

# Event Structure and State Source Dependence in Events Data Construction\*

David Sylvan<sup>†</sup>

Ashley Thornton<sup>‡</sup>

Jean-Louis Arcand<sup>§</sup>

August 24, 2023

## Abstract

In spite of their enormous breadth and sophistication, most international relations events compendia tend to underplay a key aspect of the way that events are depicted in the sources of those compendia; they also tend to underplay the reliance of those compendia on state sources. We discuss both of these issues, both in general and then in light of our own research project, which involves generating newspaper interpretations from state announcements by the central bank, and about U.S.-Russian relations. We use both our project workflow and the preliminary results of our generative text model to make specific and, we think, feasible, suggestions about how existing events data coding efforts can be complemented, and how the nature of state source dependence can be assessed across sources, though, importantly, not avoided.

Keywords: Conflict, diplomacy, journalism, events data, text analysis, methodology, machine learning, natural language processing

## 1 Introduction

Over the last several decades, there has been a striking increase in the number and rigorous construction of events data compendia, particularly, those concerned with conflict events. Researchers who want to study geographical and temporal patterns of such events, as a way of elaborating or assessing phenomena such as civil wars, escalation and de-escalation spirals, and media framing find themselves in a far different position now than would have been the case even 5-10 years ago. Especially in recent years, the internet has permitted systematic access to a range of sources which earlier researchers could only dream about; by the same token, advances in both computational technology and natural language processing (NLP) make it possible to

---

\*We thank Shekhar Hari Kumar, Maria Kamran, Daniele Rinaldo, Raphaël Leduc, Eugenia Zena, Clarisse Labbé, Claire Ransom, Alessandra Romani, and Julia Greene for their work on data construction and analysis; and Andreas Marfurt, Alireza Mohammadshahi, and James Henderson for their work on machine learning models; we are particularly grateful to Andreas Marfurt for his help on this paper. We also gratefully acknowledge the support of the Swiss National Science Foundation under grant CRSII5.180320. Prepared for delivery at the 119th Annual Meeting and Exhibition of the American Political Science Association, August 31-September 3, 2023.

<sup>†</sup>Department of International Relations / Political Science, Graduate Institute of International and Development Studies, Geneva, Switzerland; and Faculty of Governance, Economics and Social Sciences, Mohammed VI Polytechnic University, Rabat, Morocco. Email: david.sylvan@graduateinstitute.ch

<sup>‡</sup>Intrepid Project, Graduate Institute of International and Development Studies, Geneva, Switzerland. Email: ashley.thornton@graduateinstitute.ch

<sup>§</sup>Global Development Network, New Delhi, India; FERDI, Clermont Ferrand, France; Faculty of Governance, Economics and Social Sciences, Mohammed VI Polytechnic University, Rabat, Morocco; and Department of International Economics, Graduate Institute of International and Development Studies, Geneva, Switzerland. Email: jlarcand@gdn.int

deal with a massive number of news stories and indeed to convert those stories to events data automatically.

As these efforts have progressed, events data scholars have increasingly turned their attention to two aspects of data construction: increasing the number of event types and of actual events that make their way into events data compendia; and improving the granularity of those data, for example by specifying greater geographical or actor detail (Halterman et al. 2023a,b; Douglass et al. 2022; Stundal et al. 2021; Althaus, Peyton and Shalmon 2022). (These aspects, of course, are related, since certain places are likely both to have fewer events picked up by standard news sources and to be less detailed about the events which are picked up.) Although both of these issues are of considerable importance, our argument in this paper is that we also need to pay attention to two other issues: the structured nature of political events and of the role of journalism in covering those events.

To do this, we begin with some observations on what it is that journalists do when they write news stories, notably how they gloss the events that are the basis for the stories, and how they find out about those events in the first place. These points are then illustrated by an overview of the research project from which this paper is drawn, focusing both on our data construction techniques and on preliminary results from the machine learning (ML) model of those journalistic glosses. We then return to the issue of events data, making certain suggestions about how to add additional structure to the coding of political events and how to deal with the nearly inevitable dependence of journalists on state sources.

## 2 Events and Journalists

### 2.1 Relevancy criterion

In the events data world, the standard way of representing political events is with some kind of 3-tuple, wherein an actor A carries out an action (V, for verb) relative to some target B. Thus, canonically, A may attack, threaten, or praise B. Within this framework, numerous variations are possible: for example, both A and B may be collective entities or even, at least metonymically, places (“Group X occupied the town”); by the same token, V may be carried out by both A and B (“Macron and Biden negotiated”). Moreover, the 3-tuple can be augmented by various optional elements: when the event took place, where it occurred, methods that the actors employed, various numbers (explosions, deaths, size of an attacking force), the goals that the actors were pursuing when acting, the circumstances that led to the event, the justifications that the actors provided, and the issue context surrounding the event (e.g., a threat in the context of tariff negotiations is not the same as a military threat in the context of a border dispute). But in principle, none of these elements need be included for the event in question to be well-formed: one could have simply the 3-tuple: country A condemned country B; insurgents attacked government troops; police arrested demonstrators.

Our focus in this paper is on the way events are reported journalistically and enter into events data compendia; but it is worthwhile noting that the 3-tuple core is not in fact the way that most social science explanations of two-actor events conceptualize those events. Instead, standard explanations of two-actor events operate with, at a minimum, a 4-tuple conception of

events, if not a 5-tuple one, with the additional elements being the goals of at least actor A, if not of both A and B. Thus, to take a purely random example: Trump (A) demanded (V1) that the Georgia Secretary of State (B) find him enough votes so that he would win Georgia (goal-A), something that B had refused to do (V2) on the grounds that it was illegal and he would not break the law (goal-B). We could imagine representing this as a pair of connected sub-events: A-V1-B, and B-V2-A with the first sub-event's internal connections being given by goal-A and the second's by goal-B; and we could also imagine that if our principal interest was explaining the antecedents or consequences of V1, we would therefore gloss the overall event as only involving V1. In emphasizing the importance of goals, we are hardly saying anything earth-shattering: for decades, social scientists have been concerned with goal-directed actions, and indeed, the classification systems used in every events data project implicitly recognize those goals (for example, the fact that A's target was B, not C, or that the act was V1 rather than V2). However, this recognition is implicit, not explicit, and the frequent opacity or silence of events data projects on the actual rules that human coders are to employ in constructing their 3-tuples makes it harder still to posit goals after the fact.<sup>1</sup>

Nonetheless, even if one were to be satisfied with the 3-tuple schema on social science grounds, it is certainly not what journalists are doing when they write stories about events. The key term here is "stories": journalists are constructing narratives, not simply saying that some goal-less action took place. For example, here are the first three paragraphs (using Google Translate, with minor corrections, for the stories in French and German) from news stories on one morning in August:

President Biden and the leaders of Japan and South Korea agreed to expand security and economic cooperation on Friday, after a summit at Camp David that sought to forge a bulwark against North Korea and the growing influence of China.

Mr. Biden said the three leaders had agreed to hold annual military exercises and a trilateral meeting to deepen their alliance, not just this year, not just next year, forever.

The American president lauded the political courage of Prime Minister Fumio Kishida of Japan and President Yoon Suk Yeol of South Korea for overcoming long-standing resentments stemming from Japan's occupation of the Korean Peninsula before World War II to create a new alliance. (*New York Times*, 19 August 2023).

Taiwan said on Saturday August 19 that it had detected incursions by Chinese military aircraft into its air defense zone. These are military exercises that China had announced. The island's Ministry of Defense explains in a press release that it has "successively detected forty-two 'incursions' since 9 a.m. [3 a.m. French time]."

Among these assaults, 26 warplanes crossed the center line of the Taiwan Strait, the ministry said. Eight Chinese ships also took part in the maneuvers, which Beijing

---

<sup>1</sup>It is striking how rarely codebooks are publicly available, and even more so how rarely they include instructions as to how coders are to identify particular types of 3-tuples, with examples of what should and should not be done. This raises serious questions about how scholarship can proceed, questions which are hardly obviated by reference to commercial- or security-related secrecy concerns. To use a not-unfair analogy, it is as if physicists were to publish an extensive set of experimental results without letting anyone know the details of their apparatus or their experimental protocols.

said was intended to simulate “real combat conditions.” The Taiwanese ministry said that “the national army [was] monitor[ing] and us[ing] reconnaissance methods to strictly control” the situation, adding that it has sent planes and ships.

China has launched military maneuvers around Taiwan, in response to recent stops in the United States by William Lai, the vice president of the island, a territory claimed by Beijing. “The Chinese People’s Liberation Army Eastern Area Command on Saturday launched joint air and sea patrols and navy and air force military drills around the island of Taiwan,” wrote the official New China agency, quoting army spokesman Shi Yi. (*Le Monde*, 19 August 2023).

The “Joseph Schulte” has not been in the Black Sea since Friday morning. Two days after leaving Odessa, the container ship passed the Bosphorus. This is the first time a cargo ship has reached the Turkish Straits from a Ukrainian Black Sea port since Russia withdrew from the grain agreement with the United Nations.

On July 17, Russia announced its withdrawal from the agreement that had guaranteed the safe transport of agricultural goods from Ukraine across the Black Sea since last summer. As a result, Moscow declared every ship sailing to or from a Ukrainian port a legitimate war target. This poses major problems for Ukraine, which is one of the world’s most important exporters of wheat, corn and sunflower oil.

The safe passage of the Joseph Schulte is a success for Ukraine. A week ago, Kiev announced the establishment of a “humanitarian corridor” for ships from the Black Sea ports in the greater Odessa area. In addition to the port in Odessa, this also includes the ports of Pivdenni (formerly Yuzhne) and Chornomorsk. The route runs along the Ukrainian, Romanian and Bulgarian Black Sea coasts relatively close to the mainland. (*Neue Zürcher Zeitung*, 19 August 2023).

In each of these cases, the article is telling a story: there is a pre-existing situation or indeed, multiple such situations which now, because of one or more recent actions, have been complicated.<sup>2</sup> (It is in fact the way that the recent action complicates the existing situation which makes it newsworthy; cf. Gans 1979.) In the case of the first story, there are two pre-existing situations: policies of North Korea and China, and a fraught history between Japan and South Korea; the action, that of agreeing to cooperate on security issues, is cast as being aimed at countering the first situation and to some degree moving beyond the second. In the case of the second story, there are at least two pre-existing situations: Chinese claims to Taiwan and the recent visit by the Taiwanese vice president to the U.S.; the action by China is cast as an immediate, escalatory, response to the second situation and as indicative of the first one. Finally, the third story also refers to multiple situations: the war between Ukraine and Russia, the cancelation of the agreement on grain exports, and the new policy announced by Ukraine; the ship’s arrival in the Bosphorus is indicative of at least some success for Ukraine. Notice that in connecting the pre-existing situation to the recent action and the complication, the story is

---

<sup>2</sup>This terminology is deliberately drawn from Labov’s work on narrative: Labov and Waletzky 1967, 1997; Labov 2013.

attributing goals to one or more of the actors: this is explicit in the case of the first story and implicit in the case of the second and third. However, the attribution of goals is less important to the story nature of the articles than the inclusion of an additional element which is precisely what makes for the narrative complication, and hence for the story’s newsworthiness.

To get a better idea of how concern with newsworthiness leads journalists to recount events in a particularly structured way, consider what is newsworthy about each of the above events. In each of these cases, there is a fourth element: <the U.S.> (1) <forming an alliance>(2) with <two nearby countries> (3) <against China> (4); or, alternatively, <South Korea> (1) and <Japan> (2) <burying the hatchet> (3) <after decades of tension> (4); <China>(1) <intruding> (2) into <South Korea’s>(3) air defense zone <as a response to a recent high-profile visit by Taiwan to its military backer> (4); <Ukraine>(1) <was able to counter> (2) <Russia’s>(3) <recently-escalated grain trade policy> (4). For each story, if the fourth element were not present, there would be no news to report because there would be no complication (cf. Hayden White (1980) on annals and chronicles versus narratives). That fourth element might be a goal, or a response, or a temporal or historical contrast. For purposes of discussion (and to sound less like a certain French science fiction film), we will call this fourth element a relevancy link. Using that terminology and returning to our earlier example, the newsworthiness of stories about Trump’s call to the Georgia Secretary of State was less that Trump wanted to get more votes as a way of winning the election and, qua relevancy criterion, more about the illegal (RC1) and unprecedented (RC2) nature of the request, something which had never happened over the course of U.S. history.

The importance of the relevancy criterion does not mean that there may not also be other elements in a news story. Journalists can and do specify multiple details about newsworthy events, but for an event to be glossed as a news story, it must include the four elements listed above. If any one of those core elements, as we can call them, is not present, at least implicitly, the event is unlikely to be reported. We will return to this point below when we discuss how, in our research project about journalists’ interpretations of state announcements, we represent the core and non-core elements of news stories.

It might of course be argued that whatever journalists do or do not do, it has no particular bearing on how political scientists should construct events data compendia: put simply, there is no reason for us to include a relevancy criterion as a core feature of events. However, this sort of argument is belied by the actual procedures used to construct, as well as to assess the validity of, events data. Typically, there are two pathways by which information is turned into events: reliance on media sources, whether international (e.g., Reuters) or national (e.g., *Dawn*, as in the BFRS dataset developed by Bueno de Mesquita et al. 2015); and reliance on non-media sources (e.g., the SIGACTS military database discussed by Weidmann 2015, 2016, or the CINEP reports discussed by Stundal et al. 2021). Although it needs of course to be proven, let us assume for now that media sources, whether international or national, operate with some sort of newsworthy-style relevancy criterion. But arguably, the same is also true of non-media sources. Consider SIGACTS, i.e., the “significant activities” collected by the U.S. military from unit reports from its own and allied forces in Afghanistan and Iraq, mostly about activities attributed to insurgents. While there are many types of significant activities, and

while various of those activities may well be anticipated to occur with considerable frequency, the relevancy criterion is one in which they do NOT occur, however utopian this non-occurrence might be. Thus, for example, if a certain amount of time goes by without an explosion in a market, or if particular counterinsurgency sweeps are not met by sniper fire, there is nothing to report. This is not to deny that many of the SIGACTS entries are likely too low-level to be picked up subsequently by journalists and incorporated into news stories; rather, the point is that even the military's own internal reporting incorporates a relevancy criterion. For this reason, although the need for finding a means to incorporate the relevancy criterion into the coding of an event stems from the nature of journalistic glosses, the point likely holds for events derived from any source.

## 2.2 Sources: official and officials

We turn now to a second, albeit related, issue, namely, the way in which journalists find out about events in the first place. Here, it is useful to start by recalling the words of one of the most celebrated journalists of the U.S.-Vietnam War, the late David Halberstam, known for his extreme skepticism toward official press briefings by the U.S. military and embassy in Saigon. In January of 1963, the battle of Ap Bac occurred. Halberstam (2008: 88-9), whose reporting of the battle irritated large numbers of U.S. officials, recalled several years later that he spoke with U.S. helicopter pilots, with the subsequently famed U.S. colonel John Paul Vann (“a miserable damn performance, just like it always is”), with his friend and colleague, Neil Sheehan (who in turn had talked with Brig. Gen. Robert York: “What happened?” “What the hell’s it look like happened, boy? They got away – that’s what happened”), and with the commanding U.S. general, Paul Harkins (“We’ve got them in a trap and we’re going to spring it in half an hour”). The point here is not that Halberstam’s sources all put forward the same line – Vann and Harkins were polar extremes – but rather that high-quality journalism involves multiple sources, many of whom, inevitably, are officials of the state, even if they are not putting forward the official position. When one descends from Halberstam to run of the mill journalists, the reliance on state sources becomes both more pronounced and narrower, taking more effort to escape from the “Five O’Clock Follies” that were the daily press briefings in Saigon. By the time of the U.S. wars in Afghanistan and Iraq, reporters were subject to far stricter controls (including in so-called “embedded” assignments: Manning 2014: Jun. 14 ), which is one reason that many of them stuck with official briefings in Kabul and Baghdad. Analogous phenomena occur, albeit on a much smaller scale, with journalists who rely on insurgent groups for sourcing.

When one turns from violent conflicts to a broader range of events, the reliance on official sources is even stronger. In part, this is because of confidentiality and access issues: reporters will simply not know about, much less be able to observe firsthand, most types of negotiations, working group meetings, email exchanges, telephone calls, and so forth. But the reliance on official sources is also due to the fact that such sources are readily available. For example, in the United States, the White House, the State Department, and the Defense Department have frequent, regularly scheduled press briefings in which formal statements are issued and officials (by no means only press secretaries) respond to reporters’ questions; there are also formal press releases without commentary. When top officials travel, there are often special

press briefings, both on and off the record, as well as “gaggles”: briefings in which video recording is not permitted. Beyond that, many phone calls and meetings are the subject of “readouts”: summaries of the conversation. In addition, top officials make speeches, occasionally testify before legislative bodies, and are regularly interviewed, all of which give rise to verbatim transcripts. Last but not least, there are various leaks granted to small groups of, or even individual, journalists. In short, the modern state has an extensive set of sources available for, and indeed aimed at, journalists, and, as we will discuss below, we have found that foreign affairs coverage, broadly defined, revolves to a great degree around writing up the information transmitted by these sources into news stories.<sup>3</sup> Even stories about activities that do not themselves involve acts by state officials are often based to a significant degree on official sources: for example, a story in the *New York Times* (18 August) about “illegal border crossings” into the U.S. was based on Customs and Border Protection data released that day and accompanied by a quotation from an assistant secretary of Homeland Security.

This cycle of announcements and stories, in which state officials spend a considerable amount of time crafting and disseminating material to journalists, and in which, in turn, journalists themselves spend a considerable amount of time obtaining, processing, and writing up in news story form the information they receive, is nothing new. It was identified and analyzed a century ago by Walter Lippmann (1922), who in describing the effect of exposure to news stories on public opinion, coined the now-famous phrase “the manufacture of consent.”<sup>4</sup> This is not to say that news stories are deliberately or exclusively propagandistic – indeed, some reporters pride themselves on their skepticism toward officials – but simply that news stories are to a great degree constructed from the materials presented to them, already packaged, by state officials. The extent to which that construction overlaps with the source materials is an empirical question which we will address below. To be sure, it is highly likely that the overlap varies by issue area, identity of the source and of the journalistic output, and time period, but it should be understood that journalism relies heavily on state officials.

Putting together this reliance with the preceding point about the necessary inclusion of relevancy criteria all along the chain leading from information to events data means that compendia of such data ought not to be seen as partial and likely biased windows onto a landscape of actual events. Instead, they are systematic and highly stylized collections of events as originally constructed mostly by state officials.<sup>5</sup> That, it should be noted, is quite different than saying that events data compendia are ideologically biased (they may be, but this has nothing to do with their being based to a great degree on officials’ event construction). Nor should it be construed as a claim that the phenomena being collected have nothing to do with the lives, and deaths, of real human beings: quite the contrary, as these compendia reflect how those lives and deaths are understood by chains of officials and journalists. Compendia of such

---

<sup>3</sup>Although the examples in this paragraph are for the United States, most states have a similar set of disseminating practices.

<sup>4</sup>Cf. Herman and Chomsky (1988).

<sup>5</sup>Cf. the ethnomethodological conception of “accounts”: Garfinkel (2002) and, for a political science example, Nippert (2012). From this perspective, the issue with state source dependence is not so much that certain types of actual events are either being missed or distorted (Windsor 2022); rather, it is that the vast majority of those omissions and distortions are baked into the compendia-creation process. On a related issue, see the debate sparked by Kalyvas (2001).

chain-constructed events data can and should be studied to see how their patterning across time, space, and type feeds into explanations about politics, but it should be understood that the questions we can ask about that patterning are limited by the nature of the chains by which the events are constructed. These are points we will return to below; for now, however, we turn to a walk-through of how, in our project on journalistic interpretations of state announcements, we deal with both state sourcing and relevancy criteria.

## 3 Project

### 3.1 Overview

As indicated above, our project<sup>6</sup> is concerned with how journalists interpret state announcements. We know that this is a ubiquitous phenomenon, covering numerous announcements by agencies and individuals, and involving routine coverage of many of those announcements by journalists in the form of news stories. The interaction between announcements and interpretations is of considerable importance both to state officials – they make the announcements precisely so that journalists will produce news stories about them – and to various consumers of news stories (for example, stock market traders who have no time to listen to a three-hour hearing with the chair of the Federal Reserve, or members of Congress who lack the detailed knowledge of weapons systems discussed in a DoD briefing). Let us start with some definitions.

A **policy announcement** is a written statement issued by a state agency describing its actions (including comments on others’ actions) or decisions. As indicated above, announcements can take the form of written communiques, press releases, statements read before the media, background briefings (sometimes in anonymous form), and answers to journalists’ or legislators’ questions (the questions and answers are often transcribed and subsequently released in written form). Many state agencies, in many countries, make regular policy announcements, often on a pre-scheduled basis, such as weekly, monthly, or quarterly. These announcements may be made out of a sense of obligation (indeed, in some cases, they may be legally mandated, as in the case of the Humphrey-Hawkins Full Employment Act in the United States, which instructed the Board of Governors of the Federal Reserve to transmit a Monetary Policy Report twice a year to the Congress) or out of a desire to send signals to different political and economic actors. While individual agencies in different states vary in the frequency, formality, and extent of their announcements,<sup>7</sup> they nonetheless do in fact make such announcements on a regular basis and have done so for many years. As we will see below, this regularity makes it possible to construct corpora from streams of announcements and, at least for certain agencies, for those corpora to be sufficiently dense as to make ML techniques feasible.

An **interpretation** is a statement made about a policy announcement. Interpretations can

---

<sup>6</sup><https://www.intrepid-project.com/>

<sup>7</sup>For example, just comparing France with the United States, in the latter, the White House has press briefings several times a week (with press releases more frequent still) and the Federal Reserve’s Open Market Committee (FOMC) issues statements about its meetings 8 times a year; in the former, the Elysée produces press content about the president on an often-daily basis, whereas a spokesman for the prime minister presents a “compte-rendu” of the Conseil des ministres only once a week; when the Banque de France was responsible for French monetary policy, it only issued a “compte-rendu” once a year. By contrast, the European Central Bank’s press-interaction schedule is much closer to the Fed’s, as is the Bank of England’s.



be made by journalists, politicians, state officials, academics, analysts, experts, and “talking heads” of various sorts. Although interpretations may originally be made in oral form and thus can be listened to, many interpretations are written, as, for example, newspaper articles; indeed, many oral interpretations, such as legislative speeches, are at the least available as transcriptions and may originally have begun as written texts which were then read aloud. Given the significance of policy announcements, it is almost always the case that each such announcement generates a number of different interpretations: multiple news stories, political commentaries, analyst reports, or even, long after the announcements were made, academic studies.

For journalists, one of the most important interpretive forms is the news story, characterized, as per our argument above, by its four core elements. (Typically, news stories, as we use the term, are packaged within articles, with the latter containing at least one of the former. The elements of any given news story may be explicit, implicit, or a combination of the two.) Moreover, building on our earlier discussion of relevancy criteria, and taking account of the fact that most announcements are both about events and produced for journalists, we should expect that, at a minimum, announcements contain most, if not all, of the elements of news stories, and that, in many cases, they take the form of pre-written news stories. As a consequence, the journalist’s task is to gloss the elements, or the pre-written stories, in the announcements and to generate one or more news stories. We have argued elsewhere that the glossing and generation process is not a matter of applying contextual look-up rules, and perhaps ideological filters, to the announcements; rather, it is a bottom-up process in which the journalist uses syntactic, semantic, and pragmatic information from the text of the announcement, as well as from many other texts written in that language, to put together well-formed news stories (Sylvan, Arcand and Thornton 2022).<sup>8</sup>

If our argument is correct, then it should be possible to construct twinned corpora of announcements and news stories by journalists about those announcements, to use machine learning methods to train the announcements on the news stories, and hence be able to generate one or more news stories from a given announcement, both for real announcements and for counterfactual ones. In fact, as we will show, this indeed appears to be the case, and, as we will discuss in the final section of the paper, that has implications for constructing events data compendia.

One note on cases. Originally, the project was designed to explore the announcement-

---

<sup>8</sup>For example, going from this announcement:

“We already, as you know, provide non-lethal aid. We do everything we can, working with the London 11 and a group of partners, some of whom provide other things to the opposition coalition. And we will continue, as Secretary Kerry said, as President Obama said, standing next to President Hollande to look at every option that is open to us to see what else we can do to be helpful.” (State Department background briefing on Syria, February 14, 2014)

to this journalistic writeup:

“Diplomats here said the administration might consider stepping up an existing covert program to train and arm the moderate Syrian opposition or even weigh the threat of military force to compel the delivery of humanitarian aid. The senior official declined to say whether a policy shift was underway, saying options were always being reviewed.” (*New York Times*, February 15, 2014)

In this example, we can see that the journalist is not only using semantic and pragmatic information (e.g., what the U.S. and its allies are doing in Syria; what “helpful” means) but also syntactic information (the temporal scope of “will continue ... to look”).

interpretation link for two issue areas (monetary policy and foreign policy toward Russia/the Soviet Union), three countries (the United States for both issue areas, France for foreign policy, and Canada for monetary policy), both left- and right-of-center newspapers (the *New York Times* and the *Wall Street Journal*; *Le Monde* and *Figaro*; and *La Presse* and the *Globe and Mail*, respectively), and streams of announcements and interpretations that stretched from 1967 to 2017. Unsurprisingly, issues of data construction turned out to be quite involved, even prior to the arrival of the pandemic and its restrictions on archival visits, and so in the end, we ended up with two pairs of cases: monetary policy for the United States (two newspapers) and policy regarding Russia, also for the United States (again, two newspapers). A complete data set, covering the entire 1967-2018 period, has been constructed for the monetary policy-*New York Times* case, and although the Russian policy case, for the same newspaper, though for a shorter period of time, is not yet complete, it is enough so that ML training is now possible;<sup>9</sup> funding permitting, extensions, for a more limited time period, to the *Wall Street Journal*, will be done by the middle of next year.

## 3.2 Data construction

For each case, data construction involves three facets: assembling a corpus of newspaper articles on the issue area; assembling a corpus of announcements interpreted in those articles; and constructing a corpus of the articles' interpretations.

### 3.2.1 Newspaper articles

To collect candidate *New York Times* articles for the Fed case, we used a combination of two bi-grams, Federal Reserve and monetary policy, in Lexis-Nexis to download articles that might make references to Fed announcements. (Articles prior to 1980 and thus not in Lexis-Nexis required a modification of this approach.) This search strategy was motivated by limiting the number of false positives, i.e., articles related to monetary policy but not containing a reference to a current monetary policy announcement.

Once the articles were downloaded, the next step in the workflow was to human-audit the articles and keep only those containing a direct reference to a Fed announcement. We used a calendar-based approach as most potentially true positive articles are found within 2 days of a scheduled announcement. We also looked for particular terms in articles such as the FOMC or the name of the Fed chairperson at the time, which was a high probability marker of an article containing a reference to a Fed policy announcement. A second and third human audit were subsequently carried out by additional team members.

Once a list of true positive articles was assembled, the metadata for those articles were examined and opinion pieces, editorials, and (usually) extended articles in the Sunday magazine

---

<sup>9</sup>The U.S.-Russia corpus covers fewer years than the Fed case, since readily available machine-readable transcripts of White House and State Department press briefings do not extend as far back as their FOMC counterparts. As a practical matter, we truncate data construction in 1993, when the first White House website was created and when the State Department began issuing briefing transcripts on CD-ROM. However, the briefer time period (1993-2016 [following that date, the Russia election-hacking controversy overwhelms the Russia-specific articles]) is compensated for by a considerably larger number of articles per year, as well as a greater number of interpretations in each article; see below for a discussion.

were eliminated. An additional check, done by skimming over the articles themselves on the *New York Times* website, was carried out to eliminate articles which were simply transcripts of interviews, speeches, or testimony. In the end, the corpus (51 years long) contains just under 5,500 articles. After this, the articles were transformed into uniformly formatted .txt files, which subsequently were re-read to construct a master list of the articles for that year, with the entry for each article containing pointers to the announcements mentioned in the article (see below).

We are currently using a two-pass search strategy for the U.S.-Russia case. We begin with a general search with the keywords United States AND Russia AND policy, then merge the results with a specific keyword search: White House OR State Department OR <President's name> OR <Vice-President's name> OR Secretary of State [or name] OR Defense Secretary [or name] OR National Security Advisor [or name] OR Assistant Secretary of State for Eurasian Affairs [or name] or indeed any <key office bearer> AND Russia.<sup>10</sup> This two-part search ensures a reasonably extensive picture of both general foreign policy coverage and specific foreign policy coverage related to U.S. office-bearers. Following the search, a human audit is performed, along the lines of the Fed case, to see whether the articles actually touch on U.S. policy regarding Russia (for example, about what either the U.S. or other countries are doing in reaction to Russian state actions), as opposed to articles about other events, such as hockey games or power struggles in the Kremlin on which there is no U.S. announcement. At the moment, the corpus covers 24 years (1999-16) and already contains just under 2,800 articles; our projection, given the current pace of annotation, is that by the end of 2023, there should be around 4,600 articles. Keep in mind that for both cases, there is always at least one, and often more than one, interpretation in any given article.

### 3.2.2 Announcements

To assemble corpora of announcements, we began by finding the URLs of websites pertaining to various state agencies. For the Federal Reserve, this is easy (one regional Fed website, namely St. Louis, serves as a sort of repository for speeches and publications not found on the main Fed website); but for the White House and various cabinet departments, it is complicated by the fact that websites from prior presidential administrations are archived and not easily searchable. By the same token, search engines for congressional hearings, for electronic news media (even though we are excluding video- and audio-only records), and for various nongovernmental organizations (who often invite officials to speak before them) are somewhat catch-as-catch-can. Not all statements and press events are themselves conserved on even official government websites, and so we also found ourselves supplementing coverage with collections available on individual U.S. embassy websites or later Fed publications. It appears likely at this point that we will not develop procedures for gathering announcements which have themselves not been digitized. It should also be noted that the sheer length of certain announcements (e.g., the transcript of a 3-hour congressional hearing) presents particular challenges for ML parsing.

Obviously, there are large numbers of announcements that are never written about, just as

---

<sup>10</sup>Examples of other key office bearers are the Treasury Secretary, the chairman of the Joint Chiefs of Staff, the Secretary General of NATO, or the director of the CIA. It should be noted that there are significant coverage differences between searching for an individual's full name and only his/her family name; cf. Halterman et al.'s (2023b) discussion of the brittleness of dictionary-based event data coding.

there are some articles for which the announcements cannot be found. As we will discuss below, the latter, though annoying, do not pose a problem; but the former are sufficiently extensive that it makes no sense to aim at collecting them all. Instead, we perform a two-pass process: first, when the master list of articles for a given year is constructed, a team member retrieves the announcements that seem to be referred to in the articles; and subsequently, when the articles are annotated to construct the interpretations, the annotators check on the accuracy of the pointers to those announcements and add to or revise them.

More generally, it should be noted that the announcement corpus constructed for each case is explicitly linked to the article corpus. Such linkage is necessary to construct interpretation corpora (see below), but it is also a built-in germane-ness criterion simply assumed in any number of claims about putative reactions: say, stock market behavior following Fed interest rate increases, or deterrence models. (For example, Wall Street traders may not, for the most part, actually read a Fed press release: they may be reacting to an interview with an expert, or to each other's behavior. By the same token, generals or their superiors may not be attending to either the announcements or the latest moves of their adversaries, but instead executing pre-set scenarios triggered by an assessment of overall tension.) In one sense, this linkage eliminates the possibility of modeling the way in which journalists filter the information that streams across their desks, the vast majority of which, we assume, is considered by them as not immediately germane to their job of writing articles. In another sense, though, the built-in germane-ness means that the interpretation-construction task can be far more focused.

### 3.2.3 Interpretations

To construct a corpus of articles' interpretations, we annotate each article by hand, looking for self-contained portions (**Standardized sentences: StSts**) of text that seem to be about announcements; each such portion contains a news story, as we have used the term.<sup>11</sup> Once identified, each such portion is tagged with the file name of the announcement it is glossing.<sup>12</sup> After that, additional tagging takes place: for the U.S.-Russia corpus, an **actor** (usually the U.S., but possibly a group of states of which the U.S. is a member), an **act** (for example, praising, condemning, negotiating, offering to cooperate) carried out by the U.S., and a composite category, which we have labeled **Russia Link**, comprising both a second actor, Russia, and a relevancy criterion, namely what Russia had been doing either to create the situation or to complicate it. Thus, for example, the U.S. might have criticized Russia for cracking down on domestic dissent, or called on Russia to distance itself from its history of backing Assad in Syria, or negotiated with it on some issue; the U.S. could also have coordinated with allies about how

---

<sup>11</sup>Over the years of the project, we have employed a number of annotators, taking care not only to train them in annotating explicit and implicit components of interpretations, but also in having regular meetings to hash out disagreements between pairs of annotators (every article is annotated by two persons). In this way, we are able to maintain high levels of inter-annotator overlap, something which other projects at times struggle with. As regards the portioning of text, annotators are instructed to identify uninterrupted blocks of text ranging from phrases to multiple sentences in length, each of which must contain pointers to an announcement, and each of which must also contain core news story elements as per our discussion above and below. That is why these blocks of text are data containers for bona fide news stories.

<sup>12</sup>There are some intricate issues in identifying announcements when the journalist does not provide clear indications. In some cases, for example, when a journalist refers to a time period containing multiple announcements without further specification, we have not bothered to annotate the passage.

to respond to Russia regarding Iran nuclear facilities, or the annexation of Crimea. While in many cases, StSts contain text that is explicit regarding the actor, the act, and the Russia Link, this is not always true, and thus annotators are instructed to produce “labels” (to use the annotating program’s terminology) as a way of identifying news story components that are implicit in the text.<sup>13</sup>

Here, for example, is one StSt from an article appearing in the *New York Times* of June 7, 2015:

Mr. Obama began laying the groundwork on Friday in a phone call with President Petro O. Poroshenko of Ukraine, in which the White House said he had reaffirmed American support and spoken of his determination to “maintain costs on Russia and the separatists” in eastern Ukraine until they honor a cease-fire agreement that is threatening to collapse after renewed clashes last week.

Within this StSt, the token “Obama” is tagged as the actor, with a label “USA” added. The phrase beginning “he had reaffirmed” and ending “clashes last week” is tagged as the act, with a label “Will maintain sanctions on Russia until it honors ceasefire agreement for Ukraine” added. Finally, “Russia and the separatists” is tagged as the Russia Link, with a label “Russia is militarily supporting Ukraine separatists” added.

Alongside this set of tags, which are required of all bona fide StSts (i.e., they are core components of news stories), there are also optional tags; **evidence** (e.g., satellite photos) flagged by the journalist, from the announcement, as justifying the act; **motivation** for the act (what was intended to be accomplished); **temporal scope** (when the act would go into effect or how long it would last); and **attribution** (when an actor other than a designated representative of the U.S. would give his/her opinion about what should/should not be done, or have been done; this field is rarely used in the US-Russia case, presumably because Executive Branch announcers serve at the pleasure of the president; it is considerably more common in the Fed case). For each of these optional fields, annotators tag not only the wording used by the journalist but, as in the case of the core components, can and do add to those tags additional text in the form of labels, specifying what they see as otherwise only hinted at, or assumed, in the actual text of the newspaper article. While the annotators’ tagging of particular portions of text is already a vital way of identifying interpretations and their components, these “labels,” to use the annotation softwares’s term, add additional specifications to each interpretation.<sup>14</sup>

---

<sup>13</sup>Annotation of the Fed case differs in three respects from the U.S.-Russia one. First, there is obviously no Russia Link component. Second, since for the most part, the Fed is not making policy vis-a-vis another actor, but rather for the economy as a whole, the relevancy criterion for the news story revolves more around the motivations of the act – i.e., what is hoped to occur as a result – and the evidence for the act – i.e., what kind of economic news, such as housing starts or unemployment claims – led the Fed to do that. (There is another difference, having to do with the modality of the act: frequently, members of the FOMC make speeches advocating for interest rates to be raised or, less frequently, lowered, but this does not affect core news story components.) Third, the Fed case was annotated earlier in the project, and thus we were not as systematic as we could have been about including a relevancy criterion in every StSt. For an overview of the Fed case, see Marfurt et al. (2022).

<sup>14</sup>Annotation of act labels is particularly important when it comes to “should” statements: a regional Fed president, for example, who spends several paragraphs talking about inflation is clearly implying that interest rates should be raised, even though the individual might find it prudent not to say that in so many words.

As can be seen, these interpretations are act-focused, not only on what was, is, or will be done, but also on what should or should not be done. Since a single announcement can refer to more than one act, it is not only possible but common for an article to contain multiple interpretations of the same announcement (with the interpretations perhaps overlapping significantly, depending on the length of the article), and indeed for that announcement to be interpreted in multiple articles, just as it is also possible for individual articles to interpret multiple announcements. It should be noted that these types of multiplicity have several features whose significance will be discussed below: in the article, there may be references to more than one announcement by a single individual or agency; there may be references to announcements by two or more individuals or agencies; and the references may cover not only announcements made the day of, or the day before, the article's publication, but announcements made days, weeks, months, or even years before. This last type of lag, which in a sense is a way both to write and to rewrite history is particularly prevalent in U.S.-Russia articles and is worth exploring in future research; more immediately, it can be used in the ML model as an internal restriction device for the range of acts available for generating interpretations.

Several additional aspects of our annotation procedures should be noted, each of which bears on the use of human annotators. The first, and most important, is tacitness. Journalists compose their stories for a public that is assumed to have some idea about what is going on; as a result, not all the *i*'s need to be dotted or all the *t*'s crossed. An article about the Russia-Ukraine war need not mention that Western countries have been supplying Ukraine with weapons, or that certain countries are neighbors of Russia. On the other hand, it is impossible to make sense of certain stories without recalling these facts. Add the use of ironic or deliberately allusive quotations, or even of anaphora later on in long articles, and the necessity for annotators to be explicit about what is otherwise tacit is clear. As we indicated above, the annotators need not be deeply knowledgeable about the specific issues at stake; but they have to be attentive to language, and to keep asking themselves if the Russia Link entry, in particular, helps make sense of the tagged (and perhaps labeled) act-actor combination. For now, dealing with tacitness requires human annotation, even if we are hopeful that, at some point, automatic parsers can be trained to identify at least some tacit features.

A second aspect is modality. In one sense, distinguishing between acts that occurred and acts that might occur, or between what did happen and what did not happen, or, between indicative and imperative moods, is gilding the lily, given the very severe limitations of NLP machinery. However, as we will discuss below, our ML model is able to generate some modal acts and annotating articles to allow for such acts is therefore potentially useful. Note again that this is another task that lends itself well to human annotation.

Third and finally, there is the issue of context. Even casual reflection on the U.S.-Russia case will bring to mind the wide range of issues that were dealt with in the 24 years for which we have constructed data: human rights and democracy, corruption, nuclear weapons, trade and investment, NATO, Kosovo, Afghanistan, Iraq, Libya, Syria, Iran, and of course Ukraine. Compared to the Fed case, where the concern, over the entire half-century period, was to maintain price stability without permitting the economy to go into a deep recession, U.S.-Russia news stories are far more disparate. One might imagine that newspapers and other

media sources would deal with this variety by assigning the U.S.-Russia beat to a journalist with many years of experience reporting on Russia. But in fact, most *New York Times* stories are written by the White House correspondent, the State Department correspondent, the UN, or DoD, or Treasury correspondent, along with dispatches from bureau chiefs in other countries. It is also worth recalling that these positions rotate, so that, ironically, hiring students of politics as annotators of U.S.-Russia articles mirrors to a great degree the interpretive tasks faced by actual journalists. Assuredly, neither our annotators nor the Times journalists are using general, all-purpose knowledge of Russia to make sense of the announcements; rather, they are keying off of their knowledge of politics and, above all, of the English language to puzzle out what is going on in a particular context.

### **3.3 The ML task**

#### **3.3.1 General considerations**

For a given case, the data construction process thus produces three corpora: newspaper articles, interpretations of those articles, and announcements referred to in the interpretations. The task of the ML is, using both syntactic and semantic information from the announcements, to generate the interpretations, with particular attention being paid to the annotators' labels and not only to the portions of article text contained in certain fields. Before proceeding to a discussion of our ML procedures, two general observations about the ML aspect of the project should be noted.

First, the link between announcements and interpretations is, fairly often, one-to-many. As noted above, a given announcement can be referred to more than once in a given article, as well as in multiple articles. This kind of multiplicity ought not to be understood as a sign of inexactitude (although the annotators were occasionally struck by what appeared to be journalistic sloppiness), or even of differences of opinion as to the nature of the act, or of motives, or evidence. Rather, multiple interpretations of the same announcement in a particular newspaper article differ most commonly in the amount of detail they provide, as well as in the explicitness and vividness of the wording. (There can, of course, also be more analytical or speculative articles the same day, as well as follow-up articles in subsequent days.) However, interpretations may also differ with regard to modality and explicitness so that, strictly speaking, the ML task is to generate a set of possible interpretations for a given announcement, rather than only the particular ones found in the newspaper corpus. We will return to this issue below.

Second, both the number of items (articles, interpretations, announcements) in any given corpus, as well as the length of those items (number of words) is, at least by both ML and computational linguistics standards, quite small. Although, as noted above, the occasional hearing transcript may stretch on for scores of pages, in most cases, announcements rarely exceed 8 pages double-spaced. The fact that by social science data construction criteria the corpora are large, with thousands of items in each corpus, and at least in the Fed case, with the items stretching for over half a century, ameliorates matters relatively little from an ML perspective. It is for this reason that it is important to employ robust semantic information, so that terms' referents can be clearly specified and the recent or typical expectations associated with those referents made explicit.

### 3.3.2 Modeling features

We model the generation of interpretations from the announcements on which they are ostensibly based as a sequence-to-sequence generation task, employing different transformer models (Vaswani et al. 2017).<sup>15</sup> We used 80% of the interpretation-announcement pairs for training, 10% for validation and 10% for testing. Because of limitations on input tokens in the models we employed, we filtered the announcement documents into a smaller number of sentences. To capture the syntax and, more importantly, the semantics of the texts, we used BERT-style embeddings (Devlin et al. 2019); as it turned out, a jointly pretrained encoder and decoder variant of this, BART Lewis et al. (2020), worked best, particularly, in the case of Fed interpretations, at generating acts (both those mentioned in the newspaper articles and those annotated as tacit).

Our modeling approach was developed for the Fed case 12 to 18 months ago. We are now in the middle of adapting that approach to the U.S.-Russia case, a task that will continue for the rest of the year with a larger number of documents, a greater weight to Russia-Link output, and different ways of adding contextual semantics; we will also explore what happens when announcements are altered from the originals or even invented.<sup>16</sup> Nonetheless, even before such modifications are made, training a model using the same BART-based transformer yields promising results. The following pairs of paragraphs show (Rouge-1 scores, respectively, 64.76 and 70.71), first, an actual StSt, as we annotated it from a newspaper article, then a candidate StSt, generated by our model from the announcement that presumably spurred the interpretation in the article. We have put actual Russia Link labels (i.e., how our annotators characterized the [tacit] Russia-related relevance of the U.S. act) in blue and candidate labels in red.

Reference: [STD SENTENCE START] On the ABC News program "This Week" [ACTOR START] he (Powell = USA) [ACTOR END] [REFERENCE START] said [REFERENCE END], [ACT START] "We are concerned about [RUSSIALINK START] a level of authoritarianism (Russia doesnt allow free elections or a free press) [RUSSIALINK END] creeping back in the society." (The US expressed concern about Russia backsliding on democratic principles) [ACT END] [STD SENTENCE END]

Candidate: [STD SENTENCE START] In Washington, [ACTOR START] Secretary of State Colin L. Powell (USA) [ACTOR END] [REFERENCE START] said [REFERENCE END] [ACT START] [RUSSIALINK START] the United States was concerned about what he called "a creeping back into authoritarianism" in Russia (The US expressed concern about Russia backsliding on democratic principles) [ACT END] (Russia is enacting authoritarian policies) [RUSSIALINK END]. [STD SENTENCE END]

---

<sup>15</sup>For a detailed discussion of the modeling procedure and assessments of the results of the model for the Fed case, see Marfurt et al. (2022)

<sup>16</sup>As a backdoor way of handling semantics over and above the embeddings in BART, we introduced the notion of equivalence classes: phrases that, in the context of FOMC policy, had roughly the same meaning (e.g., The Fed left interest rates unchanged AND The Fed did not raise rates). Although something like this could be done for the U.S.-Russia case, the number of contexts is so much larger than in the Fed case that it may make more sense, at least as a first step, simply to have the model "look" for a proper noun like Syria or Assad or WTO. This obviously is a kludge, given that it would not tell us whether the noun in question specified the context or, instead, was used contrastively (e.g., "Libya" would not differentiate between the crisis as it unfolded and its subsequent use by the Russians as a negative analogy), but it is worth exploring.



Reference: [STD SENTENCE START] And the [ACTOR START] United States (USA) [ACTOR END] ambassador to the United Nations [REFERENCE START] called [REFERENCE END] [RUSSIALINK START] Russia’s actions in Ukraine ([Russia has militarily escalated in Ukraine](#)) [RUSSIALINK END] [ACT START] a ”threat to all of our peace and security.” (Criticized Russia’s military intervention in Ukraine) [ACT END] [STD SENTENCE END]

Candidate: [STD SENTENCE START] [ACTOR START] Ms. Power (USA) [ACTOR END] [REFERENCE START] said [REFERENCE END] [ACT START] [RUSSIALINK START] Russia’s actions ([Russia is escalating militarily in Ukraine](#)) [RUSSIALINK END] were ”a threat to the peace and security of all of us.” (The US criticized Russia for escalating militarily in Ukraine) [ACT END] (Russia is escalating militarily in Ukraine) [RUSSIALINK END] [STD SENTENCE END]

These extremely preliminary results suggest that our modeling approach has a fairly good chance of capturing the way that journalists gloss state announcements, in particular regarding the relevancy criteria used by state officials. To see this more broadly, we calculated semantic similarity (BERT-scores: Zhang et al. 2020) for pairs of announcements and the one or more StSts associated with them. This is shown below, over time, in Figure 1.

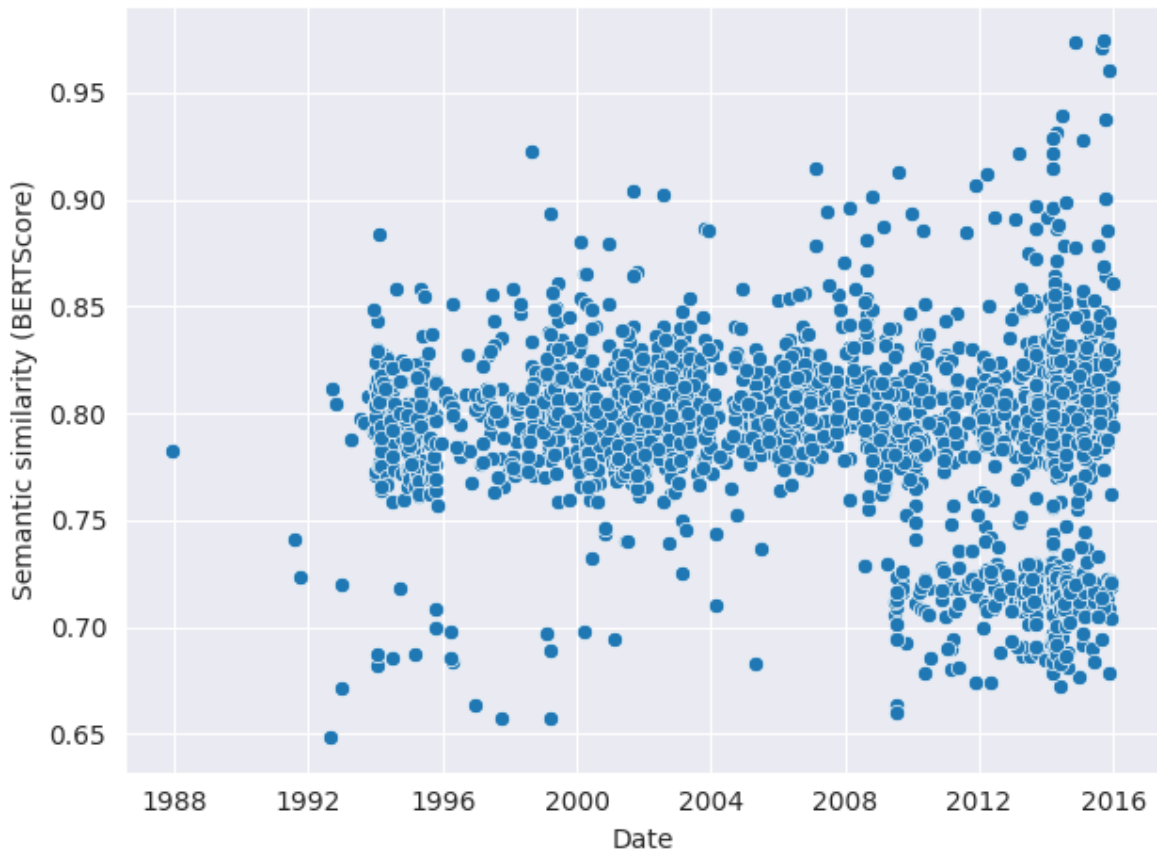


Figure 1: BERT-scores over time

As can be seen, there is a considerable degree of overlap, suggesting that there is what we could call state dependence.<sup>17</sup> We made it clear above, but it is worth reemphasizing, that such

<sup>17</sup>It is important to understand that this overlap is not an artifact of our model-training procedure, at least

overlap does not mean that journalists are self-consciously in the tank for state officials; but the overlap does mean that journalists play precisely the kind of transmission role that Lippmann wrote about.

## 4 Events Data, Take Two

Let us now return to the world of events data and see how our take on journalistic interpretation can be used to add additional structure to the coding of political events and to deal with the nearly inevitable dependence of journalists on state sources.

### 4.1 Supplementary coding

As we saw, what both state officials and journalists do when characterizing events is to include a relevancy criterion: specifying how the act in question complicates the existing situation. This can be done by specifying a goal pursued by one of the actors, or by indicating how a particular act is a response to a situation or to another act, or by establishing a temporal or historical contrast with the current situation. There are various ways of coding for this type of information. The brute-force way would be to add another slot. Thus, the U.S.-Japan-South Korea event discussed above could, in addition to the 3-tuple US-formedalliance-JapanandSouthKorea, add a fourth slot: againstChina. Similarly, to use a news item cited by Halterman et al. 2022b: 13, in which Obama and Macron met in Paris to discuss the ongoing war in Syria, one would add to US-met-France a fourth slot: discussSyrianwar. Besides an “about” slot, there are events to which one could add a “reaction” slot: for example, the news story about Chinese incursions into Taiwanese air defense space could incorporate in such a slot responseTaiwanesevisit. Nor need these sorts of additions be limited to multi-state interactions: an attack by insurgents on the state’s armed forces could add a “perspective” slot: e.g., contextbroaderoffensive, or endedlongperiodnofighting.

The problem with this way of proceeding is that one would need a large number of fourth-slot types, something seen easily from simply the four examples above (would one also need a discussXwar for every X, and analogously for every discuss and every war?). Assuming that events data scholars are committed to a pre-established set of events, this type of approach would rapidly become unwieldy.<sup>18</sup> A simpler alternative would be to realize that what a relevancy criterion does is to put into relation the “new” event with an additional event, thereby establishing the newsworthiness (complication, in Labov-speak) of the former. If, therefore, the additional event could also be coded, the coder would thus need only to specify that the “new”

---

when it comes to StSts with few or no quotations from announcements. We can imagine, somewhat in the spirit of Monty Python’s famous Hungarian phrasebook sketch, that journalists responded to State Department criticisms of Russia by writing articles about tropical islands, that they responded to White House praise of Russia by writing articles about cabarets or ice palaces, and so forth. In that case, the model might well do a good job of generating those articles without any appreciable semantic overlap between the announcements and the articles.

<sup>18</sup>In theory, it would be possible to do the kind of “bottom-up” coding we have had our annotators do, developing a fixed set of categories only after hundreds of events had been coded and the initial set of categories refined numerous times (cf. Glaser and Strauss 1967). However, that kind of approach can only work for a handful of contexts, a condition contrary to that of most events data projects.

event is linked to it, perhaps adding a particular type of linkage (e.g., reaction, anticipation, part-of).

How feasible is this alternative? In terms of task complexity, there is no particular difficulty: for the most part, one would maintain the same coding scheme, whether carried out by humans or done on an automated basis. Humans would have to be trained to spot event-coordinating phrases, such as “in response to,” but that is not complicated. The biggest change, in a sense, would be to code far enough into an article to incorporate an additional event. Our experience is that the additional event is likely to pop up within a few lines (at most one or two sentences, and quite often, the same sentence) of the “new” event. As to speed, we estimate that a trained individual coder could get through 200 linked events per week. From a budgetary point of view, that might not be feasible on a scaled-up, long-term basis, but a pilot project – say for particular contexts, such as diplomacy or multilateral security – could certainly be carried out and used for ML training purposes.<sup>19</sup> On the other hand, to incorporate our proposal in an automated coding work flow is a one-time change that offers considerable potential going forward.

What are the possible advantages of incorporating relevancy criteria into the creation of events data? Apart from the obvious one of better representing the internal structure of newsworthy events, adding linkages between events facilitates the analysis of key aspects of international politics: alliances, action-reaction cycles, proxy wars and other forms of great power intervention, and, more generally, any number of multi-actor interaction patterns (including, but not limited to, action within multilateral fora). At the moment, using 3-tuple events to address these questions is reminiscent of the attempts, decades ago, by rational choice theorists to test their theories using structural data sets such as COW.

## 4.2 Dealing with state source dependence

If it is possible, at least on a preliminary basis, to modify coding as a way of incorporating relevancy criteria, the situation as regards dependence on state sources is somewhat trickier. On our argument, one can only rarely, and with great difficulty, get at events which have not already been reported by state officials. To be sure, some state officials may well be low-level (say an embassy clerk or a staff sergeant) and/or dissident in their views; moreover, there are certainly some sources of information by people who are not part of the state. But important as these may be, the vast majority of reports used to construct international relations events stem from state officials; and if one adds soldiers and local officials, the same is also true of reports used to construct conflict events. This is why we said above that events data should, properly speaking, be construed as data about state-derived accounts of events. For this reason, it is farfetched to imagine that we can determine just how biased those accounts are as compared to events as reported by persons who are not state officials: even if the latter are picked up by a journalist, say, or by an NGO, the sheer number of times they can be pitted against the former is both too small to tell us very much and, because of that size, incapable of being used validly to correct state reports.

---

<sup>19</sup>We would also cast a vote for foreign economic policy (e.g., tariff negotiations, debt forgiveness, technological cooperation or restrictions). Not only are these events of considerable importance in international relations at least since the mercantilist era, but they are to a great degree the orphans of the events data world; there is thus to a degree less to unlearn in coding them for inter-event linkages.

What can be done, however, is to pit state sources against each other. At present, the standard practice in assembling events data compendia is to impose a “one-a-day’ deduplication” constraint in which “the same actor-recipient-location triplet can only exhibit a maximum of one event of a certain type(-mode-combination) on a particular calendar day” (Halterman et al. 2023b: 18). The reason for this is obvious: so that events are not multiply-counted. However, if there are multiple reports, each sourced differently, of what is presumably the same event, they could be assessed on any number of grounds (for example, whether certain actors are portrayed as initiators of events as opposed to targets, or how acts are described).<sup>20</sup> The crudest way of doing this would be to compare sources; for example, for a sample of events, how each of those events is constructed by wire services or newspapers from different countries. A more appropriate technique, albeit a considerably more complicated one, would tag the nationality of the sources cited in the articles: for example, the infamous “senior Administration official” in the United States as compared, say, with an official in the Quay d’Orsay or the Kremlin.<sup>21</sup> This sort of information certainly could be coded for, whether manually or automatically, but doing so would in effect involve another slot in any given event entry and, more importantly, would often require reading fairly far into the article. In addition, of course, there would have to be a concerted effort made to, in effect, parallel-code a large number of articles.

If an effort like this were, however, to be carried out, there would be a notable side-benefit over and above the determination of the particular way that one country’s officials construct events as compared with another’s. That benefit is a degree of transparency regarding the choice of sources. Assuming that there are in fact multiple sources for a significant number of events, and assuming moreover that some sort of deduplication rule is in place, entries in an events data compendium must either be stitched together from multiple sources – in which case, it would be useful to know how that stitching occurred – or else based on a single, presumably “best” source – in which case, it would be useful to know why that source was picked over the others. For now, people outside a given events data project have little to no information about these issues or, for that matter, the nitty-gritty of various other steps in the data construction process; and even when the latter type of information is available, copyright or secrecy considerations make it difficult to consult the original articles on which the compendium was built. This in turn means that coding efforts are hard to replicate (as opposed to coming up with an alternatively constructed compendium), which is worrisome. For these reasons, comparative source analysis, of the sort discussed above, has significant side-benefits.

## 5 Conclusion

As we pointed out at the start of this paper, there has in recent years been enormous progress in the number and sophistication of events data compendia constructed in political science. However, there are two problematic features of those construction efforts: an overly stripped-down conception of events, and a deep and necessary reliance on state sources of information.

---

<sup>20</sup>Chan et al. (2023) do something similar, but even if one accepts the sentiment analysis technique they employ, the problem is that they compare coverage across all events in the aggregate, rather than compare coverage within events.

<sup>21</sup>We could also imagine doing this for high-level vs. low-level officials, but the frequency with which the latter are both identified and cited is too low for such an exercise to be performed systematically.

Both of these issues stem from routine practices in the way that journalists decide what to write about and how they in fact write. For the last few years, we have been carrying out a research project addressing these issues directly. Our efforts, which permit us to generate journalistic interpretations from state announcements, suggest that journalists, in constructing news stories, include as a matter of routine relevancy criteria explaining why certain events are in fact news; we also have evidence showing just how strongly journalists rely on state officials for constructing their stories. Applying our findings to events data construction, we therefore make several suggestions about how to expand the conception of events used in constructing compendia, and how to assess the role played by different countries' state officials in the construction of what eventually becomes events data.

As indicated above, our project is not yet completed, and so our arguments in this paper should be seen as provisional. Nonetheless, all the evidence we have from our project, from studies of journalism, and from various events data construction efforts, is that it is vital for political scientists, in working with events data, to take account explicitly of its necessarily journalistic character.

## References

- Althaus, Scott, Buddy Peyton, and Dan Shalmon**, “A Total Error Approach for Validating Event Data,” *American Behavioral Scientist*, 2022, 66 (5), 603–624.
- Chan, Chung-hong, Hartmut Wessler, Marc Jungblut, Kasper Welbers, Scott Althaus, Joseph Bajjalieh, and Wouter van Atteveldt**, “Challenging the Global Cultural Conflict Narrative: An Automated Content Analysis on How Perpetrator Identity Shapes Worldwide News Coverage of Islamist and Right-wing Terror Attacks,” *International Journal of Press/Politics*, 2023.
- de Mesquita, Ethan Bueno, C. Christine Fair, Jenna Jordan, Rasul Bakhsh Rais, and Jacob N. Shapiro**, “Measuring Political Violence in Pakistan: Insights from the BFRS Dataset,” *Conflict Management and Peace Science*, 2015, 32 (5), 536–558.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova**, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in “Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)” Association for Computational Linguistics Minneapolis, Minnesota June 2019, pp. 4171–4186.

- Douglass, Rex W., Thomas Leo Scherer, J. Andrés Gannon, Erik Gartzke, Jon Lindsay, Shannon Carcelli, Jonathan Wilkenfeld, David M. Quinn, Catherine Aiken, Jose Miguel Cabezas Navarro, Neil Lund, Egle Murauskaite, and Diana Partridge, “Introducing the ICBe Dataset: Very High Recall and Precision Event Extraction from Narratives about International Crises,” arXiv preprint 2022.
- Gans, Herbert J., *Deciding What’s News: A Study of CBS Evening News, NBC Nightly News, Newsweek, and Time*, New York: Random House, 1979.
- Garfinkel, Harold, *Ethnomethodology’s Program: Working Out Durkheim’s Aphorism*, Rowman and Littlefield, 2002.
- Glaser, Barney G. and Anselm L. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine, 1967.
- Halberstam, David, *The Making of a Quagmire: America and Vietnam During the Kennedy Era*, rev ed., Lanham, MD: Rowman and Littlefield, 2008.
- Halterman, Andrew, Benjamin E. Bagozzi, Andreas Beger, Philip A. Schrodtt, and Grace I. Scarborough, “PLOVER and POLECAT: A new political event ontology and dataset,” 2023. Annual Convention of the International Studies Association.
- , Philip A. Schrodtt, Andreas Beger, Benjamin E. Bagozzi, and Grace I. Scarborough, “Creating Custom Event Data without Dictionaries: A bag-of-tricks,” 2023. Annual Convention of the International Studies Association.
- Herman, Edward S. and Noam Chomsky, *Manufacturing Consent: The Political Economy of the Mass Media*, New York: Pantheon, 1988.
- Kalyvas, Stathis N., “‘New’ and ‘Old’ Civil Wars: A Valid Distinction?,” *World Politics*, 2001, 54 (1), 99–118.
- Labov, William, *The Language of Life and Death: The Transformation of Experience in Oral Narrative*, Cambridge University Press, 2013.
- and Joshua Waletzky, “Narrative analysis,” in June Helm, ed., *Essays on the Verbal and Visual Arts*, University of Washington Press, 1967, pp. 12–44.
- and – , “Oral versions of personal experience: Three decades of narrative analysis,” *Journal of Narrative and Life History*, 1997, 7 (1–4), 3–38.
- Lewis, Mike, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer, “BART: Denoising Sequence-to-sequence Pre-training for Natural Language Generation, Translation, and Comprehension,” in Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, eds., *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 7871–7880.
- Lippmann, Walter, *Public Opinion*, New York: Harcourt, Brace and Company, 1922.
- Manning, Chelsea, “The fog machine of war,” 2014.
- Marfurt, Andreas, Ashley Thornton, David Sylvan, Lonneke van der Plas, and James Henderson, “A Corpus and Evaluation for Predicting Semi-structured Human Annotations,” in Antoine Bosselut, Khyathi Chandu, Kaustubh Dhole, Varun Gangal, Sebastian Gehrmann, Yacine Jernite, Jekaterina Novikova, and Laura Perez-Beltrachini, eds., *Proceedings of the 2nd Workshop on Natural Language Generation, Evaluation, and Metrics (GEM)*, Association for Computational Linguistics 2022, pp. 262–275.

- Nippert, Colin**, “From a Grunt’s Eye View: Towards a Grounded Theory of International Humanitarian Law Violations.” PhD dissertation, Graduate Institute of International and Development Studies 2012.
- Stundal, Logan, Benjamin Bagozzi, John R. Freeman, and Jennifer S. Holmes**, “Human Rights Violations in Space: Assessing the External Validity of Machine-geocoded Versus Human-geocoded Data,” *Political Analysis*, 2021, *31* (1), 81–97.
- Sylvan, David, Jean-Louis Arcand, and Ashley Thornton**, “Modeling Journalistic Framing of State Announcements: Using NLP to Generate Characterizations of Acts,” European Political Science Association 2022.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin**, “Attention Is All You Need,” arXiv 2017.
- Weidmann, Nils B.**, “On the Accuracy of Media-based Conflict Event Data,” *Journal of Conflict Resolution*, 2015, *59* (6), 1129–1149.
- , “A Closer Look at Reporting Bias in Conflict Event Data,” *American Journal of Political Science*, 2016, *60* (1), 206–218.
- White, Hayden**, “The Value of Narrativity in the Representation of Reality,” *Critical Inquiry*, 1980, *7* (1), 5–27.
- Windsor, Leah C.**, “Bias in Text Analysis for International Relations Research,” *Global Studies Quarterly*, 2022, *2* (1), 1–13.
- Zhang, Tianyi, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi**, “BERTScore: Evaluating Text Generation with BERT,” in “International Conference on Learning Representations” 2020.