's Daily* corpus captures a staggering 190,000 articles ranging from June 2007 to February 2020.[22] Although not strictly a policy-making or diplomatic apparatus like the MFA, the *People's Daily* serves as a trusted mouthpiece for the CCP, which

[21] Michal Mochtak and Richard Q. Turcsanyi, "Chinese Ministry of Foreign Affairs Press Conferences Corpus (CMFA PressCon)" (Harvard Dataverse, July 1, 2022).
[22] Scott Fisher, "China's People's Daily Articles from 12 May 2007 to 31 Dec 2020" (Harvard Dataverse, April 16, 2022); for more details on how the FocusData project collected its vast datasets, and how it conducted preliminary sentiment analysis, see Scott Fisher, Graig R Klein, and Juste Codjo, "Focusdata: Foreign Policy through Language and Sentiment," *Foreign Policy Analysis* 18, no. 2 (April 1, 2022).

means that anything that it puts out would have to be meticulously processed and filtered to echo official political discourse emanating from Beijing.

Surprisingly, pre-existing datasets that trace US foreign policy rhetoric are scant, if they exist at all. This means that an original data collection endeavor is warranted. In order to maximize congruence between the China and US data, the intuitive source of documents here would have been State Department press releases. However, these transcripts are scattered across multiple archived websites with non-uniform interfaces corresponding to the Clinton, Bush, Obama, and Trump presidencies. Navigating this sort of digital environment with an automated scraper to gather China-related statements would have been particularly cumbersome. Thus, this paper instead chose to take advantage of HeinOnline's diverse array of federal government databases.

Using a Python script, as well as a chrome webdriver powered by Selenium, I built a scraper that searched for and compiled any documents that included the named entities "China," "Chinese," "CCP," and "PRC" within HeinOnline's Presidential Libraries since 2000.[23] The custom dataset pooled together roughly 4,800 entries sourced from the Daily Compilation of Presidential Documents and the Public Papers of five administrations, almost all of which are transcripts of the President's public statements. Given the centrality of the White House in shaping foreign policy, these data points are likely an authoritative proxy of the US's prevailing stance towards China.

Of course, America's variegated, and often polarizing, policy-making landscape is often defined by contradictory positions and rhetoric. Members of Congress, for instance, can voice their own (and often more hawkish) opinions on China in attempts to sway foreign policy,

[23] Instead of scraping documents in their entirety, I only collected excerpts and passages around the occurrences of the named entities of interest for the sake of relevance.

although their influence in this political arena is often eclipsed by the executive branch.[24] Nevertheless, this paper also considered the Congressional Documents housed in HeinOnline's archives to be informative. With a similar scraper and identical search criteria, I compiled another dataset totaling nearly 30,300 documents, which are all primarily sourced from Congressional Research Service reports, as well as roughly fourteen volumes of Congressional Records detailing everyday activities on the Senate and House floors.

2. Data Cleaning

A) US Data

Because this paper aims to look at political rhetoric as a function of time, the documents that did not come attached with a timestamp are not informative to subsequent analysis. This made cleaning the Presidential dataset particularly challenging, as HeinOnline did not provide full dates for most Public Papers data in their search output. As a result, I pulled timestamps from the texts or subheadings of several hundred otherwise dateless entries.

Additionally, given the fact that the White House holds the reins of foreign policy, whereas Congress may often take on the role of a rowdy passenger, the proportion of texts analyzed from both samples are adjusted accordingly. This means that the combined US dataset privileges Presidential data and incorporates all of its dated entries. It then supplements documents from the Presidential Library with a random sample of 2000 Congressional data points. Figure 2 traces the frequency of US documents over time, disaggregated by the compilation or archive from which they originated.

[24] Louis Fisher, "Foreign Policy Powers of the President and Congress," *The Annals of the American Academy of Political and Social Science* 499 (1988): 148–59.

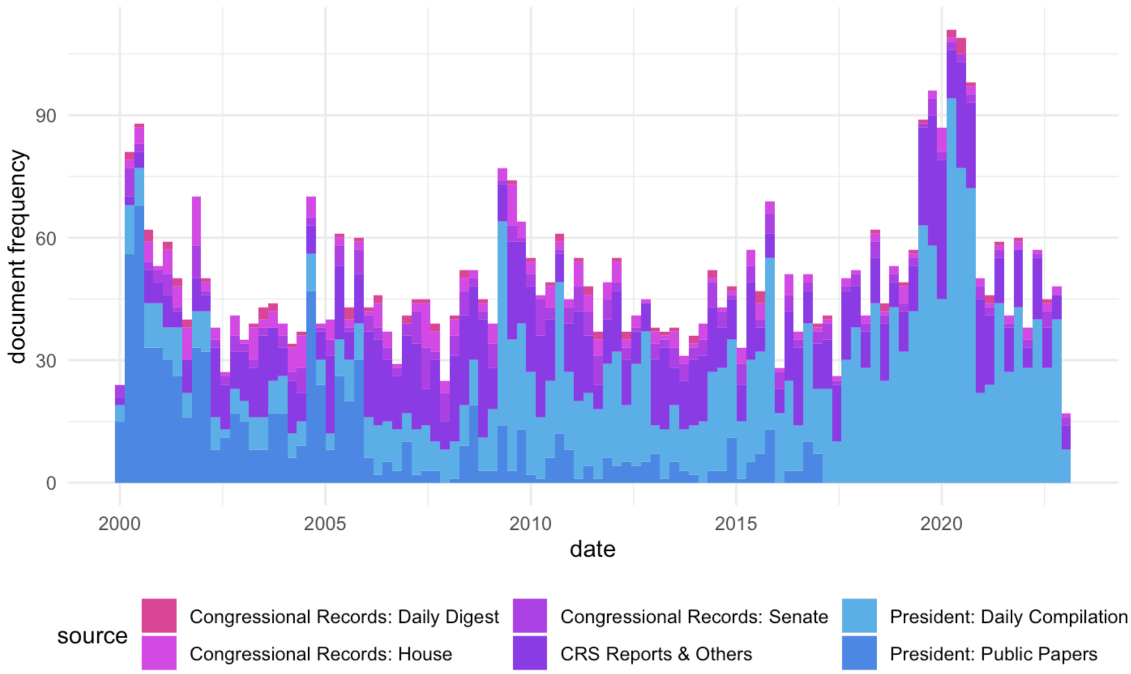

Figure 2: Relevant Document Frequency, US Data

B) Chinese Data

The first stage in the Chinese data cleaning pipeline was to extract each entry within the MFA and *People's Daily* datasets that directly mentioned the United States (or its acronyms, "US," or "U.S.").

Moreover, similar to compiling the full US dataset, it is important to recognize that the MFA enjoys much greater proximity to China's foreign policy-making processes and is thus a more authoritative representation of Beijing's true posture towards the US. Once again, it is advisable to emphasize the weight of the MFA press corpus by taking the data in its entirety, and supplementing the MFA data with a sample of *People's Daily* articles.

Processing *People's Daily* articles required an extra step, as not every publication was immediately relevant (for instance, the news outlet circulated extensive coverage of US financial market performance, which does not necessarily correspond to the Chinese government's perception of the US). The nature of this dataset thus called for an automated relevance filter operating on a machine learning algorithm, whose architecture will be covered in more detail in the sentiment analysis section. The filter trimmed approximately half of the sampled dataset, leaving behind 2480 dated entries.

We might also note that since Mochtak and Turcsanyi's China MFA dataset is populated by question-answer dyads, my analysis exclusively focuses on answers from government spokespersons in the interest of locating the most precise representation of the Chinese government's perception of the US.

Figure 3 illustrates the document count of US mentions across the MFA and *People's Daily* corpora over time relative to the total document counts, after both datasets have been thoroughly sampled, truncated, and labeled.

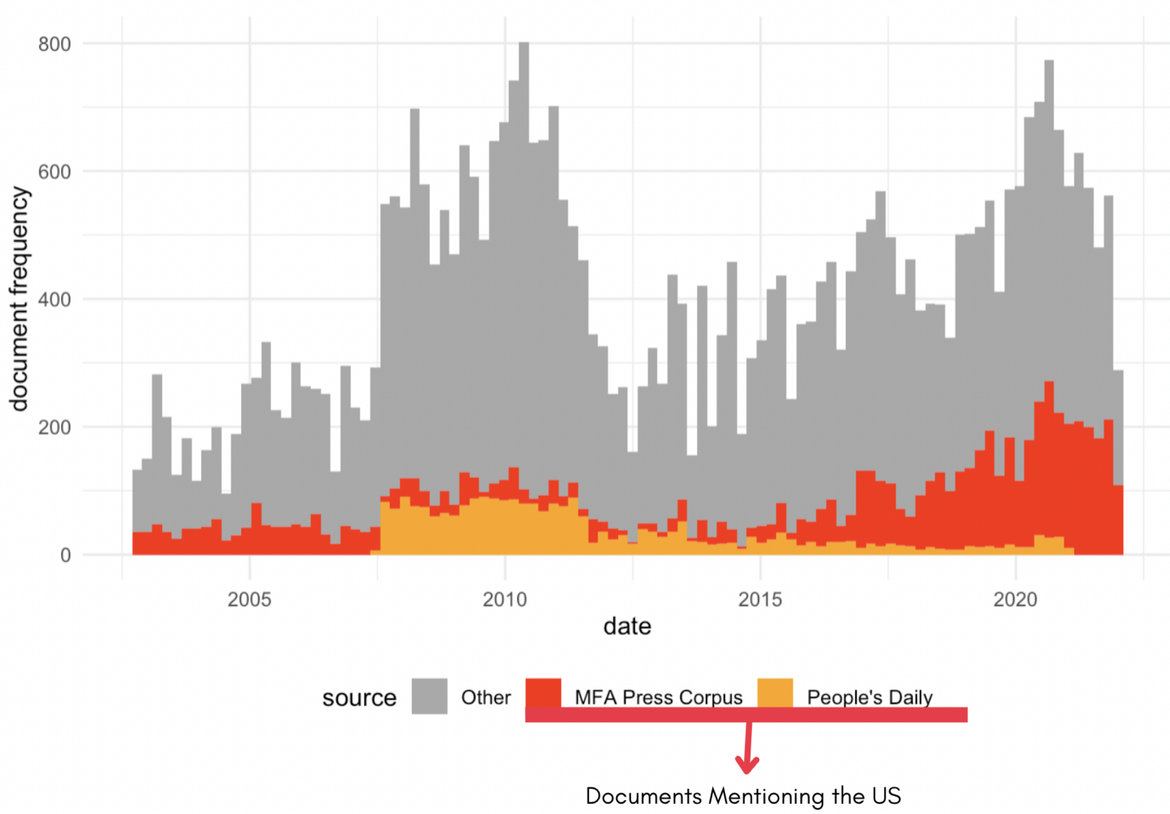

Figure 3: Relevant Document Frequency, China Data

Two factors may have pushed the counts of relevant MFA documents upwards between around 2015 and present day. First, as we can see, the proportion of MFA press transcripts mentioning the US has grown dramatically over time, indicating that the US is occupying an increasingly prominent position within the Chinese foreign policy calculus. What's more, the Chinese MFA is also likely growing more communicative—or even assertive—over time by holding more frequent and extensive press conferences, hence the rise in document frequency.

Figure 4 illustrates the composition of the US and China datasets that have passed through the various date and relevance filters. Overall, around 15,200 texts are used in subsequent analyses, of which 6000 are sourced from the MFA, 2480 are sampled from the *People's Daily*, 4760 are derived from the Presidential Library, and 2000 are sampled from Congress. In the next section, these data points of interest will be further scrutinized through sentiment and time series analysis.

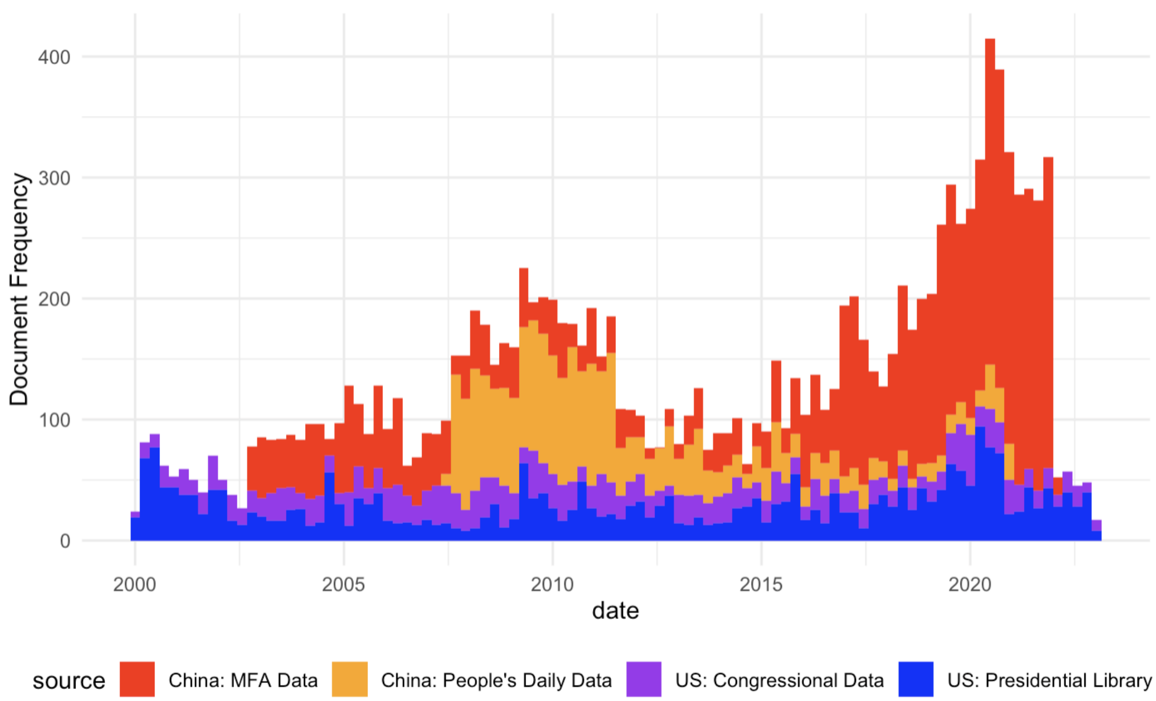

Figure 4: Composition of Relevant Data

Transforming Tones into Numbers: Sentiment Analysis

1. Sentiment Analysis: A Rich Tradition

In order to efficiently consume, parse, and draw insights from this vast set of foreign policy documents, this paper turns to sentiment analysis, a foundational methodology in the world of natural language processing whereby a computer algorithm automatically classifies the sentiment polarity of a vast collection of texts.

There is a rich tradition of applying sentiment analysis to study discourse animating US and Chinese social media platforms. Cairns and Carlson, for example, quantified the emotional outbursts rippling through Weibo (the Chinese equivalent of Twitter) during the 2012 Senkaku/Diaoyu islands dispute in order to understand the expression and censorship of nationalism in China's digital space.[25]

On the US side, Chambers et. al. looked at how Twitter users perceived foreign countries in order to better understand the public-governing elite linkage in international relations.[26] Their use of named entity recognition, as well as an extensive pipeline of relevancy filters for Twitter country data, was informative for designing my own data cleaning procedure.

NLP's footprint also extends beyond social media into political discourse analysis, whether this means looking at the sentiments in parliamentary debates to study MP attitudes

[25] Christopher Cairns and Allen Carlson, "Real-World Islands in a Social Media Sea: Nationalism and Censorship on Weibo during the 2012 Diaoyu/Senkaku Crisis," *The China Quarterly* 225 (March 2016): 23–49; the use of sentiment analysis to study popular discourse has skyrocketed as researchers become interested in the socio-political implications of COVID-19. As a case in point, see Xiaoting Lyu et al., "Sentiment Analysis on Chinese Weibo Regarding COVID-19," in *Natural Language Processing and Chinese Computing*, ed. Xiaodan Zhu et al., Lecture Notes in Computer Science (Cham: Springer International Publishing, 2020), 710–21.

[26] Nathanael Chambers et al., "Identifying Political Sentiment between Nation States with Social Media," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (EMNLP 2015, Lisbon, Portugal: Association for Computational Linguistics, 2015), 65–75; there are also plenty of studies that leverage sentiment analysis to study popular opinion during election cycles. Some examples include Brandon Joyce and Jing Deng, "Sentiment Analysis of Tweets for the 2016 US Presidential Election," in *2017 IEEE MIT Undergraduate Research Technology Conference (URTC)* (2017 IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA: IEEE, 2017), 1–4; as well as Hassan Nazeer Chaudhry et al., "Sentiment Analysis of before and after Elections: Twitter Data of U.S. Election 2020," *Electronics* 10, no. 17 (January 2021): 2082.

towards motions on the floor (Abercrombie and Batista-Navarro, Proksch et. al.), or contrasting the rhetorical strategies of Donald Trump and Hillary Clinton in the 2016 elections (Liu and Lei).[27]

Fewer scholars, on the other hand, have sought to extend the applicability of this ubiquitous NLP toolkit to the realm of foreign policy. Mochtak and Turcsanyi leveraged the Bing Liu lexicon to conduct some exploratory analysis of sentiment polarity on the MFA dataset.[28] Choi et. al. also blazed a path towards this under-studied intersection by using the AFINN lexicon to compare the sentiments of speeches from members of Congress regarding the US's ratification of two FTA deals.[29] Zeng et. al. started from a premise intimately linked with my own research question, whereby the authors sought to regress sentiment scores of US news media based on the Bing Liu lexicon (which acts as a proxy for US-China tensions) against US imports and disruptions to the global value chain during the trade war.[30]

However, scholars have yet to experiment with more robust, machine learning-driven algorithms on foreign policy texts, or parse the patterns of interaction between two countries' rhetoric by quantifying how they express emotions towards each other.

---

[27] Gavin Abercrombie and Riza Batista-Navarro, "`Aye' or `No'? Speech-Level Sentiment Analysis of Hansard UK Parliamentary Debate Transcripts," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (LREC 2018, Miyazaki, Japan: European Language Resources Association (ELRA), 2018); Dilin Liu and Lei Lei, "The Appeal to Political Sentiment: An Analysis of Donald Trump's and Hillary Clinton's Speech Themes and Discourse Strategies in the 2016 US Presidential Election," *Discourse, Context & Media* 25 (October 1, 2018): 143–52.
[28] Michal Mochtak and Richard Q. Turcsanyi, "Studying Chinese Foreign Policy Narratives: Introducing the Ministry of Foreign Affairs Press Conferences Corpus," *Journal of Chinese Political Science* 26, no. 4 (December 1, 2021): 743–61.
[29] Jaedong Choi et al., "Unraveling the Thought Probes of US Legislators on Trade Negotiations: Sentiment Analysis of the 108th and 112th US Congressional Speeches," *Pacific Focus*.
[30] Ka Zeng, Rob Wells, Jingping Gu, and Austin Wilkins, "Bilateral Tensions, the Trade War, and US–China Trade Relations," *Business and Politics* 24, no. 4 (December 2022): 403-406.

2. Which Method Works Better?

Kolchyna et. al.'s overview of sentiment analysis concisely laid out the advantages and downsides of two broad families of methods widely used for text classification.[31] On one hand, lexicon-based approaches evaluate words in a text and label their sentiment orientation based on an extensive valence dictionary. The sentiment polarity of each word then contributes to the overall sentiment "score" of the complete text. Examples of this approach abound. The AFINN lexicon compiled by Finn Nielson, as well as the Bing Liu lexicon developed by Hu and Liu place a vast array of opinion words on a numerical spectrum. VADER, on the other hand, stands out for considering words alongside syntax (i.e. the appearance of "not" would reverse the valence of the subsequent word), and has been acknowledged by scholars as a better-performing dictionary compared to its peers. However, because it considers emoticons and is particularly tailored to shorter-form, social media documents, my sentiment analysis pipeline did not incorporate VADER. Dictionaries like NRC, meanwhile, classify words among a variety of emotional states.[32] For the purposes of this paper, AFINN and Bing Liu are selected, as I am only interested in quantifying pure textual data along a binary sentiment spectrum.

Comparing texts against a predefined dictionary is very "economical," in that it puts a lower computational load on the model and typically returns reliable, consistent results. The disadvantage here, however, is that the dictionaries we use might not be well-adapted to specialized vocabulary, such as the talking points and domain-specific jargon thrown around by state leaders and diplomats.

---

[31] Olga Kolchyna et al., "Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination" (arXiv, September 18, 2015).

[32] A succinct introduction to the various methods of lexicon-based sentiment analysis can be found in M. A. Al-Shabi, "Evaluating the Performance of the Most Important Lexicons Used to Sentiment Analysis and Opinions Mining," 2020, 52-54.

Here, another family of methods—supervised machine learning approaches—become relevant. Supervised learning algorithms take a sample of "training" data, whose sentiment has already been labeled, maps each word and its frequency of occurrence onto numerical vector space (i.e. vectorization), and identifies the features separating the different classes of sentiment in the vector space. It then "infers" a function that will help classify new textual inputs. This procedure is particularly of interest for this paper, as it tailors the machine learning algorithm to consider specialized, foreign policy-oriented vocabulary that might be less compatible with traditional sentiment analysis lexicons.[33]

A well-known machine learning model is Naive Bayes, which employs Bayesian probability to determine how likely a new input text belongs to a sentiment class, given the "prior" probability of a sentiment category associated with features of the text that the algorithm has learned through training.

Another notable method is support vector machines (SVMs), which places a training data's features in a multi-dimensional space and identifies a hyperplane that can bifurcate the training data along its sentiment classes, while also maximizing the distance between each class of data on the hyperplane. SVMs thrive in environments of high-dimensionality, which is frequently the case for text classification.

Decision trees, on the other hand, create a hierarchical classification structure based on the training data categories, and process new pieces of data by passing them down a series of "nodes" that make binary decisions about which sentiment class this new data ultimately belongs to. Although highly interpretable, a single decision tree often struggles from poor accuracy.

---

[33] Olga Kolchyna et al., "Twitter Sentiment Analysis"; Binita Verma and Ramjeevan Thakur, "Sentiment Analysis Using Lexicon and Machine Learning-Based Approaches: A Survey," in *Lecture Notes in Networks and Systems*, 2018, 444-45.

Hence, quantitative researchers often use several decision trees at once—intuitively dubbed a random forest—that "vote" on the output sentiment category.

Some NLP researchers would also venture into the territory of deep learning, which features state-of-the-art methods like LSTM or neural networks. Considering the trade-off between computational cost and model efficacy, these algorithms were not included within the scope of my research, but they may be a fruitful next step in further understanding how foreign policy discourse interacts.

The scholarly consensus seems to indicate that rather than choosing a single approach, a hybrid method combining lexicon-based and supervised learning algorithms (possibly by simply taking the average of an "ensemble" of outputs from different models) typically gives even better performances.[34]

3. Sentiment Analysis in Action

This paper experiments with six overall text classification methods using R: by applying the Bing Liu and AFINN dictionaries within the lexicon-based family, as well as by using SVM, random forest, decision trees, and Naive Bayes within the machine learning family. The following figure illustrates the full design of my sentiment classification algorithm:

[34] For the viability of hybrid models, see Alessia D'Andrea, Fernando Ferri, Patrizia Grifoni, and Tiziana Guzzo, "Approaches, Tools and Applications for Sentiment Analysis Implementation," *International Journal of Computer Applications* 125, no. 3 (2015); Verma and Thakur, "Sentiment Analysis Using Lexicon and Machine Learning-Based Approaches"; for the effectiveness of taking simple ensemble averages, see Jacqueline Kazmaier and Jan H. Van Vuuren, "The Power of Ensemble Learning in Sentiment Analysis," *Expert Systems with Applications* 187 (2022).

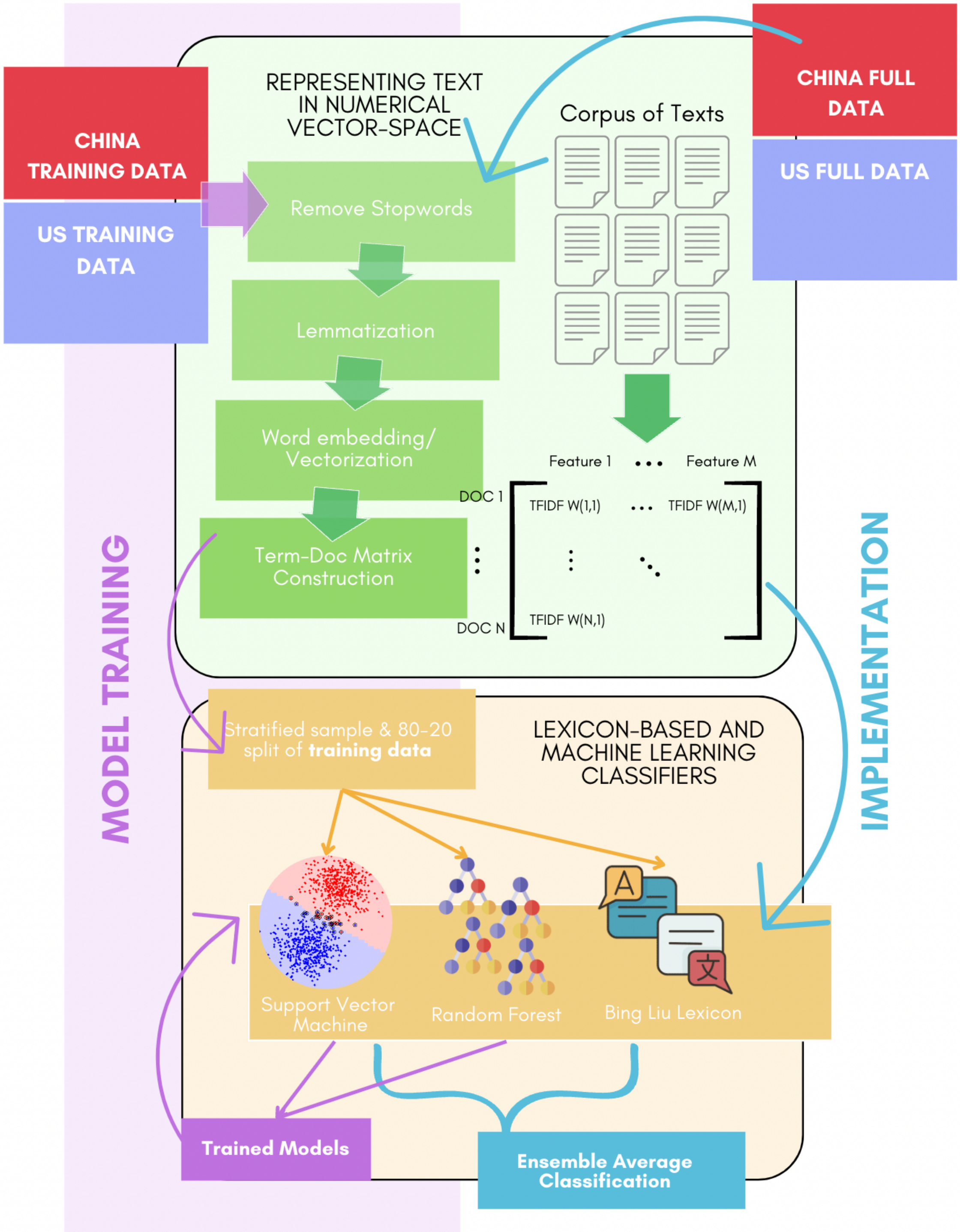

Figure 5: Design of the hybrid sentiment classification methodology

Working with Bing Liu and AFINN lexicons were relatively straightforward. This paper enlisted the syuzhet library within R to compare each text with these predefined lexicons and derive a sentiment "score."

Implementing the various machine learning algorithms, however, required jumping through a few more computational and data processing hoops. First, a random sample of 600 texts were pulled from each of the two countries' datasets (totaling 1200 documents). I then manually read through each document and coded them with three sentiment states: positive, negative, or ambiguous/neutral.

These two training datasets then passed through a stopwords filter provided by the tm library (which removes ubiquitous English words such as "a," "the," and most prepositions). Subsequently, the filtered texts are lemmatized (i.e. words built upon the same stems are grouped together), and vectorized. The two vectorized corpora are then reshaped into term-document matrices (TDM), in which each document and every unique feature (or word) in the dataset make up the rows and columns, respectively, while the value of each cell is weighted with a term frequency-inverse document frequency (TF-IDF) scheme:

$$W_{i,j} = tf_{i,j} \cdot log(\frac{N}{df_i})$$

Where W denotes the weight of the i^{th} feature within text j, tf is the frequency of the feature in text j, N is the number of texts, and df is the count of documents in which feature i is found.

Given the size and sparsity of the TDMs, it would be hard for the computer to learn anything meaningful from the training texts. Therefore, I experimented with an additional tuning parameter that discarded any features containing no more than k non-zero values to re-construct "slimmer" versions of the TDMs (k = 3 yielded optimal results for both the US and China datasets).

Because a sizable proportion of the labeled data were classified to have neutral or ambiguous sentiments, I applied three different train-test split frameworks to see which one maximized classification accuracy and reliability. The simplest approach performed an 80-20 train-test split for only positive and negative sentiment documents, whereas the second approach included three sentiment categories—a stratified sample of positive, negative, and ambiguous texts—to train the models. The third approach, entailed re-coding the three-category sentiment label as two binary classes to train the models twice.

After partitioning the data along these three schemes, I implemented each of the four machine learning methods of interest, of which SVM and random forest gave the best performances. Following conventional scholarly wisdom, I also took an ensemble average of the two top supervised models and the normalized Bing Liu scores to derive a "hybrid" sentiment score. Figure 6 records the F1 and accuracy scores for the "hybrid" scores computed by all three train-test frameworks when they are applied to both the US and China testing datasets. Evidently, approach 2, which performs a stratified sample of all three sentiment classes, puts up the best performance:

	China Test Data		**US Test Data**	
	F1 Scores	*Accuracy Rate*	*F1 Scores*	*Accuracy Rate*
Approach 1 (Trained with only positive and negative documents)	Neg: 0.5667 Ambiguous: 0.4151 Pos: 0.7925	59.04%	Neg: 0.4130 Ambiguous: 0.5373 Pos: 0.4706	48.46%
Approach 2 (Trained using all three sentiment	Neg: 0.8 Ambiguous: 0.5625 Pos: 0.793	74.67%	Neg: 0.7595 Ambiguous: 0.343 Pos: 0.588	62.16%

categories)				
Approach 3 (Trained using two sets of binary labels)	Neg: 0.6792 Ambiguous: 0.2308 Pos: 0.6897	62.11%	Neg: 0.5217 Ambiguous: 0.6195 Pos: 0.471	56.48%

Figure 6: Model Performance Metrics for Three Train-Test Approaches

For Chinese texts mentioning the US, the ensemble classifier—containing SVM, random forest, and Bing Liu—was able to achieve an accuracy of up to 74.7% when working with all three types of sentiments (as shown in the figure above).[35] The accuracy level jumps to 88% when we only evaluate the model's performance vis-à-vis positive or negative documents.

The ROC curves for each of the three classes of predictions, as well as the confusion matrix for positive and negative test data, are as follows:

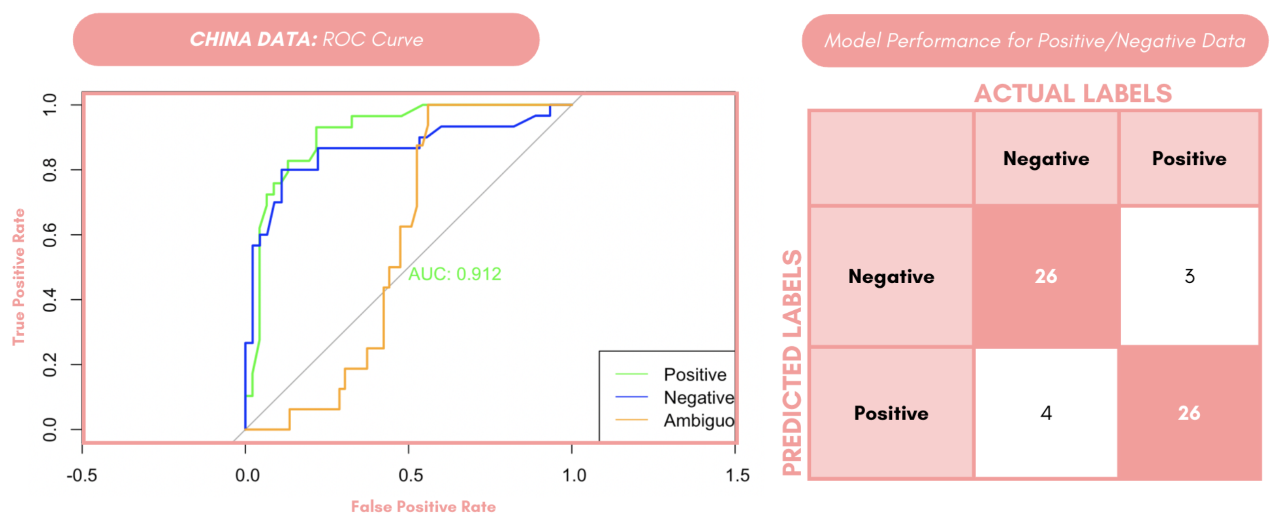

Figure 7: Chinese Data Model Diagnostics

35 This accuracy level is estimated because the hybrid label can be anywhere on a continuous numerical spectrum. Therefore, I stratified the predicted label based on the relative proportions of the actual label in order to construct a three-class confusion matrix. This means that an ROC curve, which plots the true positive and false positive rates of the model, is a more appropriate performance diagnostic tool.

The sentiment analysis model for American data, unsurprisingly, also excelled in parsing negative and positive documents while struggling to sort out ambiguous texts. The model can achieve an accuracy of around 62% for all three classes, but this statistic shoots up to 84% when isolating for unambiguous documents. The ROC curve and two-class confusion matrix for US data are captured in Figure 8:

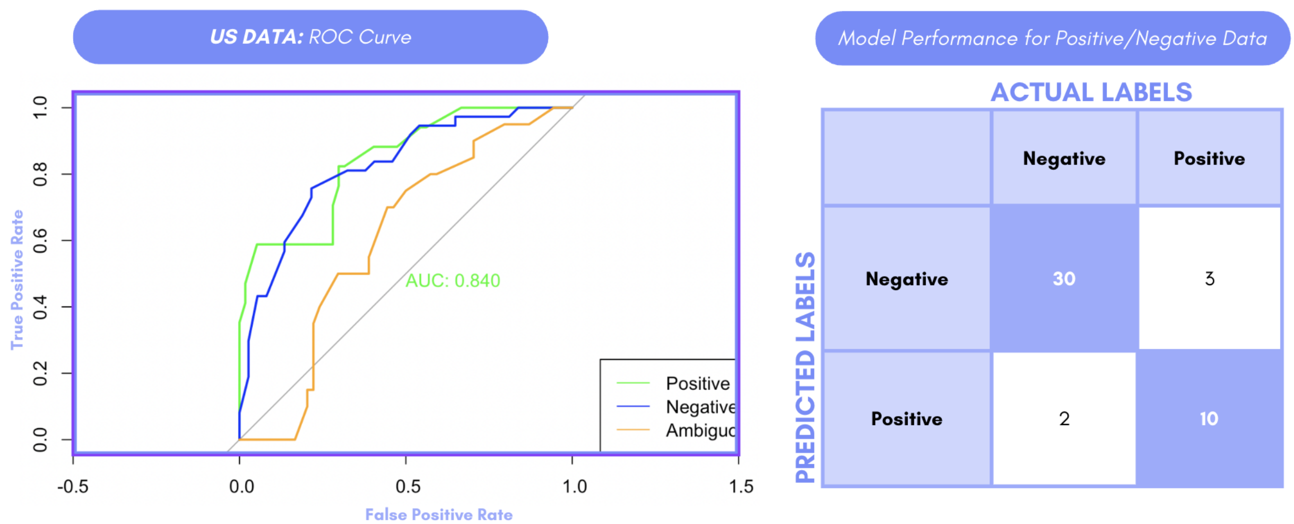

Figure 8: US Data Model Diagnostics

At this stage, all of the MFA and Presidential Library data, as well as a random sample of 2000 Congressional data points, are put into the extensive textual "make-over" pipeline consisting of the stopwords removal, lemmatization, vectorization, and TDM conversion stages. Processing *People's Daily* data, as mentioned in the previous section, turned out to be somewhat more complex due to the presence of irrelevant documents.

In order to effectively categorize the pertinence of the state-run newspaper corpus, I manually coded the 300-document sample of *People's Daily* data for their relevance, and created a variation of my sentiment classifier to be trained upon this labeled data. Specifically, this means combining the SVM and random forest engines with a custom lexicon-based filter that flagged words commonly occurring in irrelevant articles. These three methods then voted on which articles to exclude to return an updated corpus ready for sentiment analysis.

4. Visualizing Sentiments

Quantifying sentiments within these extensive corpora opens up an entire frontier of promising possibilities for analyzing patterns of interaction between US and Chinese rhetoric, the most intuitive of which is to study sentiment trends as a function of time.

As such, the sentiment labels within each of the MFA, *People's Daily*, Presidential Library, and Congressional datasets are grouped by varying intervals of time (i.e. weeks or months). This allows government discourse to be mapped onto time series that can be easily visualized, interpreted, and correlated.

In Figure 9, we can observe the fluctuations of the weekly mean sentiments within the four sets of documents. Note that sentiment scores have been smoothed by locally weighted regression (or LOESS), which allows longer-term sentiment trends to be discerned from the loud and more granular noise that would have otherwise dominated the visualization.

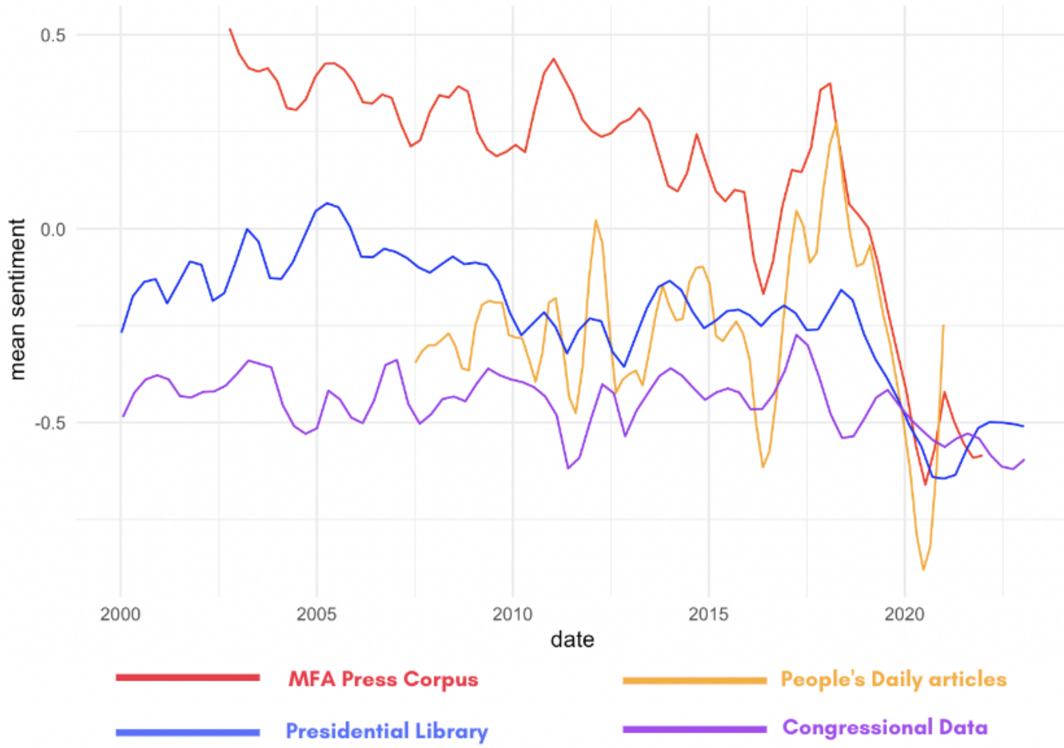

Figure 9: Sentiments of each dataset over time

Alternatively, the sentiment scores can be merged by country, which generates Figure 8. Here, to simultaneously visualize local fluctuations and trends unfolding on a macro-scale, two different span parameters for LOESS smoothing were applied:

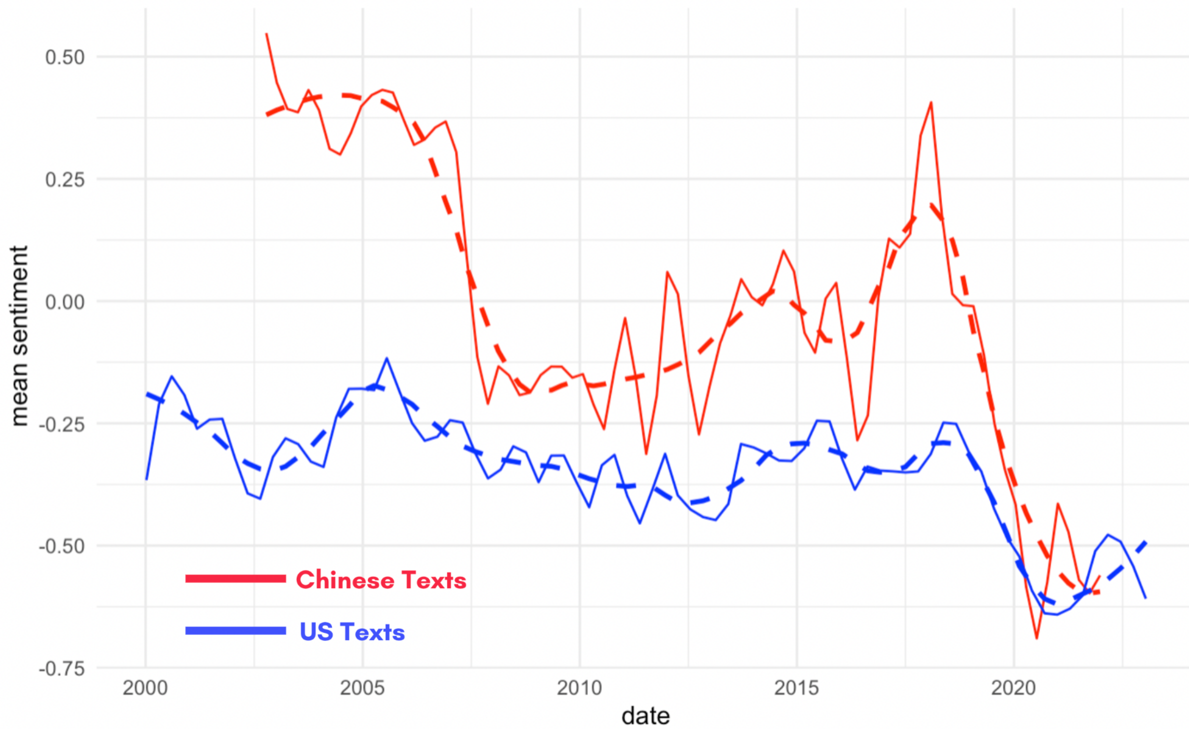

Figure 10: Chinese and US sentiments over time

Several insights can quickly be gleaned from Figures 9 and 10. First, we might note that before Chinese sentiments plunged into a freedive towards unrelenting negativity in early 2018, Beijing's rhetoric vis-à-vis its counterpart was nearly always more positive compared to Washington's attitudes towards Beijing.[36] Moreover, and perhaps not surprisingly, members of Congress have evidently disseminated a steady stream of pessimism with regards to China. Compared to the relative optimism of the MFA press corpus (at least until 2018), discussions of the US in *People's Daily* also seem to be weighed down by significantly more negativity. This is the reason for the dip in Chinese sentiment around 2007, which coincides with the time when state media data became available. The time series corresponding to this state-run newspaper, in

[36] Da Wei, in fact, also identifies 2018 as the "watershed year" in the US-China competition: "A Restructuring International Order and the Paradigm Shift in China–U.S. Relations," *China International Strategy Review* 1, no. 1 (June 1, 2019): 21.

addition, struggles to convey any coherent, meaningful information due to its volatility and the fact that it is missing data prior to 2007.

Analyzing Sentiments with an Economist's Toolkit: Granger Causality

1. The Mechanics of the Granger Test

Since this paper revolves around the central research question of whether one country's discourse takes the role of a sentiment "trend-setter," while its counterpart acts as a "trend-reactor," it may be particularly fruitful to borrow a quantitative tool that is ubiquitous—and hotly contested—in the realm of time series econometrics. Specifically, this paper will apply Granger causality to assess whether the patterns of interaction between the US and Chinese sentiments are statistically significant.

Granger causality is a statistical hypothesis test in time series analysis that postulates whether "lagged" values of one time series can possess predictive power for the values of another time series. The core inquiry of a Granger model is thus: when given time-dependent variables X and Y, can Y be substantially better predicted by a combination of values from X and Y compared to only past values of Y, or vice versa?

Mathematically, this tool can be represented as the comparison of a full bivariate model and a restricted univariate model (although the Granger test is also versatile in that it can be expanded to test a full multivariate model). Say we are interested in measuring whether US sentiment trends can be better forecasted by integrating Chinese discourse, then restricted model for US sentiments is as follows and only performs autoregression upon itself:

$$US\ Senti_t = \sum_{n=1}^{k} a_n \times US\ Senti_{t-n} + e_t$$

Where $US\ Senti_t$ denotes US sentiment towards China at time t, while $US\ Senti_{t-n}$ denotes US sentiment at time t - n. e_t, moreover, is an error term and represents residuals of the model.

The full model, on the other hand, can be written as:

$$US\ Senti_t \ = \ \sum_{n=1}^{k} a_n \ \times US\ Senti_{t-n} + \sum_{n=1}^{k} \beta_n \ \times China\ Senti_{t-n} + \epsilon_t$$

Notice here that we expand the model to also incorporate past values of Chinese sentiments. In this sense, *k*, the "lag amount" of the model, acts as a flexible parameter that can be fine-tuned to govern the scope of the past time frame that the model references. The new error term would then be ϵ_t.

In order to compare the two models, we calculate the following value and perform an F-test on it to check whether it is statistically meaningful:

$$F_{China \to US\ Senti} \ = \ ln\frac{var(e_t)}{var(\epsilon_t)} = \frac{(RSS_{restricted} - RSS_{full})/(r-s)}{RSS_{full}/(T-r)}$$

Where RSS are residual sums of squares, while *r* and *s* are the number of parameters held by the full and restricted models, respectively. *Var*, moreover, denotes variance.

Alternatively, one might employ a Wald test that situates each coefficient β_n within a chi-square distribution to check whether they differ from 0 in any significant fashion. This happens to be the inference method favored by the R time series library that my paper uses.

Because we want to understand the lagged correlation between US and Chinese sentiments towards each other in both directions, it would also clearly be advisable to flip the variables in the model laid out above and assign the Chinese time series as the response.

2. Hot Topic Among Econometricians

Granger causality has gained widespread traction among economic researchers, particularly those who want to find support for hypotheses that suggest causal links between two phenomena. A wealth of literature applies time series analysis to defend or contest the neoliberal wisdom of export-driven economic growth. Mehrara and Firouzjaee, for example, ran an enhanced version of the Granger test to suggest bi-directional causality between exports and growth among developing countries.[37] Konya applied similar methods on OECD states and found varying levels and directions of causality within a bivariate *GDP-exports* model, as well as a multivariate *GDP-exports-openness* model.[38] In many instances time series analyses and regression have given rise to flushed-out policy prescriptions. Ghosh, as well as Lean and Smyth, delved into the relationship between electricity generation and various indicators of national economic performance in India and Malaysia.[39] What they discovered was a distinct lack of "feedback" between their time series, which helps legitimize the argument that energy conservation policies do not necessarily impede growth momentums.

The analytical power of this method, however, is also a fulcrum of heated scholarly contention—specifically because Granger himself and many scholarly works that came after grew increasingly skeptical of the extent to which this test can crack the age-old statistical puzzle of causal inference. The shortfalls of the Granger test to "prove" causality is particularly acute in a bivariate model. As per Shojaie and Fox, bivariate models can be misleading or pick up

[37] M. Mehrara and Bagher Adabi Firouzjaee, "Granger Causality Relationship between Export Growth and GDP Growth in Developing Countries: Panel Cointegration Approach," *International Journal of Humanities and Social Science* 1, no. 16 (2011): 223-231.

[38] László Kónya, "Exports and Growth: Granger Causality Analysis on OECD Countries with a Panel Data Approach," *Economic Modelling* 23, no. 6 (December 1, 2006): 978–92.

[39] Sajal Ghosh, "Electricity Consumption and Economic Growth in India," *Energy Policy* 30, no. 2 (January 2002): 125–29; Hooi Hooi Lean and Russell Smyth, "Multivariate Granger Causality between Electricity Generation, Exports, Prices and GDP in Malaysia," *Energy* 35, no. 9 (September 1, 2010): 3640–48.

spurious correlations when they do not take into account other exogenous variables.[40] John Freeman also points out that the inclusion of another time series could even throw off the correlations within the existing text.[41] This paper would thus heed Freeman's advice: that quantitative modeling should be married to more theoretical or qualitative modes of inquiry. Specifically, this would entail following up the Granger tests with a review of the actual rhetoric within the documents. If I observe similar rhetorical tropes or talking points bouncing back and forth between the US and Chinese sides, then this paper has fairly strong evidence of causality between the two states' discourse sentiment trends. On the other hand, a lack of congruence between US and Chinese discourse at the textual level—even when Granger tests returns statistically significant results—would be more indicative of the fact that one state's sentiment trends can forecast that of the other, but their fluctuations may be animated by some exogenous force.

While the landscape of econometrics teems with the application of the Granger test, the crossroads between time series analysis and NLP is relatively less frequented by researchers. Baumann et. al. made strides in this intersection by considering how emotions of "valence" and "arousal" in Austrian parliamentary records and media databases correspond to shifts in topics prevalent in political discourse.[42] But for the most part, the studies that do use Granger causality on sentiment trends of textual data within the social sciences often do not venture beyond the bounds of economics or financial markets. Several papers have sought to forecast stock prices with Twitter sentiment, and Usher et. al. similarly measured whether social media sentiment

[40] Ali Shojaie and Emily B. Fox, "Granger Causality: A Review and Recent Advances," *Annual Review of Statistics and Its Application* 9, no. 1 (2022): 293.
[41] John R. Freeman, "Granger Causality and the Times Series Analysis of Political Relationships," *American Journal of Political Science* 27, no. 2 (1983): 336-37.
[42] Andreas Baumann et al., "Exploring Causal Relationships Among Emotional and Topical Trajectories in Political Text Data," in *3rd Conference on Language, Data and Knowledge (LDK 2021)* (Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021).

revolving BREXIT exit strategies is linked with the pound-euro exchange rate and the FTSE 100 index.[43]

This means that a study integrating time series analysis (like Granger causality models) with foreign policy discourse, particularly as it pertains to US-China relations, is yet to be found.

3. Granger Causality in Action

A) Examining Test Assumptions

In order to apply Granger tests to my sentiment time series, it is important to ensure that the data fed into the test are not in violation of the stationarity assumption—a prerequisite for any Granger causality output to be interpretable. This means both the US and China sentiment time series should not be defined by statistical properties that are dependent on time, which implies constant mean, constant variance, and constant autocorrelation (i.e. a consistent lag amount that can be used to autoregress current values) over time for both US and Chinese sentiments.

The augmented Dickey-Fuller (ADF) test was applied to both the US and China "parent" datasets—as well as the MFA, *People's Daily*, Presidential, and Congressional datasets—to check for compliance with the stationarity rule. When textual sentiments were averaged over monthly intervals, the time series corresponding to Chinese data were not able to "pass" the ADF test. As such, I applied first-order differencing operations to the monthly Chinese time series so as to ensure its conformity with Granger causality specifications. On the other hand, every source of data grouped by week required no further steps of modification, as they all passed the ADF with flying colors. This makes sense—because some of the longer-term fluctuations in sentiment trends become less pronounced and diluted by noise as we focus in on a shorter time frame.

[43] James Usher, Lucía Morales, and Pierpaolo Dondio, "BREXIT: A Granger Causality of Twitter Political Polarisation on the FTSE 100 Index and the Pound," in *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 2019, 51–54

B) Fine-Tuning the Test

Another core consideration when applying Granger causality is the key "lag amount" parameter that determines the number of past values the model considers to estimate the current value. In order to choose a lag that optimizes model performance, it is possible to examine how different model diagnostic indices of a Granger model—such as the AIC (Akaike Information Criterion) and HQ (Hannen Quinn) values—shift as this parameter is continuously adjusted. Given lag amounts anywhere from 1 to 15 months, lag parameters of 2 or 3 (with units in weeks or months depending on the interval used to group the sentiment data) gave the most outstanding performances.

While lags restricting the model to a narrower time frame can reduce the risk of overfitting and speed up computation, higher lag amounts may allow the Granger models to take into account a wider perspective on historical data. Thus, we can say that the lag parameter tuning is a delicate balancing process that trades bias in lower-order models for the power of higher-ordered ones. For the needs of this paper, I flexibly experiment with lag amounts anywhere between one and five based on the intuitions that 1) these values generate models that are nearly equally favored by authoritative model diagnostics, 2) these values reflect the US and China's tendency to swiftly shoot down the latest round of rhetorical provocations or embrace a gesture of reconciliation, and 3) these values are appropriate for data grouped by weeks, as any higher lag amounts would spill over into the predictive range of a model built on monthly time series. Lag values less than five are also better-suited for pairs of US and China time series grouped by month, as it is not quite plausible for historical sentiment values extending beyond a five-month threshold to contribute additional forecasting power towards the Granger models.

C) Applying the Test on Different Cuts of Data

The central research question laid out in previous sections can be probed from a variety of angles, which prompted me to assemble or cut the sentiment data in different directions. First, we have seen from the parameter experimentation and ADF tests that pairwise Granger models can be applied to both sentiments grouped by weeks and months. Moreover, this paper considers pairwise Granger causality between not only the US and China "parent" datasets, but also between the MFA and Presidential Library datasets. The reasoning here is that these two sources of foreign policy discourse are most representative of the two great powers' prevailing talking points and perceptions towards one another.

Beyond applying Granger causality over different iterations of time series pairs (US and China sentiments, or MFA and President sentiments grouped by week or month) in their entirety, I also considered how the model's performance and significance fluctuates between narrower windows of time. In other words, it is helpful to investigate how the capacity for Chinese discourse to forecast its counterpart's political sentiments towards China, or vice versa, changes when we focus upon particular time frames. To investigate this more granular variation of the original hypothesis, I made two types of "cuts" on my time series: a "backward-dating" cut, which selects windows of monthly or weekly data before a certain date, or a "forward-dating" cut, which extracts portion of the original time series after a set time. These data windows, moreover, were made along a wide range of dates anywhere from 2002 (the starting point of the China time series) to 2023 (the tail of both China and US time series).

Another way to cut the data is to filter datasets by topic. This allows us to partition the highly variegated patterns of foreign policy discourse between the two countries into more manageable segments. For example, rhetoric converging around shared interests in negotiating the Six Parties Agreement, the JCPOA, or ambitious climate deals might not resemble the furious finger-pointing and indignant ripostes that animate US-China discourse around issues like human rights. Following this line of logic, I extended my research question to check if there are any sentiment “trend-setting” and “trend-reacting” relationships that significantly, and surprisingly, deviate from the overarching norm. Each major topic is filtered out with a set of manually-compiled keywords, which are all laid out in Figure 11:

TOPIC	ISSUE	KEYWORDS
1	Trade and Finance	Trade, tariff, protectionist, WTO, import, export, countervailing duty, intellectual property, manufactur(e/er/ing), jobs, exchange rate, currency (manipulation)
2	Chinese Domestic Periphery	Human rights, Xinjiang, Uyghur, Tibet, Hong Kong
3	Taiwan	Taiwan
4	Multilateral Negotiations	Denuclearize(ation), DPRK, North Korea, Iran, JCPOA, IAEA, Paris, climate, COP, ASEAN, APEC, EAS, G20, summit, multilateral
5	South China Sea	South China Sea
6	Global Health and Epidemics	COVID, coronavirus, pandemic, epidemic, outbreak, EBOLA, SARS
7	Russia	Russia

Figure 11: Topic-Keyword Dyads in US-China Discourse

This paper also considered how the count of purely positive and negative documents fluctuated over time. Bivariate Granger tests would thus be applied on truncated sets of data from both American and Chinese sources with sentiments either above or below half a standard deviation of the mean. Rather than constructing time series with sentiments, moreover, I re-oriented the variable of interest to focus on the document count. In other words, these pairwise Granger causality tests would re-formulate the original research question to examine whether the frequency of positive- or negative-sentiment discourse production in the US can forecast, or "granger cause," the frequency with which China disseminates positivity or negativity (or vice versa).

As illustrated by Figure 12, Granger tests conducted throughout the course of this paper are distributed along a fanning, hierarchical tree structure. At the first set of diverging branches, this paper experiments with different modes of cutting the data—by time frame, by topic, or by sentiment polarity. The next level of branches determines whether the "holistic" China and US datasets would be employed, or solely the MFA and Presidential Library corpora. At the root of this tree-like architecture are the options of time series-construction via weekly or monthly intervals, as well as the directionality of the Granger model (these are the full-fledged Granger causality tests that were carried out):

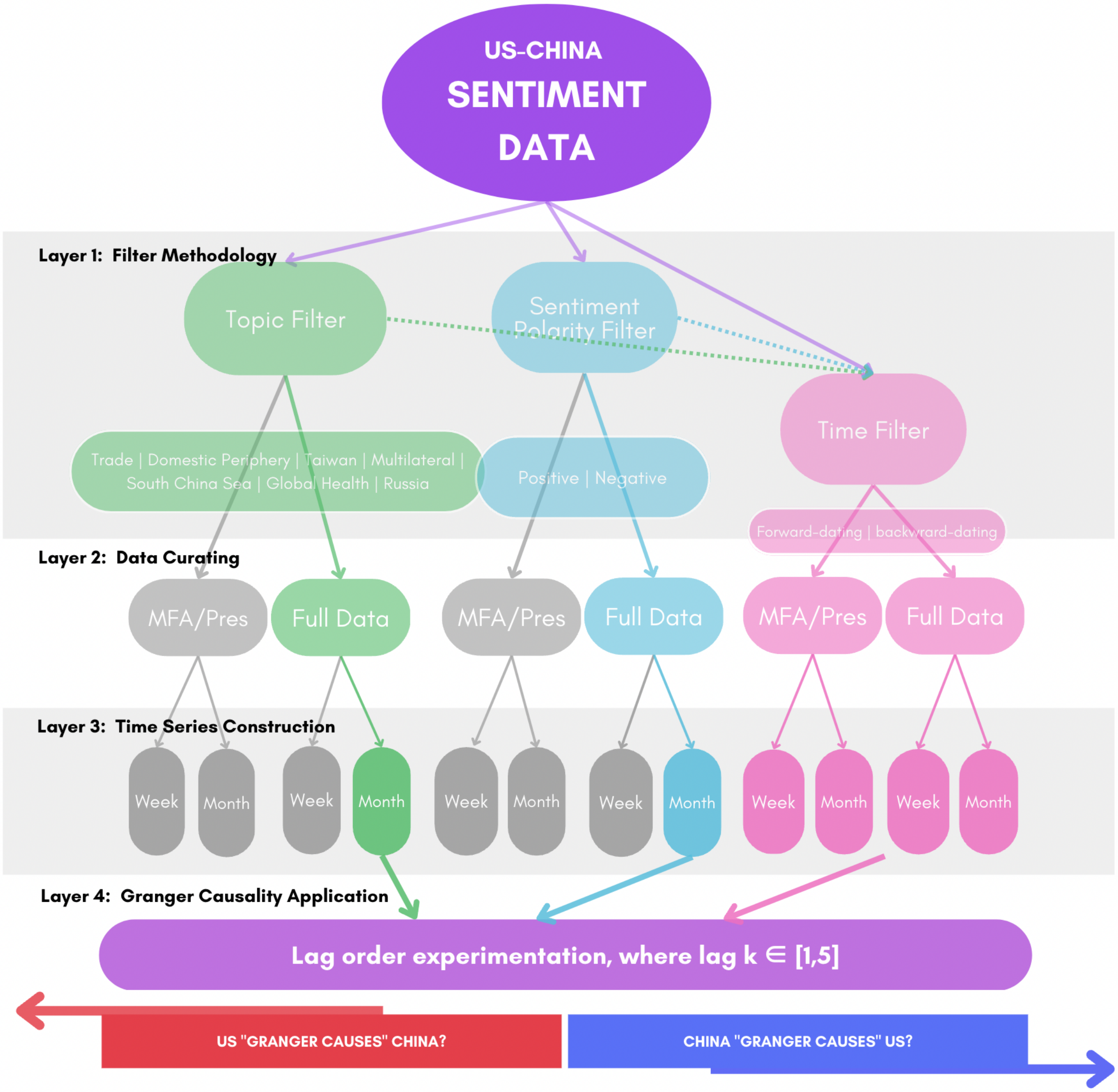

Figure 12: Bidirectional, Bivariate Granger Causality Tree

As observed, it might be important to note that time series derived from latitudinal filters (i.e. topic-specific and sentiment polarity time series) were not grouped by week, as zooming into data that has already been truncated at such a granular time frame creates a good number of missing values, which throws off the Granger model. This is not a relevant consideration for the time filter and its family of Granger tests, however, as longitudinal cutting will not reduce the density of the data and create an overflow of missing values.

The same reasoning also eliminated the “MFA/Pres” branch underneath the topic and the polarity filters, as using anything less than a full dataset would make a substantial dent in the number of available data points at a given point in time.

In certain cases, I also experimented with subsuming the time frame branch “under” the topic or sentiment polarity categories in order to see if, within particular time frames, the topic or sentiment filters generate even more insightful findings.

Beyond these exceptions, every iteration of Granger models along the bidirectional, bivariate time series tree were carried out. The findings that proved fruitful would be explored in the coming section.

What Story Does the Data Tell?: Key Findings

1. **Granger Causality with Specific Time Windows**

By correlating sentiments from the MFA and Presidential Library corpora on a weekly basis, the Granger model returns the most insight-rich, and interpretable, results. The models demonstrated bi-directional causality (i.e. Granger tests running in both directions yielded significant outcomes for the entire sequence of date windows) when I employed a lag amount of two. As the date filter is pushed back, however, the model taking MFA sentiments as a response variable began to inch towards the threshold of statistical insignificance, but the model running in the opposite direction (i.e. taking Presidential sentiments as a response) continued to show sturdy, significant behavior.

By bumping the lag parameter to three, one could observe a set of particularly interesting results, summarized by the Wald test p-values as follows (here, I compare p to an α of 0.05):

	Wald test p-values	

Using data ***AFTER...***	**Chinese MFA as response**	**US Presidential Library as response**
2002-01-01 **(i.e. entire time series)**	**2.41E-04**	**1.09E-04**
2003-01-01	2.39E-04	1.35E-04
2004-01-01	5.17E-04	3.65E-04
2005-01-01	4.94E-04	3.14E-04
2006-01-01	3.63E-04	1.46E-04
2007-01-01	4.01E-03	2.24E-04
2008-01-01	0.02	3.12E-04
2009-01-01	0.02	4.69E-04
2010-01-01	0.04	3.04E-04
2011-01-01	0.11	5.68E-05
2012-01-01	0.12	1.09E-05
2013-01-01	0.07	3.09E-07
2014-01-01	0.14	1.43E-07
2015-01-01	0.30	1.12E-06
2016-01-01	0.11	2.73E-05
2017-01-01	0.07	2.60E-04
2018-01-01	0.02	2.33E-04
2019-01-01	0.08	0.03
2020-01-01	0.98	0.31
*More significant models are coded in green		

Figure 13: Wald Test P-Value Table, Granger Causality with Forward Time Filter using MFA and Presidential Library Data

What becomes evident in Figure 13 is that when we consider data after 2008, then the Granger models begin to exhibit an interesting unidirectional causality whereby Chinese sentiments effectively forecast US sentiments, but *not* vice versa.

In this sense, for the late 2000s and the majority of the 2010s, discourse emanating from the White House about China seems to exhibit sentiments that “react” to the tones of MFA rhetoric towards the US. This phenomenon supports the idea that, for the past decade and a half, the US—or at least its head of state—has taken on the role of a discourse sentiment "trend-reactor," whereas its Chinese MFA interlocutors can be more aptly described as sentiment "trend-setters."

When I incorporated the full US and China datasets (while keeping the lag parameter at three), the unidirectional pattern of significance weakens slightly and becomes interspersed with periods of time where the Granger causality runs in either direction, although models with US data as a response were still consistently more significant:

	Wald test p-values	
Using data *AFTER*...	Chinese *full* data as response	US *full* data as response
2002-01-01 **(i.e. Entire time series)**	**0.05**	**5.20E-05**
2003-01-01	0.05	6.86E-05
2004-01-01	0.06	1.81E-04
2005-01-01	0.05	2.32E-04
2006-01-01	0.02	2.89E-04
2007-01-01	0.03	4.88E-04
2008-01-01	0.05	5.69E-04
2009-01-01	0.03	9.79E-04
2010-01-01	0.01	2.76E-03
2011-01-01	0.01	8.03E-04
2012-01-01	0.01	7.14E-04
2013-01-01	0.03	2.48E-04
2014-01-01	0.23	2.82E-04
2015-01-01	0.05	2.66E-04
2016-01-01	0.17	3.28E-04
2017-01-01	0.03	5.85E-04
2018-01-01	0.04	0.01
2019-01-01	0.06	0.02
2020-01-01	0.53	0.38
*More significant models are coded in green		

Figure 14: Wald Test P-Value Table, Granger Causality with Forward Time Filter using Full Data

Crucially, we must also note that Granger causality tests constructed from backward-dating time windows or from monthly time series are rarely meaningful. The former finding indicates that the predictive power of US and Chinese political sentiments towards each other is much more salient in the latter half of my time series compared to the first half.[44]

The inability of monthly time series to converge into significant results, on the other hand, can be primarily attributed to first-order differencing. The unvarnished, unprocessed monthly time series actually demonstrates bidirectional significance, although these outcomes are clearly not interpretable because they have infringed upon the stationarity assumption. The significance of these models, in addition, quickly dissipates once first-order differencing operations are applied. Feige and Pearce, in fact, disputed historic scholarly arguments that differencing will not disrupt previously existing and delicate causal relationships between two variables. In their own vivid words: "it is distinctly possible that the use of first difference filters … could give rise to the situation in which the baby has been inadvertently thrown out with the bathwater."[45]

2. Granger Causality by Topics

A) How Often is Each Topic Mentioned?

Beyond applying Granger causality on sentiment trends within narrower windows of time, it is also advisable to explore the second "family" of bivariate Granger models, which

[44] Another interesting observation is that while the MFA and the White House—the preeminent foreign policy-making organs of the US and China—are locked in an intimate and dynamic dialogue, Congressional and *People's Daily* texts are unsurprisingly further removed from this exchange. In fact, when Congressional data (once again transformed into a weekly time series) is correlated with Chinese sentiment trends via Granger causality, models from barely any time frames exhibited any sort of statistical significance.

[45] Edgar L. Feige and Douglas K. Pearce, "The Casual Causal Relationship Between Money and Income: Some Caveats for Time Series Analysis," *The Review of Economics and Statistics* 61, no. 4 (1979): 531.

specialize in correlating topic-specific time series. As mentioned, these topic-specific models exclusively employ the full US and China datasets to expand the number of available, post-filter data points.

First, by applying the filters introduced in the topic-keyword dyad table above, it is possible for us to visualize how the relative popularity of each issue area fluctuates over time. Figure 15 illustrates how often references to trade, Chinese domestic periphery, multilateral diplomacy, Taiwan, global health, South China Sea, and Russia surface in US documents over time. Each topic is denoted by a band in a stacked area plot, and the formerly irregular and jagged contours of the area chart have been, once again, smoothed by Loess:

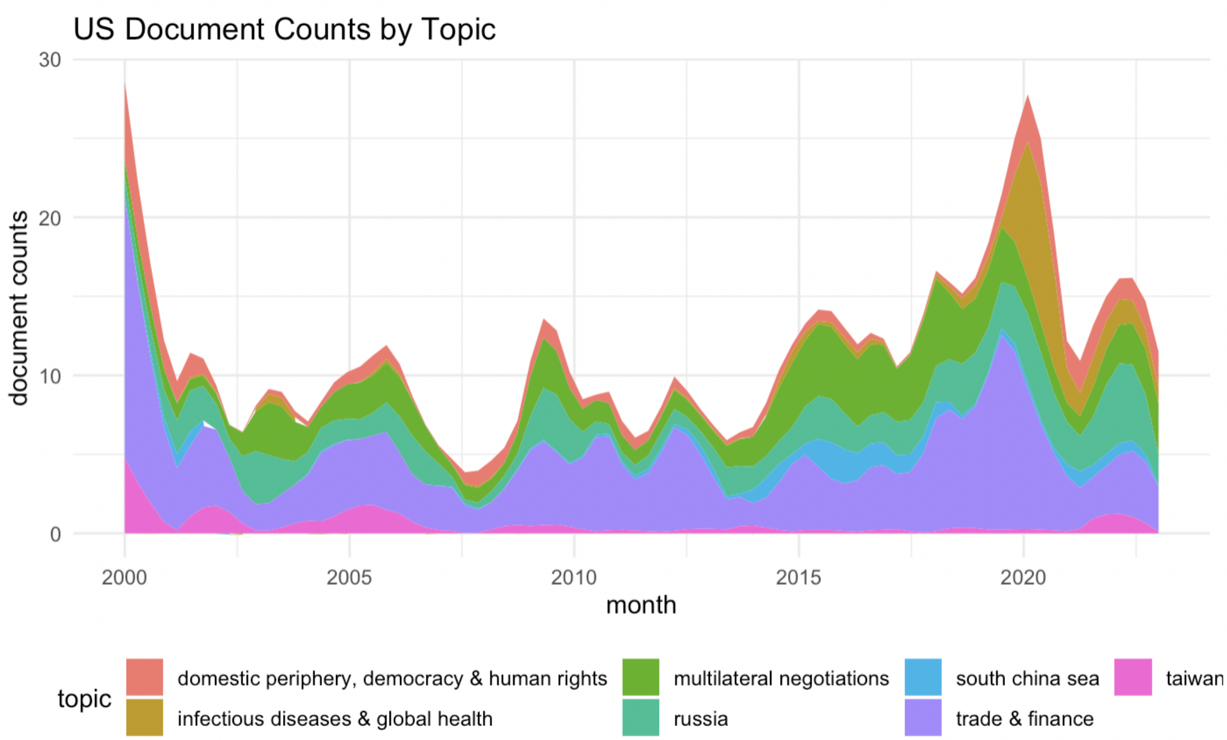

Figure 15: Document Counts by Topic Over Time, US Data[46]

[46] Of course, many documents contain overlaps of two or more topics. These would be double-counted within the area chart, although this would not affect the subsequent analysis. Moreover, topics within the figure are ordered alphabetically, and not by their relative salience.

Similarly, we can also visualize the frequency with which each of these policy themes appear in Chinese discourse vis-à-vis their great power counterpart:

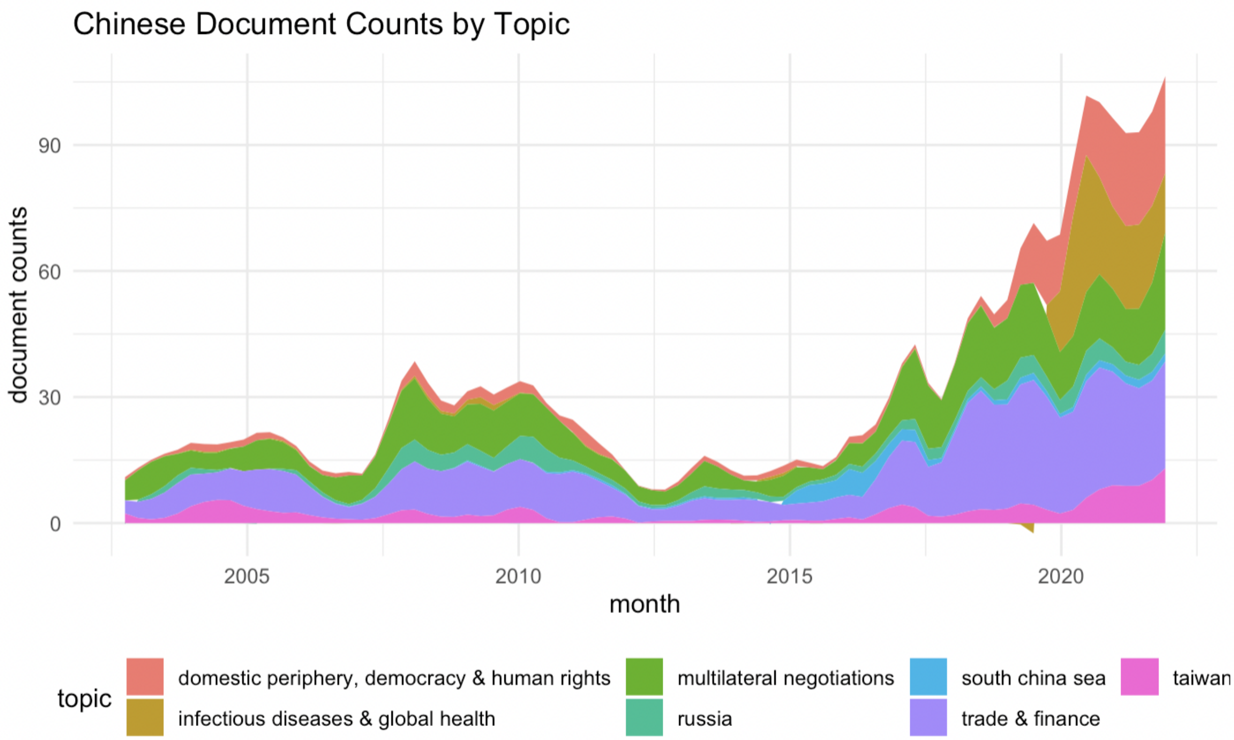

Figure 16: Document Counts by Topic Over Time, China Data

Several key takeaways can be immediately gleaned from these above plots. First, trade and finance has maintained its dominance within government discourse surrounding US-China relations for the past two decades. In the US, we see "bursts" of mentions related to this political motif around 2000, as well as between roughly 2018 and 2020. Both spikes in frequencies make intuitive sense, as the former time period coincides with China's accession to the World Trade Organization, whereas the latter burst signals the onset of the bitter trade war that broke out in early 2018, which featured fiery, mutual mudslinging and tariff-slapping between the two great powers until the Phase I trade deal temporarily put a lid on the pandemonium in January 2020.[47]

[47] For informative overviews and chronologies of the US-China relationship, particularly with regard to trade, see Chad Brown and Melina Kolb, "Trump's Trade War Timeline: An Up-to-Date Guide." *PIIE*. April 16, 2018; Council on Foreign Relations. "Timeline: U.S.-China Relations." Accessed April 17, 2023; Wang Jisi and Hu Ran, "From

While Chinese data does not extend back long enough to capture discourse surrounding the monumental WTO entry, we do observe a similar jump in late 2017 or 2018, as well as around 2016, which may coincide with Trump's relentless, protectionist tirades against unfair Chinese trade practices that he unleashed on his campaign trail, or from the lofty White House bully pulpit.

Multilateralism-themed talking points and conversations, similar to trade and finance, also appear relatively frequently in both countries. We might note with interest that the slight expansion of the multilateralism ribbon in the US area plot could be attributed to the ratification of and controversies enveloping the JCPOA. Based on a qualitative review of the US and China corpora, moreover, the overarching area of concern dominating the multilateralism topic for both countries is by and large North Korea, as well as the successive denuclearization talks that all, without exception, imploded.

Another feature to point out in Figure 16 is the fact that domestic periphery controversies—perhaps the most sensitive issue for Chinese policy-makers—seemed to steal the rhetorical spotlight starting in 2018. This burst in mentions could be driven by a number of factors, ranging from the central government's growing unease at discontent stirring in Hong Kong, to growing international scrutiny around what is happening behind barbed fences and under surveillance footage in Xinjiang. Incidentally, the meteoric surge of domestic periphery issues to the forefront of Chinese discourse matches up with an epic decline in Chinese sentiments vis-à-vis the US.

In fact, we can visualize how the frequency of domestic periphery mentions relative to the total document count in the Chinese dataset relates to the overall dynamics of Chinese

---

Cooperative Partnership to Strategic Competition: A Review of China–U.S. Relations 2009–2019," *China International Strategy Review* 1, no. 1 (June 1, 2019): 1–10.

sentiments over time. Here, the burst of domestic periphery mentions seems to occur in lock-step with overall sentiments taking the plunge:

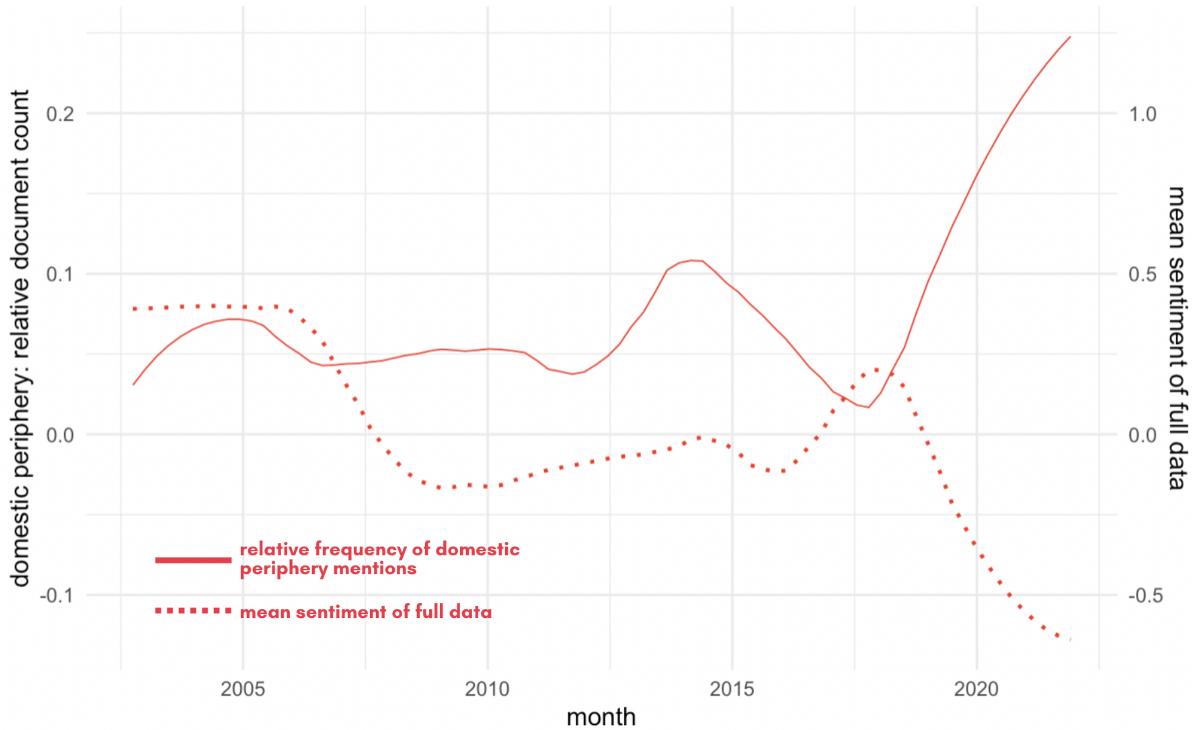

Figure 17: As Domestic Periphery Discourse Soars, Mean Sentiment Plummets

Beyond the ebb and flow of trade and finance, domestic periphery, and multilateralism, we might also note the sudden materialization of the global health band in both figures. This burst coincides with the novel coronavirus pandemic, as expected.

B) What Are the Sentiments of Each Topic?

The Granger causality models, after all, are not fitted with document counts over time, but rather time series tracing out mean sentiments by topic. The figure disaggregating the tone of US discourse by topic is thus presented below:

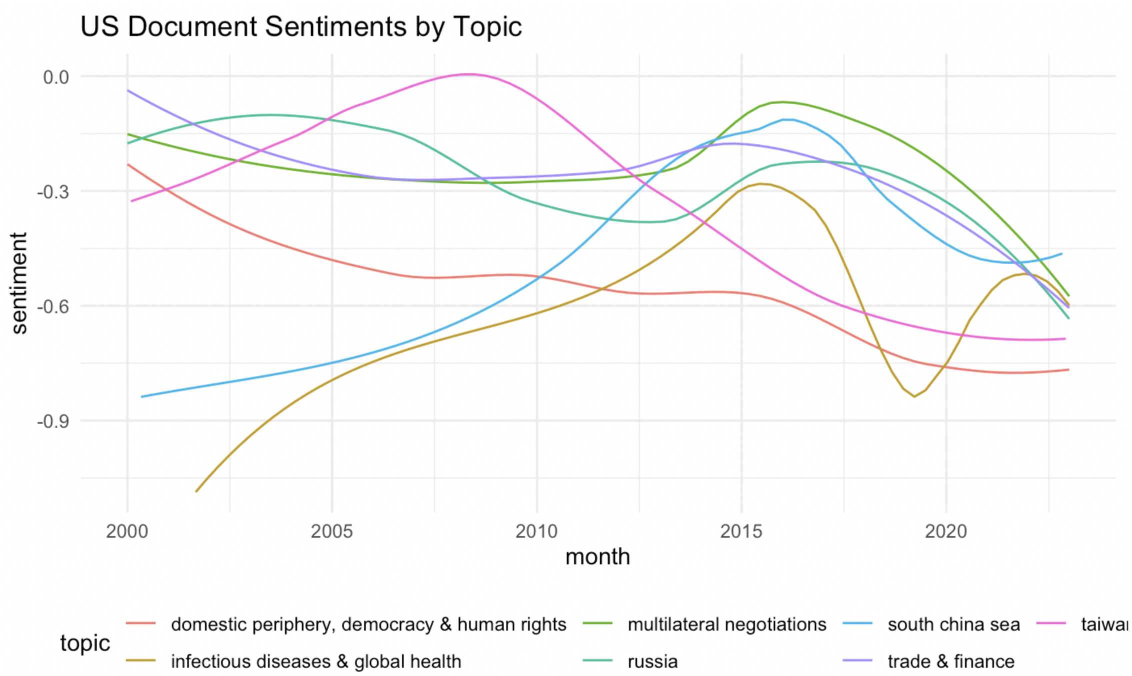

Figure 18: US Document Sentiments by Topic

The fluctuations of sentiment in the Chinese datasets, delineated by topic, are similarly visualized as follows:

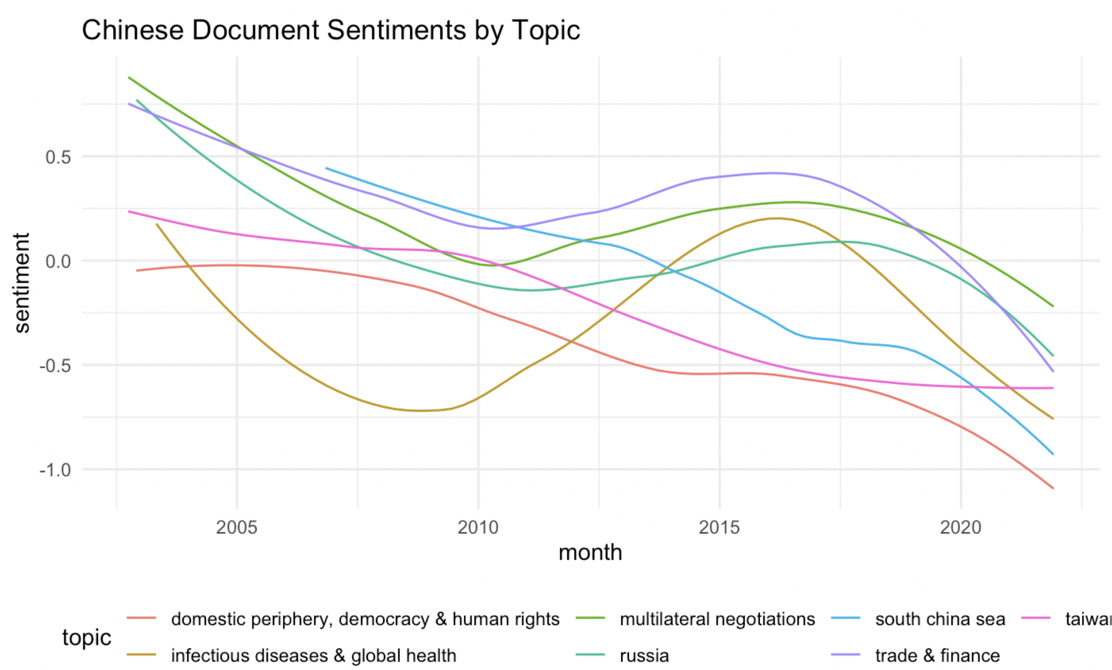

Figure 19: Chinese Document Sentiments by Topic

Once again, a Loess smoother ran through both graphs to even out the noise, although I equipped the smoothing algorithm with a particularly large span parameter in order to make the plots decipherable. In any case, the trends exhibited within both figures seem particularly troubling, as sentiments across the board and for both countries seemed to tumble into precipitous declines within the last quarter of each time series.

Another particularly informative parallel across Figures 18 and 19 is the consistent positivity associated with multilateralism, contrasted with the unrelenting negativity tied to "domestic periphery." This makes sense, as the former represents common grounds around which interests across the Pacific converge, whereas any conversation related to how China governs its deeply contested periphery activates China's political alarm bells and trespasses upon its inviolable "red line" of national sovereignty.[48] The US, on the other hand, looks on at how China is treating its formerly autonomous holdings—and how it is allegedly trampling the sacred values of "democracy" and "human rights"—with increasing disdain. When both sides feel that their counterparts have committed shocking transgressions upon their own core values, it makes sense that mutual perceptions would tank.

We might also observe that while sentiments within the Chinese data appear more harmonized across topics, the lines making up the US time series follow diverging trajectories that are entangled into hardly-interpretable knots prior to 2010, though sentiments across topics seem to have converged in recent years.

It is crucial to note that because these sentiment time series have been thoroughly smoothed, much of their nuances and meaningful fluctuations may have been glossed over. As

[48] A very clear and helpful explanation for China's sensitivity around issues related to territorial integrity is Andrew J. Nathan and Andrew Scobell, *China's Search for Security* (Columbia University Press, 2015), 193-200.

such, beyond simply interpreting the approximated curves of every topic-specific time series at once, it is advisable to isolate the time series of each topic in more detailed pairwise comparisons, while also lifting the strict smoothing parameters in order to capture more of the data's true variability.

As an example, Figure 20 plots the US and China's sentiments vis-à-vis trade next to each other using two different Loess smoothers (the dotted lines are produced with the same spans as those in the line plots that visualize all topic-specific time series).

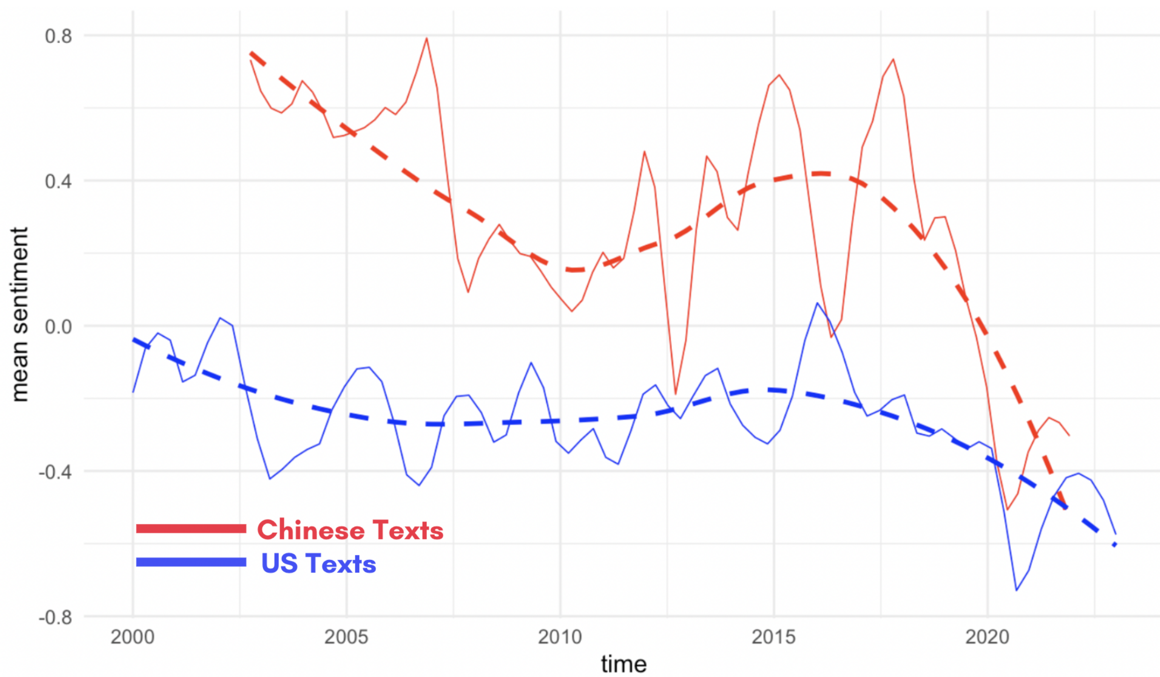

Figure 20: US and Chinese Sentiments Towards Trade & Finance Over Time

As we can see, the two sentiment trends exhibit some visible correlation with one another, but it is perhaps also striking that prior to the late 2010s when both countries' sentiments towards trade plunged towards a trough, there was a wide rift in the emotional polarity of US and Chinese discourse vis-à-vis all things trade and finance. This makes sense, as Beijing's highly effective economic growth strategy was buoyed by its exports and trade surpluses, whereas the

US's attitude towards trade with China became increasingly tainted with trepidation about the erosion of its own manufacturing capacity, as well as China's unfair IP, market access, and forced technology transfer practices.

We can observe a narrower, but nevertheless substantial gap in sentiment when it comes to discourse surrounding multilateralism, though sentiments in this arena have gradually converged over time. Similar to trade, moreover, the following pair of time series also fluctuate in sync.

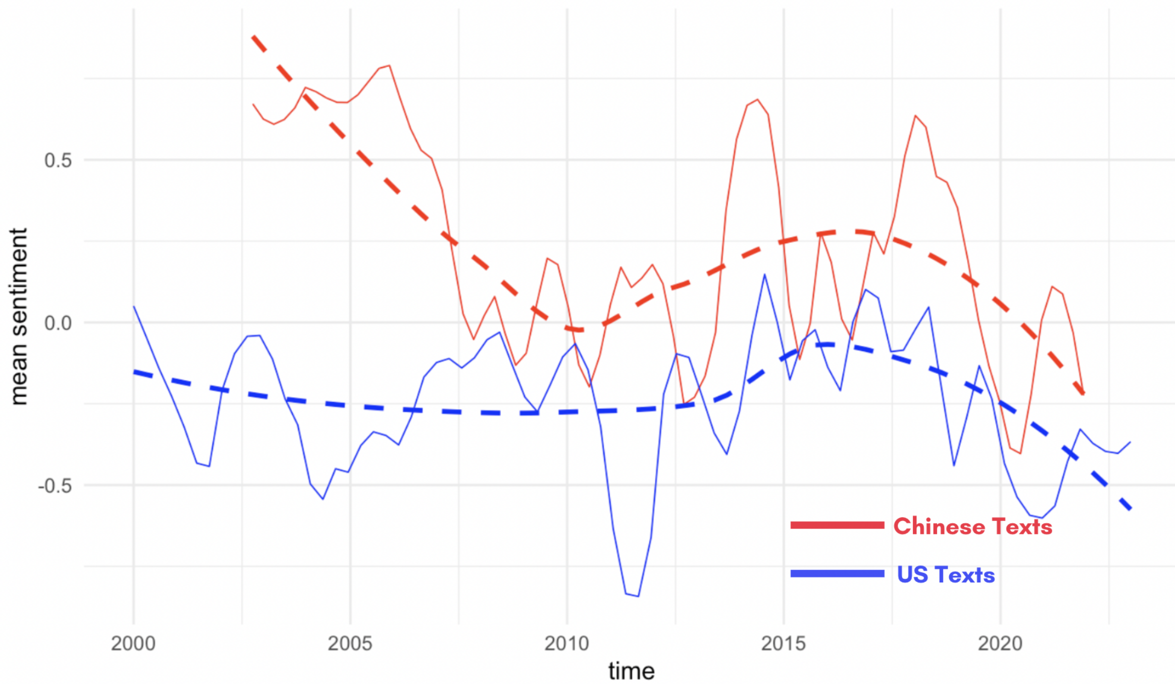

Figure 21: US and Chinese Sentiments Towards Multilateralism

Within both Figures 20 and 21, one might remark that Chinese sentiments took a steep dive off a cliff around 2007. This trend might actually not be due to any organic changes in Chinese discourse; rather, it should be attributed to the influx of *People's Daily* articles, which are generally less positive in their tones compared to the MFA corpus.

The differences in the US and China's sentiments towards the topics above is juxtaposed by the relative congruence of both countries' sentiments towards the Taiwan issue. Although with regards to Taiwan, the sentiments in bilateral discourse appear to be marred by much more negativity than in trade or multilateral negotiations, particularly in the latter half of the time series. This is not surprising, since the precarious and bitterly contested status of Taiwan constitutes the "original sin" of US-China discord. The prickly rhetorical standoff in the region, increasingly accompanied by armed muscle-flexing, seems to be a thorn in bilateral relations that cannot be excised without bloodshed.[49]

Here, it is important to make the distinction between US and Chinese discourse surrounding the question of Taiwan within the context of bilateral relations, and how the two countries perceive Taiwan per se. My data seeks to capture the former.

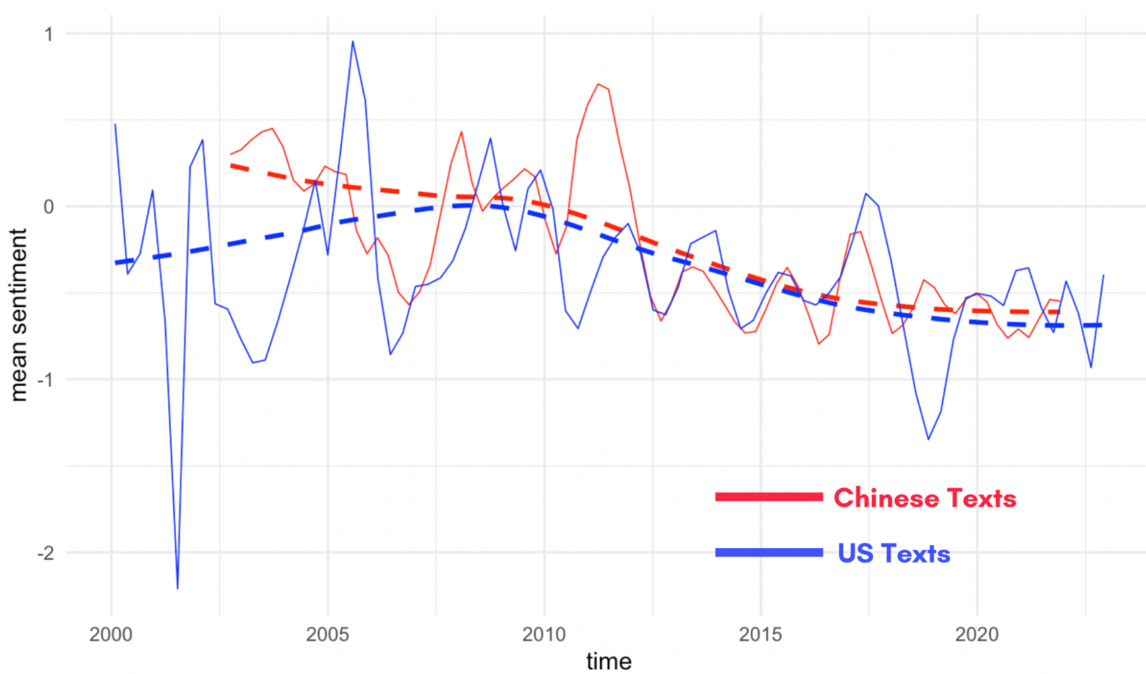

Figure 22: US and Chinese Sentiments Towards the Taiwan Issue Over Time

[49] For a detailed overview of cross-strait relations in the context of US-China diplomacy, as well as an exploration of how the precarious *status quo* came to be, see Steven M. Goldstein, *China and Taiwan* (John Wiley & Sons, 2015).

Similar to the pair of Taiwan time series, the US and China's conversations surrounding how China governs—and allegedly represses—its domestic periphery are also weighed down by unremitting negativity, which are even more pronounced than the two powers' sentiments towards Taiwan.

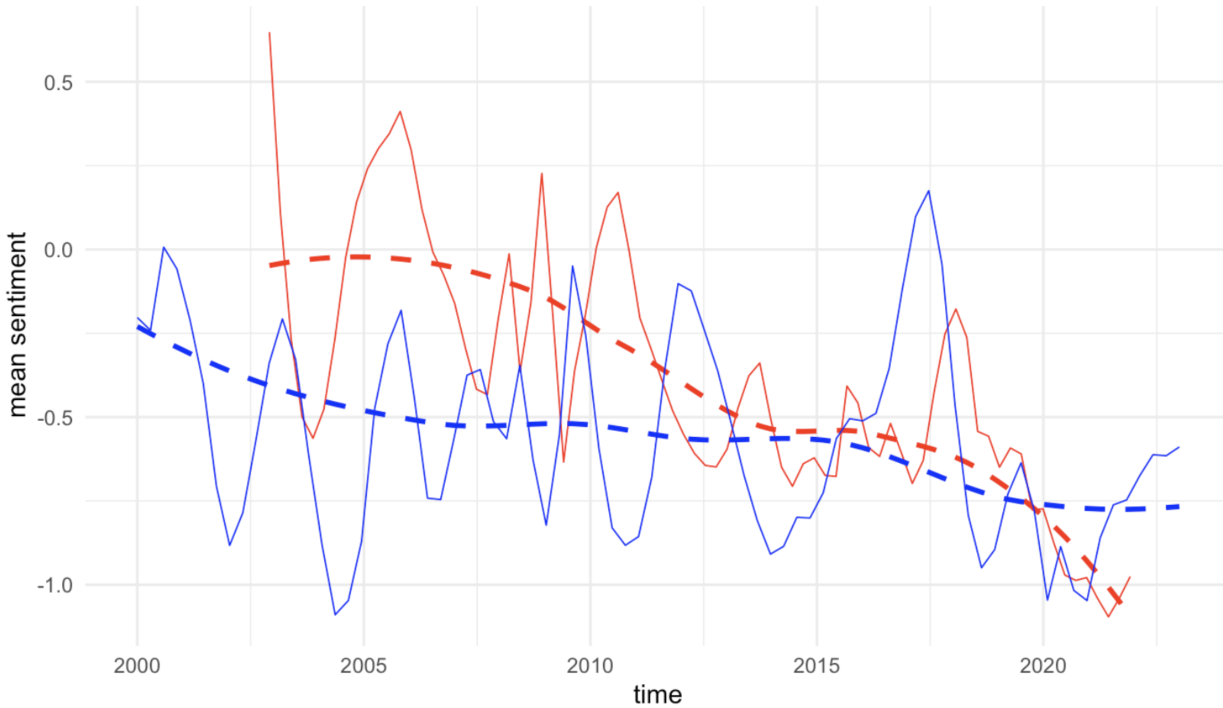

Figure 23: US and Chinese Sentiments Towards Domestic Periphery Over Time

What might be particularly noteworthy in the figure above is the sheer variability of both time series, particularly prior to 2015. This is not indicative of American and Chinese policymakers' unpredictable mood swings in the public arena, but instead reflects the relative sparsity of documents mentioning this topic.

The pessimism gripping the Taiwan and domestic periphery time series, particularly the ones tracing Chinese sentiments, reinforces our understanding that sovereignty and territorial integrity—as it applies to the country's vast Western frontier, its possession of Hong Kong, or its claim on Taiwan—constitute China's most sensitive and inviolable bottom line. Indeed, any attempt to breach China's defensive rhetorical posture on these fronts will predictably invite a lashback full of fire and fury from formerly cordial diplomats.

C) Granger Causality

Having visualized tonal trends disaggregated by issue area, this paper fitted each pair of topic-based time series with the Granger model. The Wald test results for each model are as follows:

USING FULL DATA	**Wald test p-values**	
	US Data as Response	**Chinese Data as Response**
Domestic Periphery	0.04336992	0.28190568
Multilateral Negotiations	0.2251556	0.7146137
Trade and Finance	0.04005604	0.51330884
Taiwan	0.03311907	0.80575678
South China Sea	0.3977521	0.6437452
Global Health	0.5635938	0.8228761
Russia	0.5350707	0.2429181

Figure 24: Wald Test P-Values, Granger Test by Topic with Full Data[50]

Following the patterns of interaction between overall US and Chinese sentiment time series, the US appears to once again be the "reactive" party across all the topics that returned significant models—specifically, Chinese domestic periphery issues, trade and finance, as well as Taiwan. Meanwhile, China's discourse functions more as a "trend-setter," who can effectively help forecast the foreign policy sentiment of its interlocutor across the Pacific. In this sense, the one-directional significance suggests a lagged correlation between US and Chinese sentiments vis-à-vis the three topics mentioned above, whereby the former time series trail behind the latter.

[50] When conducting Granger tests for monthly data filtered by topic, all sub-data sets generated showed significance for—or were on the threshold of showing significance for—the ADF test (i.e. they passed for all intents and purposes). As such, no differencing mechanisms were deemed necessary to prepare the data for Granger causality.

This phenomenon, however, partially flipped on its head when I once again took forward-dating cuts on the data. When only data from 2006 or later were considered, and when this time filter was gradually moved up, the Granger causality direction of the domestic periphery model flipped. This finding indicates that in these latter periods, it was China, instead of the US, that had become the sentiment "trend-reactor."

For instance, the meaningfulness of models fitting time series only from 2015 or up are as follows:

USING DATA AFTER 2015	**Wald test p-values**	
	US Data as Response	**Chinese Data as Response**
Domestic Periphery	0.3776805	0.01933579
Multilateral Negotiations	0.4900904	0.1644173
Trade and Finance	0.0187027	0.136422
Taiwan	0.97406771	0.06629449
South China Sea	0.5834748	0.3797923
Global Health	0.8848808	0.7980965
Russia	0.636824252	0.004278079

Figure 25: Wald Test P-Values, Granger Test by Topic with Data After 2015

While the trade-related Granger model held steady with US data as a response, the domestic periphery model, as mentioned, flipped directions. The Taiwan model, on the other hand, fell into an awkward middle ground whereby it was no longer significant with US sentiments as a response variable, but it inches towards statistical significance with China data as response. Interestingly, for data after 2015, we see a high degree of model significance within the Russia topic flowing in the direction from US data as an explanatory time series, and China data as response.

In a word, this extensive set of topic-based Granger causality tests demonstrates that US sentiments consistently and reliably reacts to those of China in the domain of trade. When it comes to discourse surrounding China's treatment of its domestic periphery, which also lies adjacent to the sensitive issues of democracy and human rights, then the Chinese government takes on the "trend-reactor" role, particularly anytime after 2006.

One consideration that is important to note here, however, is that the significance of the models is *not* simply a surefire measurement of the extent of interaction between the two countries' discourse. Rather, it could be tugged in different directions, or its reliability could be compromised, by the number of data points it can work with. For instance, because the South China Sea and the Taiwan time series are inundated with missing values, it becomes hard for Granger models to harness past values as covariates.

The application of topic filters offers powerful insights into how my research question can be understood, and ways in which the present research can be enhanced. Notably, time series generated by topic can highlight the salient correlation between the two countries' discursive behavior vis-à-vis domestic periphery issues and trade. Granger causality tests of these two pairs of time series also highlight the fact that the direction of such correlations may run in different directions (i.e. whereas US sentiments may be more "reactive" in the realm of trade, they are more "proactive" in Chinese domestic periphery, particularly at later dates). But beyond offering clear-cut answers, the methodology of topic-specific time series analysis prompts us to ask more piercing questions; for instance, can the US and China datasets be expanded to construct higher-resolution, higher-density, and more representative topic-specific models? And would these models demonstrate more sturdy meaningfulness for discourse surrounding issues like Taiwan or Russia?

3. Granger Causality by Valence

The third way to filter my time series data is to extract values that lie on the two ends of the sentiment spectrum. Then, it is possible to construct time series with the frequency of documents classified exclusively as positive, or negative, in tone.

For instance, by cutting both China and US sentiment data at 0.1 standard deviations above the mean, we can visualize the count of relatively positive documents as a proportion of the total size of either country's corpus over time:

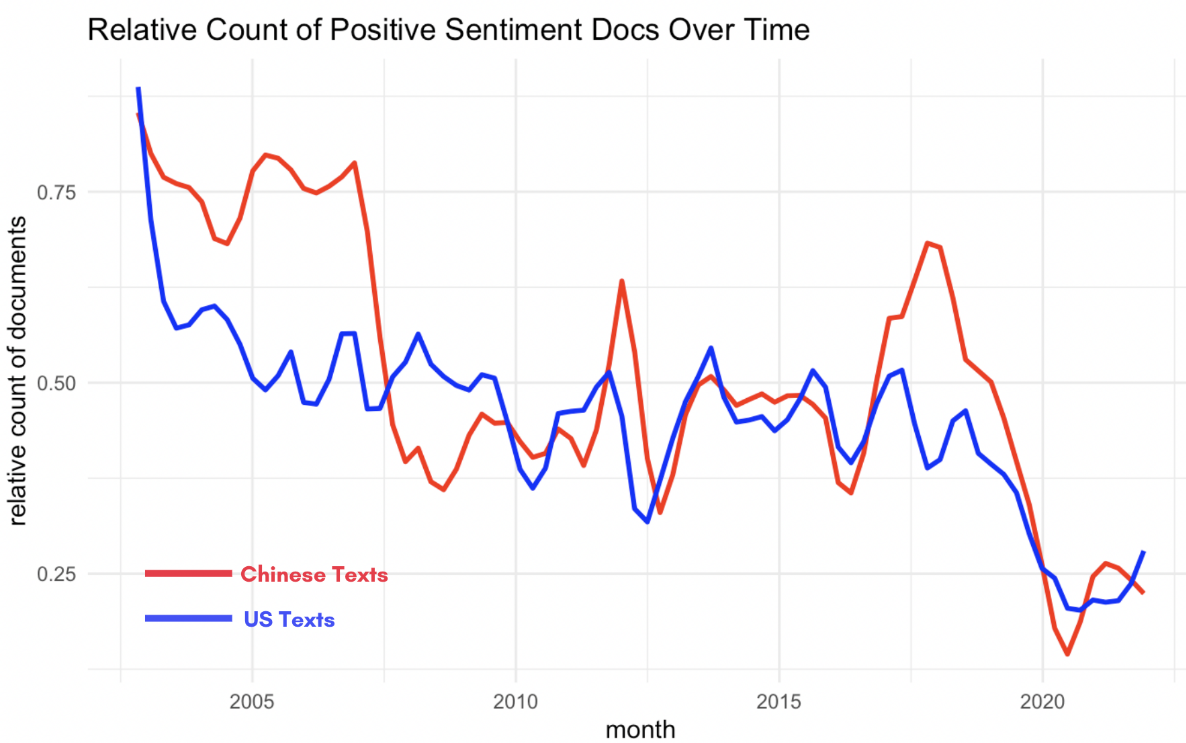

Figure 26: Relative Count of Positive Documents Over Time

The steep decline in documents that express positive emotions—particularly in the post-2018 era—lines up with the salient trends in the overall time series, although the dwindling positivity of Chinese foreign policy discourse is preceded by a "bump" of optimism around 2018 that also defined the contours of the overall time series. It would also be possible to apply

Granger tests to this data (note that instead of using relative counts, the Granger tests took in absolute counts of positive sentiment documents), although no statistically significant results emerged from a lag parameter of two. This insignificance persisted when the lag was adjusted between one and five, or when the direction of the model was flipped. In this sense, we might say that the lagged correlation between US and Chinese positive document counts is flimsy at best, or perhaps non-existent, and that the pair of time series tracing the frequency of positive tones are ineffective predictors of each other.

On the other hand, the behavior of the negative sentiment time series was much more interesting. The figure below traces the relative document counts of sentiments dipping below a tenth of a standard deviation of the mean:

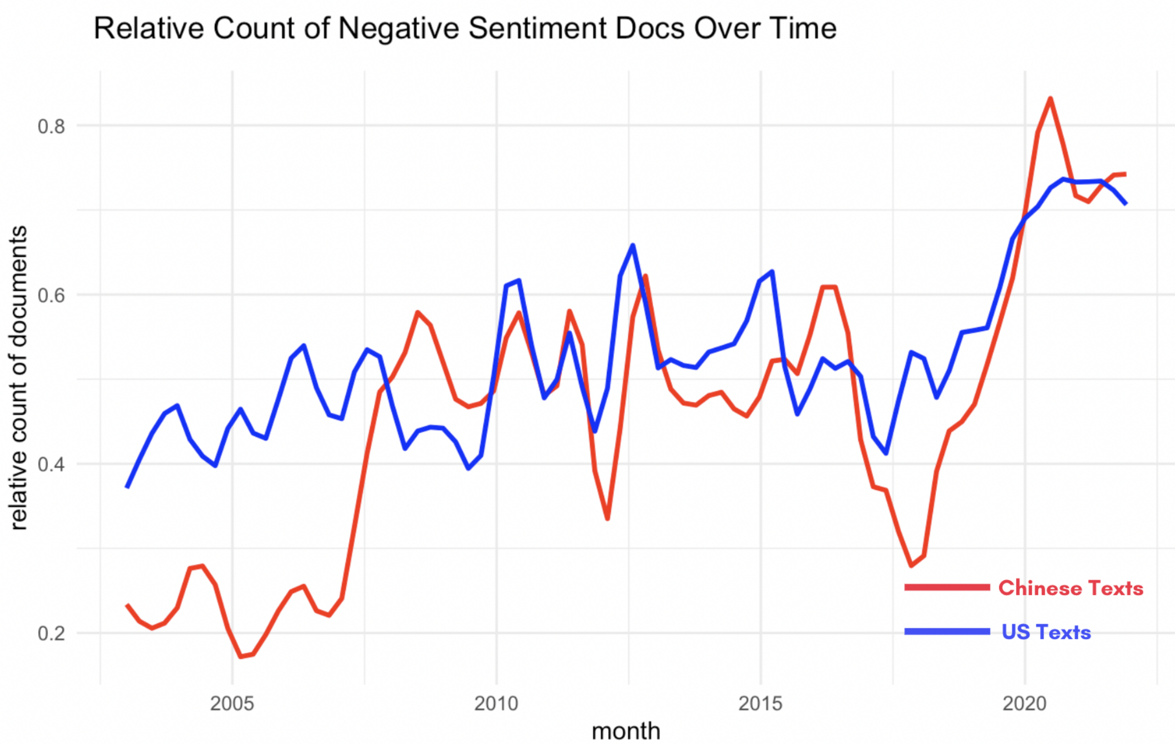

Figure 27: Relative Counts of Negative Documents Over Time

This visualization once again corroborates the trends we witnessed in the overall time series—specifically, one might note that negativity has skyrocketed within the past five years.

This epic deterioration in bilateral ties and the toxic rhetoric it spawns, however, is preceded by a burst of positivity from the Chinese side in around 2017, which is captured by the local trough of the red time series.

Compared to the positive sentiment document counts, moreover, these time series showed much more promise when they were fitted with Granger models.

Indeed, when Granger causality tests were applied to monthly time series, which have been reshaped by differencing to fit the stationarity assumption, *only* the model with Chinese data as a response was meaningful (Wald p-value = 0.01364, compared to p = 0.8161 when US was assigned as response within a model with lag parameter 2). This outcome, moreover, persisted when the lag value was continuously adjusted, when the threshold for negativity was adjusted between a factor of 0.05 and 0.3 of a standard deviation below mean sentiment, and when a forward-dating scheme was applied to only test for more and more recent segments of the time series.

When the Granger causality results for positive and negative sentiment data are contrasted, it becomes quite apparent that any sort of correlation between US and Chinese foreign policy discourse towards each other can be attributed to fluctuations of the extent to which both sides express *negativity*, as opposed to positivity. In other words, whereas US and Chinese foreign policy-makers tend to spew rhetoric marked by negative tones in sync with one another, the patterns of interaction in how they communicate more optimistic rhetoric are less congruent and more ambiguous. Possibly, we can interpret this phenomenon as the two great powers' sensitivity towards and proclivity to react to aggressive discursive posturing.

4. Overview of Meaningful Findings

Upon carrying out all three "families" of Granger causality tests, it is also possible to contextualize the meaningfulness of all results within the Granger model tree illustrated in Figure 12. Within this flowchart, all the Granger test lineages that produced significant p-values are highlighted, and the direction of the Granger causality are indicated at the root-level of the tree:

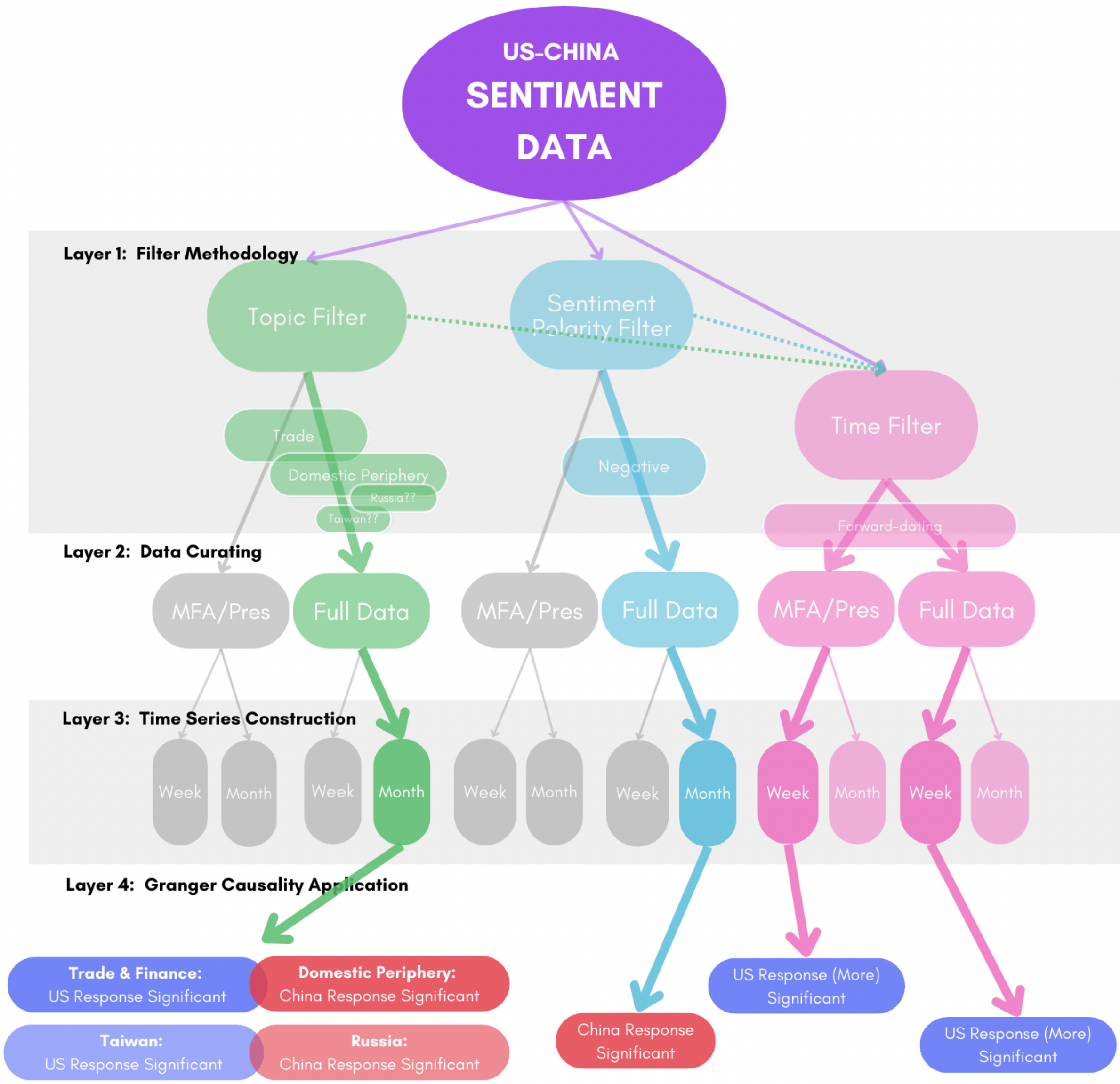

Figure 28: Bivariate Granger Causality Tree with Summary of Findings[51]

[51] Note in this figure that the Taiwan and Russia models under the "topic filter" lineage are purposefully shaded with lighter colors. This is because their significance is not stable when the model time frame is tweaked.

In sum, when it comes to Granger tests based on sentiment polarity (rather than on document counts), US data appears to be meaningfully *reactive* on many more fronts than Chinese data, particularly when the models only consider more recent time periods. The natural implication here is that this paper's quantitative methods provide strong evidence of Chinese political discourse's proclivity to assume "trend-setting" behavior, which is contrasted by the US's tendency to react to its counterpart.

Two phenomena, however, deviate from this more general conclusion. First, we might note that more often than not, Chinese sentiments vis-à-vis its highly controversial governance (and tainted human rights record) in Xinjiang, Tibet, and Hong Kong tend to be *reactive* to those of the US. This finding can be synthesized with the fact that in Figure 25, Chinese sentiments vis-à-vis the Taiwan question is also inching towards significant reactivity. Placed within the context of China's hyper-sensitivity towards any breach of its territorial integrity, the patterns of interaction within these time series may point towards the CCP's inclination to demonstrate virulent defensiveness, but not assertiveness and proactivity, when it detects foreign rhetorical forays into its far-flung domestic periphery, or its inviolable territorial claims.

The second anomaly in my overall findings is that the frequency with which China broadcasts negative discourse is also reactive. The significance of the Granger model produced by a negative polarity filter is particularly interesting—because its counterpart with a positive polarity filter produced no meaningful results. Once again, this piece of finding is highly indicative of the fact that expressions of *negativity* between the US and China are much more highly correlated than expressions of *positivity.*

Discussion of Key Findings

Although the Granger models helped create a more coherent, and interesting, narrative from a crowded and jumbled set of sentiment polarity scores, these meaningful p-values must be married to some qualitative evaluation. Indeed, as mentioned in the Granger methods section, Granger causality can lend credence to the forecasting or predictive powers of a time series for another, but it cannot tell us anything about the causal relationship between the two variables. It is therefore advisable at this point to probe beyond the numbers in order to map the fluctuations of the sentiment time series with real, textual features of foreign policy discourse.

1. Mapping the Topography of Sentiment Time Series

Of course, it would be difficult to survey every salient textual feature across two decades of US-China foreign policy discourse, but it would be useful to review the texts that form the most glaring topographic features of the time series. Figure 29 maps significant events in the history of US-China relations that coincide with the prominent topographic features in each of the Presidential, Congressional, MFA, and *People's Daily* datasets. In addition, it also situates all four time series in the context of four US presidential administrations and two generations of Chinese leadership.

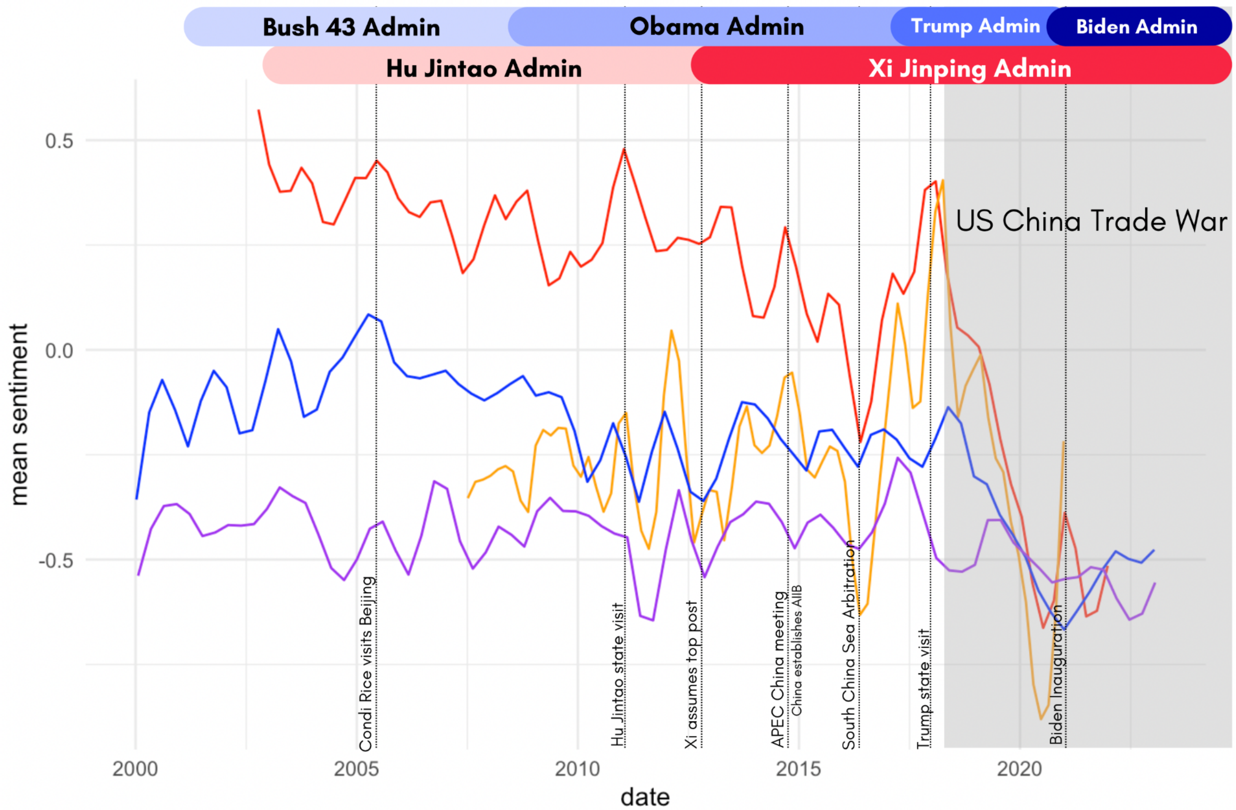

Figure 29: Mapping Significant Events Along US-China Sentiment Time Series

If US texts directly acknowledge or even respond to their Chinese counterparts, or vice versa, then we have fairly strong evidence of *causal* linkage between US and Chinese discourse.

Perhaps the most acute feature dominating the bilateral discursive landscape—and threading through the entire course of my quantitative analysis—is the theme park rollercoaster behavior of US and Chinese sentiments from 2017 to 2018. Indeed, foreign policy sentiments in both countries, and in particular China, climbed to local peaks in 2017 only to come crashing down at unrelenting speeds the next year. Professor Evan Medeiros terms this topographical marvel the "Trump bump." When former President Trump was swooped into the top post, vowing to break from the traditional, establishment Washington playbook, the Chinese government perceived an opportunity to woo, or even co-opt, this political newcomer for its own benefit. However, after Trump unveiled his harsh tariffs in early 2018 in spite of China's soaring rhetorical gestures, the Chinese took a hard pivot in their discourse vis-à-vis the US, switching out tentative conciliatory outreach for virulent finger-pointing. Subsequently, all hell broke loose across the Pacific.

Indeed, this dramatic episode, whereby rhetorical rapprochement was followed by an all-out war of words, enjoys a rich presence in my datasets. MFA spokesperson Hua Chunying, for instance, stacked one cordial talking point on top of another in her press briefing on November 7, 2017, apparently rolling out a rhetorical red carpet ahead of Trump's anticipated state visit to Beijing. The following is an excerpt of Hua's statement:

> The Chinese side is willing to work with the United States to act upon the important consensus reached by the two heads of state to focus on cooperation and properly manage differences in the spirit of mutual respect and mutual benefit so as to ensure the sustained, sound and steady growth of China-US relations. This not only serves the fundamental interests of the two peoples but also meets the shared expectations of countries around the world and the international community.

Showered with the pomp and circumstance, as well as the full, luxurious trappings of Chinese state hospitality, the Trump team reciprocated with positive rhetorical gestures of their own two days later:

> To both the American delegation and to the Chinese business representatives here, your discussions greatly strengthen our partnership and provide a critical bridge between our business community and yours … Both the United States and China will have a more prosperous future if we can achieve a level economic playing field.

In fact, the President also inserted his own signature, colloquial lexicon into the record, noting that there is "great chemistry" between the two sides, and that "we're going to do tremendous things for China and the United States."[52]

At this point, sentiments from both sides were climbing to a warm and optimistic peak. Even as the Trump White House aired its grievances about issues such as the gaping trade deficit, it was quick to assign the blame to its American predecessors while absolving the Chinese of any apparent responsibility. A speech on March 5, 2018, for example, complained

---

[52] On the temporary warming of relations between Trump and Xi, see "Trump China Visit: US Leader Strikes Warmer Tone with Xi Jinping," *BBC News*, November 9, 2017, sec. China.

that “We lost $500 billion [to China]. How previous Presidents allowed that to happen is disgraceful.” About a month later, the President elaborated upon this point:

> I have great respect for President Xi. Two of the most incredible days of my life were spent in China … But … they've created a trade deficit, and I really blame our representatives and, frankly, our preceding Presidents for this … We'll be negotiating with China. Again, our relationship is very good with China, and we intend to keep it that way. … We intend to get along with China, but we have to do something very substantial about the trade deficit.

However, amidst these rhetorical niceties, the Office of the US Trade Representative had been scrutinizing China’s trade practices at the President’s behest and promulgated their findings in a virulent, bombshell Section 301 investigation report in late March of 2018, which accused China of forced technology transfer practices, cyber-enabled IP theft, discriminatory licensing, as well as acquisitions of US assets masterminded by the state.[53] After receiving news of this report, which seemed to be a harbinger for a colossal wall of tariffs on billions of dollars’ worth of Chinese exports, the Chinese snapped back with heated indignation:

> Certain American official's remarks that "the era of economic surrender is over" sounds quite awkward and it seems that they were trying to shift the blame. In fact, it would be more appropriate to say that it's time for the US economic intimidation and hegemony to end. The US side should be clear that what international trade needs in the 21st century is rules rather than power.

The Trump presidency in turn dropped its cooperative guise and put its counterpart on full rhetorical blast. In late July, the President alleged that China’s “very abusive trade practices” and its IP theft propelled it to “go after our soybean farmers” and to “make $500 billion … off the big fat sloppy US.” Another Trump quote from two months later makes clear that any remnants of the veil of cordiality had been torn and torched:

> China is engaged in numerous unfair policies and practices relating to United States technology and intellectual property, such as forcing United States companies to transfer technology to Chinese counterparts. For months we have urged China to change these unfair practices and give fair and reciprocal treatment to American companies. But so far China has been unwilling to change its practices.

[53] Nina M. Hart and Brandon J. Murrill, “Section 301 Tariffs on Goods from China: International and Domestic Legal Challenges,” *Congressional Research Service*, April 5, 2022.

At this point, what had been a promising series of dialogue had collapsed into a grim spiral of finger-pointing, indignant threats, and protectionist outbursts.

Parsing the texts making up this particular topographical feature renders two pieces of insight salient. First, we get qualitative corroboration that the US, at least from late 2017 to 2018, does seem to take on a more "reactive" role in the discursive relationship, whereas China more often sits in the driver's seat. Indeed, studying both sets of data at a micro-level suggests that China—to some extent—does initiate the tentatively cordial gestures as well as the virulent insult-hurling.

But this qualitative survey also demonstrates that a causal link between the two sentiment trends is ambiguous, or perhaps even elusive. Although the subject matter of the two governments' discourse often overlapped, an unequivocal indication that statements from the US dataset were directly replying to their Chinese counterparts (or vice versa) was lacking. Other segments of the time series dyad, moreover, offered even less topical or thematic overlap, and the portfolio of topics and events that made up the two datasets would be situated within non-intersecting discursive planes.

In the absence of any unambiguous causal linkage, how do we interpret the meaningfulness of Granger causality models? To pose this question in more qualitative terms, how can we situate the formation, promulgation, and interaction of two distinct strains of foreign policy discourse within a broader geopolitical canvas dominated by the clash of great powers?

2. What Are Ways to Explain Foreign Policy Discourse?

A) Discourse as a Reflection of Reality

The most straightforward interpretation of foreign policy discourse for the purposes of this paper, of course, is simply that rhetoric is a good proxy for how policymakers from one country perceive their counterparts, and for the actual warmth of bilateral ties. This explanation is informed by Biwu Zhang's *Chinese Perceptions of the US*, as well as David Shambaugh's *Beautiful Imperialist*, which both enlist elite discourse ranging from state-affiliated academic publications to newspapers as a medium that reflects the abstract concept of how China perceives the US. In fact, this paper cited their work when discussing how its findings can be applied to lift the veils of perceptions between great powers.

The "Trump bump" case study also supports this explanation on certain fronts. Amidst a state visit embellished with rosy platitudes, showstopping fanfare, and lavish ceremonies, bilateral sentiments reached local peaks. However, these sentiments soon after soured and plunged into freedives when the battlehorns for an impending trade war rang across the Pacific. By 2020, sentiments for both US and Chinese discourse had hit a low ebb, which fits well with consensus within foreign policy circles that bilateral ties, as well as mutual perceptions, have reached an unprecedented, freezing temperature.

When we apply this explanation, then the patterns of interaction picked up by Granger models serve as a proxy for the relationship between how the US perceives the state of its relationship with China, and vice versa. In this sense, the US's more "reactive" posture vis-à-vis Chinese sentiment trends demonstrates that China is often the first-comer in perceiving upturns, tensions, and discord in bilateral ties.

What must be highlighted here, however, is that China's "frontman" role in shaping new perceptions or setting new sentiment trends in foreign policy discourse is *not* always indicative of its tendency to escalate or de-escalate bilateral tensions. Turning back to the Trump case study, we can see that although US sentiments were reacting to Chinese sentiments, Chinese foreign policy discourse itself is a product of the visceral reaction against US *actions* (i.e. President

Trump's decision to declare a trade war via the USTR's Section 301 report). This means that although my paper's quantitative modeling techniques could capture bilateral perceptions and measure the warmth of US-China ties, it cannot take into account how concrete *actions*—as opposed to words—may stir up and muddy the foreign policy arena.

B) Discourse as a Constitutive and Productive Player

While explanation *A* encompasses a traditional understanding of foreign policy discourse as a passive reflection of reality, explanation *B* lends an ear to critical discourse analysis (CDA) and identifies discourse as a site where norms and meanings are actively constituted. In the words of Roxanne Doty, "statement making … is *productive*," and "language is seen as a set of signs which are part of a system for generating subjects, objects, and worlds."[54] Jennifer Milliken identifies two characteristics of discourse as viewed from a constructivist perspective—that it acts "as a system of signification" which can "construct social realities," and that it is imbued with "productivity," which at times means "operationalizing a particular regime of truth."[55]

All this poststructuralist jargon points to the idea that foreign policy discourse is not just a transparent looking glass through which geopolitical truths are reflected; rather, it is an active participant in international relations of its own right that can signal and produce new norms, new behaviors, and new roles.

A case in point here is the reconstitution of Chinese foreign policy discourse under the Hu Jintao administration. Throughout the past decades, China had been operating under a "minimalist" foreign policy doctrine encapsulated by Deng's slogans "hiding and biding" (韬光养晦), as well as "not taking the lead" (不当头). However, China's miraculous growth spurt,

[54] Roxanne Lynn Doty, "Foreign Policy as Social Construction: A Post-Positivist Analysis of U.S. Counterinsurgency Policy in the Philippines," *International Studies Quarterly* 37, no. 3 (1993): 302.
[55] Jennifer Milliken, "The Study of Discourse in International Relations: A Critique of Research and Methods," *European Journal of International Relations* 5, no. 2 (June 1, 1999): 229.

coupled with an increasingly complex external environment, pushed Beijing to search for a new *modus operandi* in the international arena. In this sense, Hu's new umbrella slogan for Chinese diplomacy, "harmonious world" (和谐世界), is a powerful message to the world that China was ready to come out of its metaphorical shell and enter a new developmental stage. Hu's foreign policy discourse, therefore, is not just a straightforward reflection of Chinese ambitions, but it is actively constituting a new role for the country to grow into. "Harmonious world," and by extension Xi's "community of common destiny" (命运共同体), signal new geopolitical norms that China is trying to construct.[56]

Similarly, on the American side, the Obama-era mantra "pivot to Asia," as well as the idea of "America's Pacific Century" (which crystallized in a *Foreign Policy* essay penned by Secretary Clinton), both actively seek to construct new realities on the ground. The "pivot," therefore, constitutes a signpost orienting America towards a more proactive and enhanced presence in the Asia-Pacific theater.[57] Similar to Hu's "harmonious world" or Xi's "common destiny," the Obama administration's foreign policy discourse has actively produced new norms and new dynamics within US-China diplomacy.

From this viewpoint, discourse is imbued with a sort of power to constitute and legitimize "new regimes of truth." According to Semenov and Tsyvk, moreover, the Chinese pay particularly close heed to the extent of their "discourse power" (话语权), which represents their ability to speak up on the international stage to popularize and legitimize their narratives and their worldviews.[58]

The notion that discourse is somehow charged with norm-constituting power has enormous implications for this paper. It pushes us to interrogate whether there is a link between

[56] Su Hao, "Harmonious World," 29-32.
[57] Janine Davidson, "The U.S. 'Pivot to Asia,'" *American Journal of Chinese Studies* 21 (2014): 77–82.
[58] Semenov and Tsvyk, "The Approach to the Chinese Diplomatic Discourse," 566.

being a “trend-setter” in foreign policy sentiments, and being in a position to steer the locomotive of discourse and harness all its constitutive power. Can the directionality of the Granger model point out whether the US or China possesses the powers of shaping discourse? And by extension to this question, does the US’s tendency to sit in the reactive seat serve as any indication that China has harnessed its “discourse power” with more success?

C) Discourse as Performance and Theater

Explanations *A* and *B* both contribute to the utility of the Granger models and affirm their applicability. The former suggests that lagged correlation between sentiment trends can be explained by changing mutual perceptions between the two states, whereas the latter argues that the direction of Granger causality can be explained by weighing relative “discourse power.” This third explanation, on the other hand, acknowledges the fact that there are exogenous factors that may detract from meaningful interpretations of my Granger models. Namely, foreign policy discourse can, at times, be detached from reality because it not only serves reflective, communicative, or norm-setting functions, but it is also highly *performative*.

For instance, discourse can serve this theatrical performance function for a country’s broad populace. Hua Chunying, the MFA spokesperson mentioned in the Trump case study above, has often been labeled by pundits as a quintessential “wolf warrior” diplomat for the virulent, anti-Western fervor that animates her rhetoric. According to the scholars Sullivan and Wang, Hua’s brand of fiery foreign policy discourse is incentivized by “cyber-nationalist” pressures emanating from the bottom-up.[59] But discourse does not always correspond to reality. Poh and Li’s assessment of the Xi administration notably leans toward the idea that China’s

[59] Jonathan Sullivan and Weixiang Wang, “China’s ‘Wolf Warrior Diplomacy’: The Interaction of Formal Diplomacy and Cyber-Nationalism,” *Journal of Current Chinese Affairs* 52, no. 1 (April 1, 2023): 68–88.

increasingly ambitious slogans, threats, and promises belie the abundance of foreign policy carry-over between Xi and his more mellow predecessors.[60]

The intimate linkage between popular sentiment and foreign policy rhetoric in America is also becoming ever more salient. Wojczewski's paper, for instance, studies how former President Trump weaponized the dichotomy between "self" and "other"—which often runs parallel to the delineation between American and Chinese—to whip up "populist-nationalist" fervor on the campaign trail.[61] A wide array of scholars, however, have pointed out that Trump has moderated or even pivoted away from some of his more unorthodox and rabble-rousing promises once he transitioned from the campaign trail to the White House.[62] This phenomenon once again demonstrates that the performative element of foreign policy rhetoric can detract from its link to reality.

What's more, a policymaker's rhetoric can also serve performative purposes for colleagues or superiors. Notably, an increasingly popular theory among China watchers is that faced with a slew of bureaucratic pressures, the MFA needed to deploy its most firebrand "wolf warriors" to the front lines of political discourse in order to avoid being shunted into irrelevance.

According to Zhao and Gao, as well as Jing Sun, the embattled MFA is being squeezed by bureaucratic pressures from all sides. For one, Xi is centralizing the decision-making pipeline by setting up Leading Small Groups (LSGs) with himself at the nexus, while excising the MFA from the political equation. At the same time, other agencies like the Ministry of Commerce (MOFCOM), the National Development and Reform Corporation (NDRC), and the People's

[60] Poh and Lee, "A China in Transition," 95-97.
[61] Thorsten Wojczewski, "Trump, Populism, and American Foreign Policy," *Foreign Policy Analysis* 16, no. 3 (June 1, 2020): 292–311.
[62] As a case in point, see Matthew Hill and Steven Hurst, "The Trump Presidency: Continuity and Change in US Foreign Policy," *Global Affairs* 6, no. 1 (January 1, 2020): 1.

Liberation Army have been encroaching upon the MFA's turf.[63] To stay afloat, the agency has to resort to jingoistic outbursts on the international stage to showcase its loyalty, as well as its ability to faithfully follow Xi's vision to the letter. As Peter Martin put very eloquently in his book *The Civilian Army*, the virulent "Wolf Warrior" rhetoric stems from the diplomatic corps being "unable to extricate themselves from the constraints of a secretive, paranoid political system which rewards unquestioning loyalty and ideological conviction."[64]

This bureaucratic turf wars theory also lends some credence to the idea that foreign policy discourse may not fully adhere to explanations *A* and *B*, and that it is always important to account for exogenous and domestic variables into the research calculus.

The Long (and Promising?) Road Ahead

In many ways, this paper represents a hopeful starting point to a winding journey of academic inquiry. It demonstrates that the immense foreign policy discursive landscape can be surveyed, extensively sampled, and then quantified. By turning complex and emotionally-charged words into interpretable vectors, moreover, this paper offers a "proof of concept" that we can apply a quantitative toolkit to the tough, abstract puzzles in foreign policy—like what are the patterns of interaction between the discourse of two states, and even how great powers perceive or misperceive each other.

Importantly, we see two paths dominating the future research roadmap. The former possibility entails creating a more extensive, and more comprehensive dataset. This means harnessing the full computational potential of a larger server and mining textual data pertaining

[63] Jing Sun, "Growing Diplomacy, Retreating Diplomats – How the Chinese Foreign Ministry Has Been Marginalized in Foreign Policymaking," *Journal of Contemporary China* 26, no. 105 (May 4, 2017): 420, 422; Kejin Zhao and Xin Gao, "Pursuing the Chinese Dream: Institutional Changes of Chinese Diplomacy under President Xi Jinping," *China Quarterly of International Strategic Studies* 01, no. 01 (April 2015): 47.
[64] Peter Martin, *China's Civilian Army: The Making of Wolf Warrior Diplomacy* (Oxford University Press, 2021), 5.

to a wider range of foreign policy stakeholders, such as the Politburo, the State Council, and MOFCOM in China, the National Security Council along with State and Defense in the US, or even think tanks and authoritative voices in the academy in both states. Indeed, a larger corpus of textual data at disposal means a more representative estimate of how one country perceives the other. What's more, the data-driven mode of inquiry into foreign policy discourse can transcend the US-China relationship. As formerly inaccessible textual data gets uploaded to the digital domain at an exponential rate, the horizons of the present methodology's applicability is virtually limitless. It is possible for us to collect data on any state's (or non state actor's) discourse vis-à-vis any other entity or topic and absorb the insightful stories that numbers can tell.

The second path along the future research roadmap entails conducting textual analysis with increasingly sophisticated methods. An enhanced textual classifier, for example, may integrate the algorithms developed by Chamber's et. al. to evaluate sentiment scores not for the entire text, but rather for a target named entity found within the text. Sentiment classification can also transcend the simple and rather limited binary spectrum ranging from positive to negative valence. To even more precisely capture how one country perceives another through a piece of discourse, we could ask an enhanced computational methodology to code a wide array of emotional frames, including anger, surprise, hopefulness, or perhaps the rhetoric is defined by more complex themes like persuasion, the sense of being threatened, or more ambivalent emotional baggages. Beyond situating texts along a tonal spectrum, it is also useful to consider how *stances* (i.e. expressions of affirmation or negation regarding a particular political issue or entity) are reflected within different texts.

These enhanced classifiers, in addition, could integrate the computational prowess of some state-of-the-art algorithms ranging from Stanford's deep learning-powered Core NLP API to Google's BERT model. By using these newer and flashier technologies to capture a multidimensional array of sentiment profiles and rhetorical themes, it is imminently possible to build a powerful model that can capture the complexities of how one country perceives another with record efficiency.

Academic heavyweights in the world of IR, like Bob Keohane and John Mearsheimer, have made it known that they believe discourse analysis to be bad scholarship. They claim that parsing discourse is bad science lacking testable theories and entirely untethered from empirical observation.[65] However, as this paper has shown, an investigation of how and why language matters in foreign policy can be integrated with the robustness and rigor of data science. In fact, the interdisciplinary approach to discourse can lay the foundation for a powerful foreign policy toolkit that functions as a real-time thermometer for bilateral relations and perceptual realities. A framework for scholarly inquiry propped up by data and NLP, therefore, may be a promising new frontier in the realm of IR.

As the spar over words continue to heat up the Pacific, and as initiatives for constructive dialogue between the US and China keep getting derailed by rabble-rousing nationalisms, it is becoming more important than ever for us to salvage what is left of this incredibly consequential—and potentially fruitful—dialogue. At this critical foreign policy juncture, will the two states barrel ahead with their virulent finger-pointing, or will they study the highly interactive patterns of bilateral discourse and replace self-absorption with empathy? The long road ahead for both states can be very promising, but only if we better understand how we talk about them, and how they talk about us.

[65] Milliken, "The Study of Discourse Analysis in International Relations," 227.

Works Cited

Abercrombie, Gavin, and Riza Batista-Navarro. "`Aye' or `No'? Speech-Level Sentiment Analysis of Hansard UK Parliamentary Debate Transcripts." In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA), 2018. https://aclanthology.org/L18-1659.

Ahmad, Munir, Shabib Aftab, Iftikhar Ali, and Noureen Hameed. "Hybrid Tools and Techniques for Sentiment Analysis: A Review" 8, no. 4 (2017).

Baumann, Andreas, Klaus Hofmann, Bettina Kern, Anna Marakasova, Julia Neidhardt, and Tanja Wissik. "Exploring Causal Relationships Among Emotional and Topical Trajectories in Political Text Data." In *3rd Conference on Language, Data and Knowledge (LDK 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

Bhatia, Aditi. "Critical Discourse Analysis of Political Press Conferences." *Discourse & Society* 17, no. 2 (March 1, 2006): 173–203. https://doi.org/10.1177/0957926506058057.

Brown, Chad, and Melina Kolb. "Trump's Trade War Timeline: An Up-to-Date Guide." *PIIE*. April 16, 2018. https://www.piie.com/blogs/trade-and-investment-policy-watch/trumps-trade-war-timeline-date-guide.

Cairns, Christopher, and Allen Carlson. "Real-World Islands in a Social Media Sea: Nationalism and Censorship on Weibo during the 2012 Diaoyu/Senkaku Crisis." *The China Quarterly* 225 (March 2016): 23–49. https://doi.org/10.1017/S0305741015001708.

Callahan, William A. "Tianxia, Empire and the World: Soft Power and China's Foreign Policy Discourse in the 21st Century." *British Inter-University China Centre Working Paper Series* 1, no. 2 (2007): 1-24. https://bpb-eu-w2.wpmucdn.com/blogs.bristol.ac.uk/dist/0/168/files/2012/06/01-Callahan.pdf.

Cao, Qing. "Confucian Vision of a New World Order?: Culturalist Discourse, Foreign Policy and the Press in Contemporary China." *International Communication Gazette* 69, no. 5 (October 1, 2007): 431–50. https://doi.org/10.1177/1748048507080868.

Chambers, Nathanael, Victor Bowen, Ethan Genco, Xisen Tian, Eric Young, Ganesh Harihara, and Eugene Yang. "Identifying Political Sentiment between Nation States with Social Media." In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 65–75. Lisbon, Portugal: Association for Computational Linguistics, 2015. https://doi.org/10.18653/v1/D15-1007.

Choi, Jaedong, Rinseo Park, Young Min Baek, and Jae Mook Lee. "Unraveling the Thought Probes of US Legislators on Trade Negotiations: Sentiment Analysis of the 108th and 112th US Congressional Speeches." *Pacific Focus*. Accessed March 25, 2023. https://doi.org/10.1111/pafo.12220.

Council on Foreign Relations. "Timeline: U.S.-China Relations." Accessed April 17, 2023. https://www.cfr.org/timeline/us-china-relations.

Davidson, Janine. "The U.S. 'Pivot to Asia.'" *American Journal of Chinese Studies* 21 (2014): 77–82.

D'Andrea, Alessia, Fernando Ferri, Patrizia Grifoni, and Tiziana Guzzo. "Approaches, Tools and Applications for Sentiment Analysis Implementation." *International Journal of Computer Applications* 125, no. 3 (2015): 26-33. https://www.ijcaonline.org/research/volume125/number3/dandrea-2015-ijca-905866.pdf.

Doty, Roxanne Lynn. "Foreign Policy as Social Construction: A Post-Positivist Analysis of U.S. Counterinsurgency Policy in the Philippines." *International Studies Quarterly* 37, no. 3 (1993): 297–320. https://doi.org/10.2307/2600810.

Feige, Edgar L., and Douglas K. Pearce. "The Casual Causal Relationship Between Money and Income: Some Caveats for Time Series Analysis." *The Review of Economics and Statistics* 61, no. 4 (1979): 521–33. https://doi.org/10.2307/1935784.

Fisher, Louis. "Foreign Policy Powers of the President and Congress," *The Annals of the American Academy of Political and Social Science* 499 (1988): 148–59. https://www.jstor.org/stable/1045825.

Fisher, Scott. "China's People's Daily Articles from 12 May 2007 to 31 Dec 2020." Harvard Dataverse, April 16, 2022. https://doi.org/10.7910/DVN/HWNN7S.

Fisher, Scott, Graig R Klein, and Juste Codjo. "Focusdata: Foreign Policy through Language and Sentiment." *Foreign Policy Analysis* 18, no. 2 (April 1, 2022). https://doi.org/10.1093/fpa/orac002.

Freeman, John R. "Granger Causality and the Times Series Analysis of Political Relationships." *American Journal of Political Science* 27, no. 2 (1983): 327–58. https://doi.org/10.2307/2111021.

Ghosh, Sajal. "Electricity Consumption and Economic Growth in India." *Energy Policy* 30, no. 2 (January 2002): 125–29. https://doi.org/10.1016/S0301-4215(01)00078-7.

Goldstein, Steven M. *China and Taiwan*. John Wiley & Sons, 2015.

Hao, Su. "Harmonious World: the Conceived International Order in Framework of China's Foreign Affairs." *Foreign Affairs* 87, no. 1 (2008): 29-55. http://www.nids.go.jp/english/publication/joint_research/series3/pdf/3-2.pdf.

Hart, Nina M., and Brandon J. Murrill. "Section 301 Tariffs on Goods from China: International and Domestic Legal Challenges." *Congressional Research Service*, April 5, 2022. https://crsreports.congress.gov/product/pdf/LSB/LSB10553.

Hill, Matthew, and Steven Hurst. "The Trump Presidency: Continuity and Change in US Foreign Policy." *Global Affairs* 6, no. 1 (January 1, 2020): 1–3. https://doi.org/10.1080/23340460.2020.1726788.

Jervis, Robert. *Perception and Misperception in International Politics: New Edition*. Princeton University Press, 2017.

Jisi, Wang, and Hu Ran. "From Cooperative Partnership to Strategic Competition: A Review of China–U.S. Relations 2009–2019." *China International Strategy Review* 1, no. 1 (June 1, 2019): 1–10. https://doi.org/10.1007/s42533-019-00007-w.

Joyce, Brandon, and Jing Deng. "Sentiment Analysis of Tweets for the 2016 US Presidential Election." In *2017 IEEE MIT Undergraduate Research Technology Conference (URTC)*, 1–4. Cambridge, MA: IEEE, 2017. https://doi.org/10.1109/URTC.2017.8284176.

Kazmaier, Jacqueline, and Jan H. Van Vuuren. "The Power of Ensemble Learning in Sentiment Analysis." *Expert Systems with Applications* 187 (2022). https://doi.org/10.1016/j.eswa.2021.115819.

Kolchyna, Olga, Tharsis T. P. Souza, Philip Treleaven, and Tomaso Aste. "Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination." arXiv, September 18, 2015. http://arxiv.org/abs/1507.00955.

Kónya, László. "Exports and Growth: Granger Causality Analysis on OECD Countries with a Panel Data Approach." *Economic Modelling* 23, no. 6 (December 1, 2006): 978–92. https://doi.org/10.1016/j.econmod.2006.04.008.

Lean, Hooi Hooi, and Russell Smyth. "Multivariate Granger Causality between Electricity Generation, Exports, Prices and GDP in Malaysia." *Energy* 35, no. 9 (September 1, 2010): 3640–48. https://doi.org/10.1016/j.energy.2010.05.008.

Liu, Dilin, and Lei Lei. "The Appeal to Political Sentiment: An Analysis of Donald Trump's and Hillary Clinton's Speech Themes and Discourse Strategies in the 2016 US Presidential Election." *Discourse, Context & Media* 25 (October 1, 2018): 143–52. https://doi.org/10.1016/j.dcm.2018.05.001.

Liu, Ke, and Fang Guo. "A Review on Critical Discourse Analysis." *Theory and Practice in Language Studies* 6, no. 5 (May 17, 2016): 1076. https://doi.org/10.17507/tpls.0605.23.

Lyu, Xiaoting, Zhe Chen, Di Wu, and Wei Wang. "Sentiment Analysis on Chinese Weibo Regarding COVID-19." In *Natural Language Processing and Chinese Computing*, edited by Xiaodan Zhu, Min Zhang, Yu Hong, and Ruifang He, 710–21. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-60450-9_56.

Martin, Peter. *China's Civilian Army: The Making of Wolf Warrior Diplomacy*. Oxford University Press, 2021.

Mearsheimer, John J.. *The Tragedy of Great Power Politics*. W. W. Norton & Company, 2001.

Mehrara, M., and Bagher Adabi Firouzjaee. "Granger Causality Relationship between Export Growth and GDP Growth in Developing Countries: Panel Cointegration Approach," *International Journal of Humanities and Social Science* 1, no. 16 (2011): 223-231. https://www.semanticscholar.org/paper/Granger-Causality-Relationship-between-Export-and-Mehrara-Firouzjaee/087d122898e4eac586dec2c571707ceb67034083.

Milliken, Jennifer. "The Study of Discourse in International Relations: A Critique of Research and Methods." *European Journal of International Relations* 5, no. 2 (June 1, 1999): 225–54. https://doi.org/10.1177/1354066199005002003.

Mochtak, Michal, and Richard Q. Turcsanyi. "Chinese Ministry of Foreign Affairs Press Conferences Corpus (CMFA PressCon)." Harvard Dataverse, July 1, 2022. https://doi.org/10.7910/DVN/BAKGET.

Mochtak, Michal, and Richard Q. Turcsanyi. "Studying Chinese Foreign Policy Narratives: Introducing the Ministry of Foreign Affairs Press Conferences Corpus." *Journal of Chinese Political Science* 26, no. 4 (December 1, 2021): 743–61. https://doi.org/10.1007/s11366-021-09762-3.

Mokry, Sabine. "China's Foreign Policy Rhetoric between Orchestration and Cacophony." *The Pacific Review* (February 20, 2023): 1–28. https://doi.org/10.1080/09512748.2023.2175895.

Morning Consult. "U.S.-China Relations Barometer." Accessed April 16, 2023. https://morningconsult.com/us-china-relations-barometer/.

Nathan, Andrew J., and Andrew Scobell. *China's Search for Security*. Columbia University Press, 2015.

Poh, Angela, and Mingjiang Li. "A China in Transition: The Rhetoric and Substance of Chinese Foreign Policy under Xi Jinping." *Asian Security* 13, no. 2 (May 4, 2017): 84–97. https://doi.org/10.1080/14799855.2017.1286163.

Proksch, Sven-Oliver, Will Lowe, Jens Wäckerle, and Stuart Soroka. "Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Legislative Speeches." *Legislative Studies Quarterly* 44, no. 1 (2019): 97–131. https://doi.org/10.1111/lsq.12218.

BBC News. "Trump China Visit: US Leader Strikes Warmer Tone with Xi Jinping." November 9, 2017, sec. China. https://www.bbc.com/news/world-asia-china-41924228.

Semenov, Alexander, and Anatoly Tsvyk. "The Approach to the Chinese Diplomatic Discourse." *Fudan Journal of the Humanities and Social Sciences* 14, no. 4 (December 1, 2021): 565–86. https://doi.org/10.1007/s40647-021-00321-x.

Scobell, Andrew. "Perception and Misperception in U.S.‐China Relations." *Political Science Quarterly (Wiley-Blackwell)* 135, no. 4 (December 2020): 637–64. https://doi.org/10.1002/polq.13099.

Shambaugh, David. *Beautiful Imperialist: China Perceives America, 1972-1990*. Princeton University Press, 1991. https://doi.org/10.2307/j.ctv1ddd0sc.

Shojaie, Ali, and Emily B. Fox. "Granger Causality: A Review and Recent Advances." *Annual Review of Statistics and Its Application* 9, no. 1 (2022): 289–319. https://doi.org/10.1146/annurev-statistics-040120-010930.

Sullivan, Jonathan, and Weixiang Wang. "China's 'Wolf Warrior Diplomacy': The Interaction of Formal Diplomacy and Cyber-Nationalism." *Journal of Current Chinese Affairs* 52, no. 1 (April 1, 2023): 68–88. https://doi.org/10.1177/18681026221079841.

Sun, Jing. "Growing Diplomacy, Retreating Diplomats – How the Chinese Foreign Ministry Has Been Marginalized in Foreign Policymaking." *Journal of Contemporary China* 26, no. 105 (May 4, 2017): 419–33. https://doi.org/10.1080/10670564.2016.1245895.

United States Department of State. "Secretary Antony J. Blinken, National Security Advisor Jake Sullivan, Director Yang And State Councilor Wang At the Top of Their Meeting." Accessed April 16, 2023. https://www.state.gov/secretary-antony-j-blinken-national-security-advisor-jake-sullivan-chinese-director-of-the-office-of-the-central-commission-for-foreign-affairs-yang-jiechi-and-chinese-state-councilor-wang-yi-at-th/.

U.S. Congressional Documents, distributed by HeinOnline. Accessed February 2, 2023. https://heinonline.org/HOL/Index?collection=congrec.

U.S. Presidential Library, distributed by HeinOnline. Accessed January 30, 2023. https://heinonline.org/HOL/Index?collection=presidents&set_as_cursor=clear.

Usher, James, Lucía Morales, and Pierpaolo Dondio. "BREXIT: A Granger Causality of Twitter Political Polarisation on the FTSE 100 Index and the Pound." In *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 51–54, 2019. https://doi.org/10.1109/AIKE.2019.00017.

Verma, Binita, and Ramjeevan Thakur. "Sentiment Analysis Using Lexicon and Machine Learning-Based Approaches: A Survey." In *Lecture Notes in Networks and Systems*, 441–47, 2018. https://doi.org/10.1007/978-981-10-8198-9_46.

Wang, Zheng. *Never Forget National Humiliation: Historical Memory in Chinese Politics and Foreign Relations*. Columbia University Press, 2014.

Wei, Da. "A Restructuring International Order and the Paradigm Shift in China–U.S. Relations." *China International Strategy Review* 1, no. 1 (June 1, 2019): 21–32. https://doi.org/10.1007/s42533-019-00004-z.

Wojczewski, Thorsten. "Trump, Populism, and American Foreign Policy." *Foreign Policy Analysis* 16, no. 3 (June 1, 2020): 292–311. https://doi.org/10.1093/fpa/orz021.

Wright, Thomas. "The US and China Finally Get Real with Each Other." *Brookings* (blog), March 22, 2021. https://www.brookings.edu/blog/order-from-chaos/2021/03/22/the-us-and-china-finally-get-real-with-each-other/.

Zeng, Ka, Rob Wells, Jingping Gu, and Austin Wilkins. "Bilateral Tensions, the Trade War, and US–China Trade Relations." *Business and Politics* 24, no. 4 (December 2022): 399–429. https://doi.org/10.1017/bap.2022.8.

Zhang, Biwu. *Chinese Perceptions of the U.S.: An Exploration of China's Foreign Policy Motivations*. Lexington Books, 2011.

Zhao, Kejin, and Xin Gao. "Pursuing the Chinese Dream: Institutional Changes of Chinese Diplomacy under President Xi Jinping." *China Quarterly of International Strategic Studies* 01, no. 01 (April 2015): 35–57. https://doi.org/10.1142/S2377740015500037.

Zheng, Yongnian, and Sow Keat Tok. "Harmonious Society and Harmonious World: China's Policy Discourse Under Hu Jintao." *Briefing Series* 26 (2007): 1-12. https://www.nottingham.ac.uk/iaps/documents/cpi/briefings/briefing-26-harmonious-society-and-harmonious-world.pdf.

Zhu, Lei, and Wei Wang. "A Critical Discourse Analysis of the US and China Political Speeches—Based on the Two Speeches Respectively by Trump and Wang Yi in the General Debate of the 72nd Session of UN Assembly." *Journal of Language Teaching and Research* 11, no. 3 (May 1, 2020): 435. https://doi.org/10.17507/jltr.1103.12.

田桂秋. "基于评价理论的中美外交演讲对比研究." 硕士, 中国海洋大学, 2014. https://oversea.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201501&filename=1014369058.nh&uniplatform=OVERSEA&v=9jhztGzm1x-sa1OecVFnri0MVb8R6IxOq2G8OgiMWCXw0pvUvmCItVCNfVuZ-CkF.

罗瑞丰. "中美国家元首外交演讲话语的批评话语分析." 博士, 上海外国语大学, 2020. https://doi.org/10.27316/d.cnki.gswyu.2020.000319.

赵静, and 詹全旺. "中美外交部发言人在例行记者会上使用模糊限制语的对比分析." 阜阳师范学院学报*(*社会科学版*)*, no. 3 (2018): 43–49. https://doi.org/10.14096/j.cnki.cn34-1044/c.2018.09.