Foreign Policy and Journalistic Assumptions: Incorporating Background Semantics Into Machine Learning Models of Event Interpretation*

David Sylvan† Jean-Louis Arcand‡ Ashley Thornton§

March 25, 2022

Abstract

When journalists characterize events as part of a pattern – a speech, for example, seen as the latest escalation in a conflict, or a smiling joint press conference categorized as evidence that each side made concessions – the characterization cannot possibly specify all the assumptions that underlie this interpretation. To some degree, this is because of writing conventions for how news stories are put together, but it is also due to journalists assuming a certain background semantics: about a country's history, for example, or a leader's political style, or for that matter, about the meaning of certain specialized words or phrases ("testing"; "sovereignty"). Thus, understanding how journalists interpret events is not in the first instance a matter of scholars slotting phenomena into pre-existing schemas such as "cold war" or "great power confrontation," but of specifying the background semantics assumed in who-what-where-when-why accounts of those phenomena. We illustrate the importance of this specification by building machine-learning models to generate elite newspaper summaries of policy announcements by the Federal Reserve about monetary policy, contrasting models with different types and degrees of semantic detail, and with an eye to extending those models to generate newspaper summaries of policy announcements by the White House and the State Department about Russia.

Keywords: Foreign policy, political economy, methodology

1 Introduction

This paper, and the project underlying it, starts from the observation that state agencies regularly issue announcements about the policies which they are carrying out, and that those announcements have significant political and economic consequences even before the policies can be partly or completely implemented. These consequences are quite varied, depending, but only to some degree, on the type of announcement and, more generally, on the issue domain (a

*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 in data construction and analyis; and Andreas Marfurt, Alireza Mohammadshahi, and James Henderson for their work on machine learning models; and we gratefully acknowledge the support of the Swiss National Science Foundation under grant CRSI I5_180320. Prepared for delivery at the Annual Convention of the International Studies Association, March 28 - April 2, 2022.

†Department of International Relations / Political Science, Graduate Institute of International and Development Studies, Geneva, Switzerland. Email: david.sylvan@graduateinstitute.ch

‡Department of International Economics, Graduate Institute of International and Development Studies, Geneva, Switzerland. Email: jean-louis.arcand@graduateinstitute.ch

§Intrepid Project, Graduate Institute of International and Development Studies/IDIAP, Switzerland. Email: ashley.thornton@graduateinstitute.ch

point to which we return below). Thus, investors may react to certain announcements about interest rates or the curent monetary policy stance (e.g., quantitative easing), or for that matter to announcements about the escalation of a military standoff, by buying or selling stocks, bonds, or other financial instruments; diplomats may shift or reinforce their negotiating positions in reaction to an announcement about an impending vote in the Security Council; lobbyists may concert their efforts to overturn a given policy in reaction to any number of different types of announcements; and of course, members of the public may decide to vote for or against a given political party on the basis of what is, or is not, being said by spokespersons for the governing coalition. All of these specific phenomena have been studied for many years.

What those studies miss, however, is the precondition necessary for each of these phenomena to occur, namely, that for state agencies' announcements to be reacted to, they must first be interpreted and that those interpretations are carried out, more often than not, by individuals specialized in following those agencies. For example, stock brokers have neither the time nor perhaps the technical command of macroeconomic theory to read through a 20-page memorandum in which the Federal Reserve's Open Market Committee (FOMC) details what went on at its monthly meeting some six weeks earlier. Instead, they have to rely on wire service and analysts' reports or claims by experts about what the Fed's announcement means: what the Fed is doing now, or planning on doing in the future, or will not be doing, what its priorities are, and so forth. Similarly, most diplomats do not necessarily read through every word of the White House press secretary's daily briefing; instead, they read the *New York Times* articles on that subject as well, of course, as memos by colleagues charged with keeping tabs on the White House (the latter, in fact, usually rely on the press as well). Our concern here is with these necessary and ubiquitous acts of interpretation by various specialists, particularly journalists assigned to the coverage of particular state agencies.

In our view, it is impossible to imagine the modern state without both announcements and interpretations of those announcements. On the one hand, the range of activities dealt with by the state is far greater than a century ago, with many of those activities being carried out by multiple agencies and, to boot, being highly technical in nature. On the other hand, those same activities are aimed directly at numerous "attentive" publics: from investors and officials of foreign governments to legislators and potential voters. In this regard, journalists are privileged interlocutors, focused on intently by politicians and agency personnel, with the announcements being crafted for maximum impact and with those announcements themselves being the object of additional announcements (e.g., "off the record" briefings of journalists about press releases). Journalists, in this sense, are interpreters who, on a daily basis, select, gloss, and repackage state announcements. This paper is a progress report about a research project on that interpretation process, and in particular, on how, in that project, we address the issue of journalists' (and elites', more generally) background semantics.

Specifically, the paper is divided into several parts. We begin by discussing, in an abstract fashion, the concepts of announcement and interpretation, and argue for an alternative way of researching interpretation. Ever since Lippmann's seminal work (1922) on public opinion, the standard account of how journalists carry out interpretation is that they engage in mapping, taking announcements and rewriting them, so to speak, as a series of claims about what is, or what might be, or what will not be, happening, and why. This mapping is specific, both with respect to domain (for instance, an announcement by a central bank is assumed to be conditioned by fear of inflation, whereas an announcement about a military ally is assumed to be directed implicitly at an actual or potential adversary) and, within domains, by topical context (for example, in the single domain of foreign policy, both the referent and the valence of terms such as "tightening" or "hardliner" differ by issue [say, trade negotiations vs. blockades], speaker, and time period). There are, however, several distinct problems with the mapping account of journalistic interpretation, notably that there are simply too many possible contexts (about a leader, a country, a time period, an act or a set of acts, etc.) to be specified a priori. Moreover, the background semantics that journalists (and for that matter, readers) are required to have are difficult to specify as a list, no matter how extensive, of knowledge claims. We thus propose a non-mapping, machine-learning (ML), textual entailment approach that bootstraps up from semantic (and also syntactic) information in both the text and in multiple other texts.

We then turn to the details of our alternative, laying out the data construction of both interpretations and announcements, as well as the specific method of assessing how the former are linked to the latter. We discuss some specific issues about the kinds of syntactic and semantic information needed in the ML process, then go over how our ML model-building efforts have dealt with those issues. Although the data for evaluating both the validity of our proposed interpretation-announcement link, as well as the proto-hypotheses on domain specificity, are still in the process of being constructed, preliminary analysis is intriguing.

2 Interpretation: mapping vs. entailment

To start with, some terminology. A **policy announcement** is a written statement issued by a state agency describing its actions or decisions. These 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 considerably in the frequency, formality, and extent of their announcements,[1] they nonetheless do in fact make such announcements on a regular basis for many years. As we will see below, this regularity makes it possible to construct cor-

[1] For example, just comparing France with the United States, in the latter, the White House has press briefings several times a week and the FOMC issues statements about its meetings 8 times a year; in the former, the spokesman for the prime minister presents a "compte-rendu" of the Conseil des ministres once a week, whereas when the Banque de France was responsible for French monetary policy, it only issued a "compte-rendu" once a year.

pora 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 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. As we argued above, journalists' interpretations are of particular interest; and as we will see below, if individual announcement corpora can be aligned with associated interpretation corpora, such as streams of newspaper articles about the announcements, we thus would have naturally-occurring training data for ML models of interpretation.

One standard way to conceptualize the link between announcements and interpretations is to see the latter as a mapping of the former onto an already-structured domain. For example, if the chairman of the Federal Reserve gives a press conference and says that "It's just a lot of people who need to get back to work, and it's not going to happen overnight, it's going to take some time," and if a journalist glosses this as the Fed being "willing to allow inflation to move higher without reacting" (*New York Times*, 17 March 2021), then the journalist can be seen as mapping the chairman's language about people needing to get back to work onto a domain characterized by a presumed tradeoff between inflation and unemployment, an additional presumed Fed preference for low levels of inflation, a further presumed Fed default policy of raising interest rates to maintain those low levels, and finally, in that domain, a surprising inference (hence, "news"): namely, that the chairman's words seemed to be signaling that the Fed is acting in a way opposite to its default policy.

This way of thinking about interpretation as mapping is reminiscent of classic work on belief systems as "an interrelated set of affect-laden cognitions concerning some aspects of the psychological world" (Abelson and Carroll 1965: 24; see also Abelson 1971, 1973, 1979), such that particular events are mapped onto instances of generic states of affairs (e.g., North Korea attacks South Korea in June 1950 is an instance of Communist aggression against the Free World), with the default action implications of the latter (e.g., if there is Communist aggression against the Free World, then stand firm) being in turn remapped onto specific policy positions in this particular situation (e.g., send troops to South Korea as an instance of standing firm). Unsurprisingly, the mapping approach to interpretation has been picked up in computer science (e.g., Carbonell 1978), political psychology (e.g., Jost, Federico and Napier 2009), foreign policy analysis (e.g., Taber 1992), and, of particular note for this discussion, studies of journalism (e.g., Gans 1979; Herman and Chomsky 1988; Gamson and Modigliani 1989; Carvalho 2007; Matthes 2009).

There are two interrelated problems with thinking about interpretation as mapping. The journalist is acting akin to a function that goes from the words of the announcement to the rewrite of the news article, but what actions or perceptions by the journalist would that function entail? We might imagine that the journalist is using topical rules of some sort: for example, if a Fed chairman talks about pressure on wages, he is referring to inflation, whereas if he talks about the economy being weak, he is referring to unemployment. However, those rules will not be all-purpose, because the chairman could specify that in the current context, pressure on wages is likely to be short-term and limited. One would thus add a second set of contextual rules; but this does not solve anything, because if there are multiple possible contexts, then one needs meta-contextual rules, *ad infinitum*. Quite apart from the practical impossibility of storing and searching among even a large finite number of rules, the more serious problem is that there is no evidence at all to suggest that many contexts had any rules specifying the conditions of their use: which financial journalist, prior to 2020, could possibly have had rules for how to interpret Fed action during a worldwide pandemic?

Note that the contextual rules problems that arise in thinking about interpretation as mapping revolve around the issue of domain specificity. Consider, for example, a statement by the Fed chairman about "tightening": normally, this would be mapped onto a monetary policy issue domain, concerned with interest rates and the money supply. But if the comment occurs when talking about different topics such as bank supervision, or exchange rate policy, or financial sanctions against officials of foreign governments, then the topical context, and hence the meaning of "tightening," would be different. Indeed, even in the topical context of monetary policy, if the Fed were trying to induce banks to lend more freely, had made funds available for low-interest loans, and then discovered that banks were not in fact lending as much as desired, it might "tighten" the conditions under which it made those funds available to banks. Thus, the practical impossibility of specifying a set of rules that both define all conceivable contexts for mapping terms and the conditions under which one (or more?) contexts apply is due to the fact that interpretation is at the least domain-, and most likely topic-specific. Perhaps unsurprisingly, most linguistic studies which directly address the content and differences between domains (e.g., work on semantic frames: Fillmore, Johnson and Petruck 2003) tend to proceed in either an extremely abstract (e.g., in Framenet, Attaching, or Communication) or a generic (e.g., conflict, economic consequences, human interest, or morality: Semetko and Valkenburg 2000) fashion, or else simply omit "most of the words [e.g., 'blatant'] that one confronts in naturally occurring text" (Pavlick et al. 2015: 408).[2]

In fairness, it should be pointed out that these problems, which have been known in the field of artificial intelligence for decades (e.g., Davis and Marcus 2015), have sparked numerous proposals, ranging from "case-based" or analogical approaches to approaches based on combining "common-sense" reasoning with neural networks (e.g., Hwang et al. 2020). Our approach, though closer to the second, is based critically on textual entailment, i.e., the semantic implications between texts (Dagan, Glickman and Magnini 2006). Recall the above example about Fed "tightening." Here, what a journalist would presumably do is to use a) the everyday meaning of "tightening," b) sentences before and after about what will not be done, c) syntactic infor-

[2]Similar problems are apparent in the foreign policy literature on "substitutability" (e.g., Most and Starr 1984; Palmer and Bhandari 2000). On the other hand, as we will discuss below, the combinatorial power of ML techniques makes it possible, at least in principle, to use certain types of frame information to address contextual specificity issues.

mation about the modality of the verb tighten, its direct object, and its target, and d) prior information (perhaps in the announcement, or in earlier texts) about what the Fed is currently doing about monetary policy, and, perhaps, about the Fed chairman's presumed preferences, in order to construct a summary of the who-what-when-why-type information expressed in the phrases or sentences about tightening. However, since the text of the chairman's statement contains large amounts of syntactic and semantic information, even when that statement is not trying to be vague or obfuscatory,[3] the number of combinatorial possibilities for the use of that information is simply too great to be specified *a priori*.

But beyond the semantic information contained in the text of the announcement and, perhaps, the texts of earlier announcements, the journalist will also need to access a larger set of background semantics in order to make sense of the announcement. Consider our earlier allusion to the Fed's actions when the Covid-19 pandemic first hit the United States. On 9 April 2020, i.e., some 5 weeks after the Fed began adapting policy to the pandemic, it announced a raft of new policies, including facilities for channeling credit to banks making small business loans and for buying short term notes from states and municipalities. In the announcement, the Fed's chair, Jerome Powell, stated that "Our country's highest priority must be to address this public health crisis, providing care for the ill and limiting the further spread of the virus." For a journalist to interpret this particular sentence, it would, at the minimum, have required the background semantics of how a public health crisis, combated by social distancing and involving many persons flooding hospitals, would impinge on the economy (for example, via higher unemployment rates, lower tax revenues, and increased demand on public health facilities). And in fact, without missing a beat, the *New York Times* article a few hours later referred explicitly to "the severe damage to the economy as quarantines keep workers at home and grind entire sectors to a standstill." This goes well beyond domain specificity, and points to the need to incorporate an entire range of background semantics.

Nor is this an isolated case. A dozen or so years before the pandemic, the Fed, faced with the 2008 financial crisis, began carrying out "quantitative easing" and other policies designed to deal with "zero lower bound" phenomena that, until then, were discussed by only a handful of interpreters. Or take Russian foreign policy: since the invasion of Ukraine, US announcements about Russia have addressed the issue of Ukraine being "a hotbed for neo-Nazism," to quote from a State Department press release on 3 March. In other words, journalists will, as a matter of routine, need a range of background semantics, across widely varying issue domains, in order to interpret state announcements. Added to the meta-contextual rules problem discussed above, it should therefore be clear that in order to incorporate those background semantics into a model of journalistic interpretation, one cannot even try to start listing all the information drawn on by journalists.

Instead, to capture how journalists key off of multiple textual elements to interpret announcements, we propose to combine human coding to specify particular interpretations in news

[3] The former chairman of the Federal Reserve Board, Alan Greenspan, famously said that "I guess I should warn you, if I turn out to be particularly clear, you've probably misunderstood what I said." Much the same can be said of diplomatic announcements: officials spend considerable amounts of time using coded language designed to be understood by insiders but not necessarily by others, as in the Treaty of Breda (10 CTS 231) ending the Second Anglo-Dutch War, which famously ceded the island of Manhattan without ever saying so.

articles about announcements, then to use ML techniques, incorporating both background semantics from a wide range of texts and at least elementary syntactic information. The resulting model, which will be trained on coding-announcements pairs, will be used to generate, and assess the accuracy of, codings from other pairs (and, down the road, to generate codings from new or counterfactual announcements). Below, we will discuss this approach in some detail, but note for now that issues of domain and topical specificity are to a considerable degree obviated: the researcher need not establish in advance a taxonomy of interpretive contexts, nor the features of any such contexts. To be sure, a wildly heterogeneous corpus of news articles may make it more difficult both to code and to use ML techniques on that coding, but that is an empirical question, and in fact, the approach enables human coders' intuitions at least about domain specificity (say in sorting articles as between those on the Fed and those on relations with Russia) to be checked, an issue to which we return at the end of this paper.

At heart, our approach glosses interpretation as an interlocked set of entailment relations between texts (Dagan et al. 2006; Gupta et al. 2014; Henderson 2017; Pasunuru et al. 2017; Tatar et al. 2008). The journalist is given the text of the announcement, attends to particular portions of that text, and rewrites those portions in order to clarify matters of who, what, when, where, why, and how. In doing this, she uses syntactic and semantic knowledge about the text of the announcement, texts of other announcements, and large numbers of other texts that are not announcements at all.[4] However, because, in Chomsky's famous phrase, there is an uncountably large amount of information in these texts, there is no way to adjudicate among the interpretive possibilities on a priori grounds; instead, we use ML approaches to train representations of announcement portions on actual newspaper interpretations of those portions, with the latter having been human coded in order to specify both their extent and composition, as well as the additional interpretive components left implicit in the articles.

As indicated above, the ML techniques will use not only semantic but syntactic information as well. We will return to this point several times below, but for now, simply consider two examples of journalistic interpretation, one about a Federal Reserve press release, the other about a State Department background briefing on Syria. Example 1:

> "The F.O.M.C. meeting ended at 12:55 PM; there is no further announcement." (Federal Reserve press release, December 22, 1998)
>
> "The Federal Reserve voted today to hold interest rates steady, judging that three rate reductions in the fall had given the economy a big enough kick to keep it growing at a healthy pace into the new year." (*New York Times*, December 23, 1998)

In this example, we can see that the journalist is using information about what the F.O.M.C. can do (e.g., raise interest rates) as well as what it usually does (i.e., make no announcement when it is neither raising nor lowering rates); the journalist also uses information about the world (how the economy is doing) and what the Fed did over the last few months (three reductions in the fall). Now example 2:

> "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

[4]The journalist will also use pragmatic information, but that is beyond the scope of this project.

> 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)
>
> "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 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"). It is precisely this type of NLP information that we want to employ, and that standard "bag of words" methods will tend to miss.

3 An alternative approach

The two aspects of our alternative involve human construction of data on, and machine learning of the relation between, announcements and interpretations. Originally, the research project was designed to explore the announcement-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 Russian policy, 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 we expect to have the Russian policy case, for the same newspaper, though for a shorter period of time, completed by the end of this year;[5] extensions, for a more limited time period, to the *Wall Street Journal*, will be done by the middle of next year.

[5]The 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 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.

3.1 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.

3.1.1 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 candidate true positive articles are found within 2 days of a scheduled announcement. We also looked for particular terms in the article 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 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.[6] 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

[6]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.

corpus covers 16 years (2001-16) and already contains 3,389 articles; if, as per footnote 4, the remaining 8 years (1993-2000) cover U.S. policy toward Russia with the same frequency, we will end up with close to the same number of articles (and, as discussed below, a greater number of interpretations) as for the Fed case.

3.1.2 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 there are some articles for which the announcements cannot be found. The latter, though annoying, do not pose a problem, as we will discuss below; 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 second 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 relevancy 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 irrelevant. In another sense, though, the built-in relevancy means that the interpretation-construction task can be far more focused.

3.1.3 Interpretations

To construct a corpus of articles' interpretations, we annotate each article, looking for portions of text that seem to be about announcements. Each such portion is in turn tagged with minimal journalistic information: a) a pointer to the document containing the announcement,[7] a textual sub-passage containing b) the journalist's identification of who (e.g., the Fed or the U.S.) is either acting or whose (possible) action is being discussed, another sub-passage containing c) the nature of the action (e.g., raising interest rates; imposing sanctions), and, d) in the case of policy toward Russia, a final textual sub-passage containing an explanation of why the act in question is relevant to U.S. policy toward Russia (Russia Link). In addition, there are optional fields in the textual portions: e) evidence (e.g., the rate of inflation) flagged by the journalist, from the announcement, as justifying the act; f) motivation for the act (what was intended to be accomplished); g) temporal scope (when the act would go into effect or how long it would last); and h) attribution (when an actor other than the Fed 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). Since in many cases, the journalist's wording may be obscure, or because she may be assuming knowledge that is not explicitly stated, annotators systematically add their own characterization of both the action and the Russia Link; they may also add their own characterization of the other fields.

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. As we will discuss below, this last type of lag is of particular interest, signaling far more than just the range of the journalist's memory or the extent of her files.

3.2 The ML task

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' character-

[7] 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.

ization entries and not only to the portions of article text contained in certain fields. Several general observations about the ML aspect of the project should be noted; below, we will go over the details of the ML, with special emphasis on the incorporation of semantic and syntactic information.

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 differ most commonly in the amount of detail they provide, as well as in the explicitness and vividness of the wording. 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 develop robust semantic information, so that terms' referents can be clearly specified and the recent or typical expectations associated with those referents made explicit.

Third, it is important to understand the biases inherent in using newspaper articles as the basis for interpretations. Journalism involves "deciding what's news," to use the title of the Gans book (1979) mentioned above, which is to say that continuation of the status quo ante is not normally written about. This may be one reason why, when nothing new happens on U.S.-Russia relations, there are few articles about the announcements of routine activities. On the other hand, even when the Fed decides not to change interest rates, that is still news, albeit less written about. As we will see below, this points to a possible distinction in the two issue areas; for now, it means that journalists are faced with less of a filtering issue regarding Fed announcements than in the case of U.S.-Russia policy; from an ML perspective, the latter are therefore likely to be considerably more complex than the former.

4 Syntactic and semantic information

We indicated above that for ML techniques to succeed in generating interpretations from announcements, they would need to incorporate both syntactic and semantic information. Consider first the issue of syntax. On 10 July 1986, the Fed put out a press release which began in this way: "The Federal Reserve Board announced a reduction in the discount rate from 6.5

percent to 6 percent, effective on Friday, July 11, 1986. The action, conforming in part to recent declines in a number of market interest rates, was taken within the framework of the generally accommodative stance of monetary policy that has prevailed for some time. More specifically, the action appeared appropriate in the context of a pattern of relatively slow growth, comfortably within capacity constraints, in the United States and in the industrialized world generally. That pattern has been accompanied by relatively low prices of a number of important commodities and greater stability in prices of goods generally."

The next day, the *New York Times* published an article whose lede was as follows: "Citing a 'relatively slow' economy and low inflation, the Federal Reserve Board today cut its benchmark lending rate to 6 percent from 6 1/2 percent. The reduction, the third this year, brought the crucial discount rate to its lowest level since the start of 1978." Putting aside, for the moment, semantic considerations throughout the announcement and the article (e.g., "generally accomodative stance ... that has prevailed for some time" as roughly equivalent to "The reduction, the third this year"), there is a clear syntactic dependency relationship in the third sentence of the press release, whereby "the action" (linked anaphorically to the preceding and, more importantly, the first sentence) is coordinated in that sentence with "a pattern of relatively slow growth" and, in the following (fourth) sentence (via "a pattern"), with "relatively low prices." In other words, in order to generate the "because" of the third and fourth sentences (arguably, a vital part of the journalist's interpretation), the ML parser needs to have syntactic information inputted to it.

More subtly, but of perhaps greater importance, the parser also needs to have information about the modality of certain verb phrases. For example, on 20 March 2014, Obama announced that his administration had "been working closely with our European partners to develop more severe actions that could be taken if Russia continues to escalate the situation. As part of that process, I signed a new executive order today that gives us the authority to impose sanctions not just on individuals but on key sectors of the Russian economy. This is not our preferred outcome. These sanctions would not only have a significant impact on the Russian economy, but could also be disruptive to the global economy." The first quoted sentence has a verb in the conditional tense ("could") which is clearly linked to a possible situation, namely that of further Russian escalation. Interestingly, although the *New York Times* article captured the conditional nature of the broader sanctions mentioned by Obama, it did not tie them to further Russian escalation. Indeed, of the first 10 Google results (searching obama ukraine sanctions preferred), only Reuters and C-SPAN captured that linkage; the BBC and the *Atlantic* interpreted the announcement the same way as the *New York Times*, i.e., simply mentioning that Obama had given himself the power to impose additional sanctions; and *USA Today* conflated the possible sanctions with the new sanctions Obama announced that day that he was putting into effect. These disparate interpretations were particularly striking in light of a not-for-attribution general telephone briefing that took place less than five minutes after Obama finished making his statement; in that briefing, "senior administration official number one" reiterated that the executive order was a way of "preparing for potential future consequences on the Russian government ... if Russia further escalates this situation."

Our intent in this example is not to criticize journalists for not listening carefully (presumably their attention was somewhat distracted by the sanctions that were actually imposed, by speculation about how the Russians would react to those sanctions, and by the continuing drumbeat of speculation about whether the U.S. would send military aid to Ukraine), but to indicate that modal verbs lend themselves to multiple interpretations, not only because of simplification (e.g., dropping the "if" clause) or because of the logical possibilities implied in such verbs (e.g., that Obama has not ruled out imposing those additional sanctions), but because of the cognitive difficulty of parsing modal sentences. For this reason, particularly in the case of sentences with modality, what we are aiming for from the ML process is that it generate a range of possible interpretations: what the Fed, or the U.S. government, is doing, what those actors might do, and what those actors are probably not going to do.

Thus, the announcement input to the ML trainer will encompass three types of syntax-like (if not all strictly syntactic) tags: dependency relationships, modality, and cross-sentence anaphora. Announcements (or 500-token chunk windows, for particularly lengthy announcements) will be trained on the multiple interpretations linked to each announcement. The results presented below are based on only one of those tags, dependency relationships (see Mohammadshahi and Henderson 2021 for details on the approach), but we will endeavor to add the other two types in the coming months.

Consider now (once again) the issue of semantics. In the spring of 2014, one of the many actions taken by the United States, along with NATO, during the Ukraine crisis was the sending of fighter jets to patrol the airspace above the Baltic republics and Poland. This action was presented by U.S. officials as a response to various Russian moves regarding Ukraine (annexation of Crimea, presumed support for separatist armed groups in the eastern part of the country). Interestingly, the intended relevance of U.S. and NATO action was almost never specified in the newspaper articles: the history of Russian relations with these countries, the fact of their NATO membership, and the possibility that Russia might carry out actions against these countries similar to what it had been doing in Ukraine. As we indicated above, some of this information is added by annotators as part of the data construction process in which they specify the contents of a given newspaper article's interpretations. However, information of this sort is, deliberately, minimal, just enough to make sense of the text of the articles. The issue that then arises is how to incorporate this sort of information, as input from the announcements, to the ML process – but without falling into the infinite regress, as per our earlier discussion of mapping, of writing down hundreds or thousands of stylized facts known by practitioners of foreign policy and the journalists who cover their actions.

We are currently considering incorporating two forms of information available from the text of the announcements, as well as a third type of information available from other texts. The first of the three types is semantic roles (e.g., who did what to whom, where and when); the second is named-entity recognition. Just how internally structured and how heterogeneous the former need to be is an open question; our hope is that the multiple entailment relations across sentences within texts, as well as across texts, will permit us to avoid specifying large numbers of roles (although cross-context applicability is a potential problem). To be specific, the parser will, focusing in particular on verbs, query propositional dictionaries for the possible semantic roles of each subject or object of the core verb in each sentence (for details, see Carreras and

Màrquez 2004), then tag the relevant words or phrases in a fashion akin to that of the syntactic tagging discussed above. That tagged information will then be combined with the syntactic tags as input to the ML trainer.[8]

As regards the second type of possible semantic information, the hope is that specifying a wide variety of entities – not just countries, agencies, international organizations, and firms, but also individuals and material objects (e.g., F-15 fighter jets; Iranian gas turbines; M-2; 90-day Treasuries) – will facilitate entailment relations. Thus, to return to the fighter-jet example above, associating sending fighter jets to Poland and the Baltics with reassurance against Russian aggression should be possible with purely textual information. It should be noted, however, that, in doing so, the parser will have to be inputted with syntactic information as well (so that the interpretation of sending of F-15s is distinguished from the interpretation of stating, as Obama often did, that the Ukraine situation could not be resolved by sending "lethal" military aid to Ukraine itself). For now, however, specification of entities is a task for the future.

As can be seen, the type of semantics given by semantic role labeling is broad but thin, specifying roles but without domain specificity. As such, it misses the broad range of entailments suggested to any experienced reader, by a word, from the many textual contexts (qua words in the same sentence) in which the reader has already encountered that word. To begin to capture this background semantics (a better term might be back door semantics), we therefore used an additional type of information from Google's Bidirectional Encoder Representations from Transformers (BERT; Devlin2019). This technology, which has only been available for the last five years or so, has already revolutionized NLP modeling (for example, Mu et al. 2021; Cheng et al. 2021), by dint of being able to assign a matching score (BERTscore, based on associated words across a massive textual database; Zhang et al. 2020) to pairs of sentences, for example, "The weather is cold today" as a reference and "It is freezing today" as a candidate. In fact, we use BERT at two stages of the ML process: an initial filtering of announcements to decide which sentences to pick (necessary given that some announcements are buried deep within lengthy reports, press conferences, or legislative hearings); and, more importantly, in the StSt generation stage, where it is used to encode the filtered announcement.[9]

In principle, the combination of syntactic and semantic tagging, on the one hand, and ML, on the other, should permit contextual differentiation. While there is no doubt that the ratio of likely topical contexts to the number of interpretations in each domain is extremely high by computational linguistic standards, our expectation is that the relatively limited vocabulary of journalistic articles on a particular beat (this is even more true of domain context), the equally limited range of attributed motives in that beat, and the possibility of trying large numbers of quite intricate combinations of tagged data will palliate this difficulty. Although we have not yet incorporated full verb-based semantic tagging, we have incorporated syntactic tags into the BERT-filtered and -assessed ML model, and can now say a bit about the results. For now,

[8] It should be noted that both the semantic and syntactic tags are built on the Penn Treebank, i.e., a set of human annotations of some million words from *Wall Street Journal* newspaper articles in 1989; the original set of annotations were syntactic, and, to construct the propositional dictionaries, semantic tags were then added via additional human annotations.

[9] This encoding, in effect, is a set of contextual word vectors for each individual word in the filtered announcement. We then operate on those vectors in a decoding stage, using an "attention" algorithm to determine exactly how much of each vector to look at in order to generate the next word.

standard overlap scores (e.g., ROUGE) are of little value, given their all-or-nothing quality as regards matching (for this reason, we are also now developing more nuanced, human-assisted metrics). But we can get a flavor of how the model does and does not succeed by displaying worse vs. better pairs of actual and generated StSts.

One such pair is this:

Reference:

[STD SENTENCE START] mr . [ATTRIBUTION START] martin [ATTRIBUTION END] , who is one of the most optimistic of fed governors about the economy s ability to grow without raising inflation , [REFERENCE START] said [REFERENCE END] that there could be opportunities for the [ACTOR START] federal reserve [ACTOR END] to [ACT START] move to a more accommodative posture (might loosen monetary policy , in future) [ACT END] [SCOPE START] from time to time [SCOPE END] and thus [MOTIVE START] help to avoid slipping into a growth recession [MOTIVE END] . [STD SENTENCE END]

Candidate:

[STD SENTENCE START] paul a . [ATTRIBUTION START] volcker [ATTRIBUTION END] , chairman of the [ACTOR START] federal reserve board (fed) [ACTOR END] , [REFERENCE START] said [REFERENCE END] that the central bank [ACT START] would not ease monetary policy (will not tighten monetary policy) [ACT END] [SCOPE START] soon [SCOPE END] , but he said [EVIDENCE START] the economy was [EVIDENCE END] . [STD SENTENCE END]

By contrast, a second pair is this:

Reference:

[STD SENTENCE START] in mid - afternoon yesterday , this optimistic view of the credit market outlook was rein forced by news that the [ACTOR START] federal reserve [ACTOR END] open market committee [ACT START] had voted [REFERENCE START] may 26 [REFERENCE END] to give priority to the objective of [MOTIVE START] moderating pressures on financial markets [MOTIVE END] even though that might temporarily entail higher growth rates in the monetary aggregates than were considered appropriate for the longer run (might loosen monetary policy) [ACT END] . [STD SENTENCE END]

Candidate:

[STD SENTENCE START] the [REFERENCE START] summary [REFERENCE END] of the [ACTOR START] committee s (fed) [ACTOR END] meeting showed that [ACT START] the growth of money supply was aggregates (loosened monetary policy slightly) [ACT END] . [STD SENTENCE END]

What can be seen in these snippets is that although the model correctly is able to generate both the terminology and the general sense of the acts indicated in the announcement (keep in mind that phrases in parentheses are the human annotators' specification of the general thrust of the relevant StSt components), it falls down when it comes to issues of modality and temporality: in the first example, something that might happen in the future is turned into a fairly definite claim about the future (in negative language); in the second example, another contingent possibility is turned into a definite claim about what was already done. Whether this issue is most easily handled by adding modality information to the synactic tags, or, instead, by assuming that the semantic tags will deal with it indirectly is an issue we are now exploring.

Of course, one of the key differences between monetary policy and foreign policy, as issue domains, is that the latter is far more open-ended than the former. Not only are there more topics dealt with, but journalistic articles tend to quote many more announcements and range much longer back in time than for monetary policy. As a result, we should expect that the ML model will have to be altered.

5 Conclusion

Newspaper interpretation of policy announcements can be understood as a matter of textual entailment, rather than a rule-based mapping of statements from one structured domain to another. We have laid out a hybrid human coder/ML training methodology, using both syntactic and semantic information, to explore that entailment, and have begun to assess it using two cases. Preliminary evidence to date suggests that the methodology is a feasible way of understanding the announcement-interpretation connection, and we will pursue this when comparing across issue domains.

References

Abelson, Robert P., "The Ideology Machine," 1971. Annual Meeting of the American Political Science Association.

_ , "The Structure of Belief Systems," in Roger Schank and Kenneth Mark Colby, eds., *Computer Models of Thought and Language*, New York: W.H. Freeman & Co., 1973.

_ , "Differences Between Belief and Knowledge Systems," *Cognitive Science*, 1979, *3* (4), 355–366.

_ **and J. Douglas Carroll**, "Computer Simulation of Individual Belief Systems," *American Behavioral Scientist*, 1965, *8* (9), 24–30.

Carbonell, Jaime G. Jr., "POLITICS: Automated Ideological Reasoning," *Cognitive Science*, 1978, *2* (1), 1–15.

Carreras, Xavier and Lluís Màrquez, "Introduction to the CoNLL-2004 Shared Task: Semantic Role Labeling," in "Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004) at HLT-NAACL 2004" Association for Computational Linguistics Boston, Massachusetts, USA May 6 - May 7 2004, pp. 89–97.

Carvalho, Anabela, "Ideological Cultures and Media Discourses on Scientific Knowledge: Rereading News on Climate Change," *Public Understanding of Science*, 2007, *16* (2), 223–243.

Cheng, Quanying, Yunqiang Zhu, Jia Song, Hongyun Zeng, Shu Wang, Kai Sun, and Jinqu Zhang, "Bert-Based Latent Semantic Analysis (Bert-LSA): A Case Study on Geospatial Data Technology and Application Trend Analysis," *Applied Sciences*, 2021, *11* (24).

Dagan, Ido, Oren Glickman, and Bernardo Magnini, "The PASCAL Recognising Textual Entailment Challenge," in Quionero-Candela J., Dagan I., Magnini B., and d'Alché-Buc F., eds., *Machine Learning Challenges: Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Textual Entailment*, 2006, pp. 177–190.

Davis, Ernest and Gary Marcus, "Commonsense Reasoning and Commonsense Knowledge in Artificial Intelligence," *Communications of the ACM*, August 2015, *58* (9), 92–103.

Fillmore, Charles J., Christopher R. Johnson, and Miriam R.L. Petruck, "Background to FrameNet," *International Journal of Lexicography*, 2003, *16* (3), 235–250.

Gamson, William A. and Andre Modigliani, "Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach," *American Journal of Sociology*, July 1989, *95* (1), 1–37.

Gans, Herbert J., *Deciding What's News: A Study of CBS Evening News, NBC Nightly News, Newsweek, and Time*, New York: Random House, 1979.

Gupta, Anand, Manpreet Kaur, Shachar Mirkin, Adarsh Singh, and Aseem Goyal, "Text Summarization through Entailment-based Minimum Vertex Cover," in "Proceedings of the Third Joint Conference on Lexical and Computational Semantics (*SEM 2014)" Association for Computational Linguistics and Dublin City University Dublin, Ireland August 2014, pp. 75–80.

Henderson, James, "Learning Word Embeddings for Hyponymy with Entailment-Based Distributional Semantics," 2017.

Herman, Edward S. and Noam Chomsky, *Manufacturing Consent: The Political Economy of the Mass Media*, New York: Pantheon, 1988.

Hwang, Jena D., Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi, "COMET-ATOMIC 2020: On Symbolic and Neural Commonsense Knowledge Graphs," *arXiv preprint arXiv:2010.05953*, 2020.

Jost, John T., Christopher M. Federico, and Jaime L. Napier, "Political Ideology: Its Structure, Functions, and Elective Affinities," *Annual Review of Psychology*, 2009, *60*, 307–337.

Lippmann, Walter, *Public Opinion*, New York: Harcourt, Brace and Company, 1922.

Matthes, Jörg, "What's in a Frame? A Content Analysis of Media Framing Studies in the World's Leading Communication Journals, 1990-2005," *Journalism & Mass Communication Quarterly*, 2009, *86* (2), 349–367.

Mohammadshahi, Alireza and James Henderson, "Recursive Non-Autoregressive Graph-to-Graph Transformer for Dependency Parsing with Iterative Refinement," *Transactions of the Association for Computational Linguistics*, 03 2021, *9*, 120–138.

Most, Benjamin A. and Harvey Starr, "International Relations Theory, Foreign Policy Substitutability, and Nice Laws," *World Politics*, April 1984, *36* (3), 383–406.

Mu, Youqing, Hamid Tizhoosh, Rohollah Tayebi, Catherine Ross, Monalisa Sur, Brian Leber, and Clinton Campbell, "A BERT model generates diagnostically relevant semantic embeddings from pathology synopses with active learning," *Communications Medicine*, 07 2021, *1*, 11.

Palmer, Glenn and Archana Bhandari, "The Investigation of Substitutability in Foreign Policy," *Journal of Conflict Resolution*, February 2000, *44* (1), 3–10.

Pasunuru, Ramakanth, Han Guo, and Mohit Bansal, "Towards Improving Abstractive Summarization via Entailment Generation," in "Proceedings of the Workshop on New Frontiers in Summarization" Association for Computational Linguistics Copenhagen, Denmark September 2017, pp. 27–32.

Pavlick, Ellie, Travis Wolfe, Pushpendre Rastogi, Chris Callison-Burch, Mark Dredze, and Benjamin Van Durme, "FrameNet+: Fast Paraphrastic Tripling of FrameNet," *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers)*, 2015, pp. 408–413. Beijing, China, July 26-31, 2015.

Semetko, Holli A. and Patti M. Valkenburg, "Framing European Politics: A Content Analysis of Press and Television News," *Journal of Communication*, June 2000, *50* (2), 93–109.

Taber, Charles S., "POLI: An Expert System Model of U.S. Foreign Policy Belief Systems," *American Political Science Review*, December 1992, *86* (4), 888–904.

Tatar, Doiina, Andreea Diana Mihis, and Dana Lupsa, "Text Entailment for Logical Segmentation and Summarization," in Epaminondas Kapetanios, Vijayan Sugumaran, and Myra Spiliopoulou, eds., *Natural Language and Information Systems. NLDB 2008. Lecture Notes in Computer Science, vol. 5039*, Springer, 2008.

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.

	Min	Max	Median	Q1	Q3	Median absolute deviation	Mean	Standard deviation	95% confidence interval	Skewness
FOMC-2015: $N = 49$										
Number of reference documents in the article (Count)	1	7	2	1	2	1.48	1.92	1.26	0.36	1.88
Distance between earliest reference document and article (Number of days)	0	365	2	1	15	1.48	17.78	53.62	15.40	5.79
US-RUSSIA (2015): $N = 126$										
Number of reference documents in the article (Count)	1	7	2	1	2.75	1.48	2.03	1.30	0.23	1.62
Distance between earliest reference document and article (Number of days)	0	2287	2	1	13	1.48	101.79	310.92	54.82	4.44