Examining Political Rhetoric with Epistemic Stance Detection

Ankita Gupta[◊] **Su Lin Blodgett**[♡] **Justin H Gross**[◊] **Brendan O'Connor**[◊]

[◊]University of Massachusetts Amherst, [♡]Microsoft Research
{ankitagupta,jhgross,brenocon}@umass.edu
sublodge@microsoft.com

Abstract

Participants in political discourse employ rhetorical strategies—such as hedging, attributions, or denials—to display varying degrees of belief commitments to claims proposed by themselves or others. Traditionally, political scientists have studied these epistemic phenomena through labor-intensive manual content analysis. We propose to help automate such work through epistemic stance prediction, drawn from research in computational semantics, to distinguish at the clausal level what is asserted, denied, or only ambivalently suggested by the author or other mentioned entities (*belief holders*). We first develop a simple RoBERTa-based model for multi-source stance predictions that outperforms more complex state-of-the-art modeling. Then we demonstrate its novel application to political science by conducting a large-scale analysis of the Mass Market Manifestos corpus of U.S. political opinion books, where we characterize trends in cited belief holders—respected allies and opposed bogeymen—across U.S. political ideologies.

1 Introduction

Political argumentation is rich with assertions, hypotheticals and disputes over opponent's claims. While making these arguments, political actors often employ several rhetorical strategies to display varying degrees of commitments to their claims. For instance, political scientists have studied the *footing-shift* strategy, where actors convey their own beliefs while claiming that they belong to someone else (Goffman, 1981; Clayman, 1992). Sometimes they may attribute their beliefs to a majority of the population via *argument from popular opinion* (Walton et al., 2008). Actors can also resort to *hedging*, stating their own beliefs, but qualified with a partial degree of certainty (Fraser, 2010; Lakoff, 1975; Hyland, 1996) or express simple *political disagreements*, contradicting claims made by their opponents (Jang, 2009; Klofstad et al., 2013; Frances, 2014; Christensen, 2009).

Traditionally, political scientists and other scholars have manually analyzed the impact of such strategies and argumentation on audience perception (Clayman, 1992; Fraser, 2010). Recent advances in natural language processing (NLP) and digital repositories of political texts have enabled researchers to conduct large-scale analyses of political arguments using methods such as subjectivity analysis (Liu, 2012; Pang and Lee, 2008), argument mining (Trautmann et al., 2020; Toulmin, 1958; Walton, 1996), and opinion mining (Wiebe et al., 2005; Bethard et al., 2004; Kim and Hovy, 2004; Choi et al., 2005). While these approaches primarily concern argument structure and normative attitudes, we propose a complementary approach to analyze sources' *epistemic* attitudes towards assertions (Langacker, 2009; Anderson, 1986; Arrese, 2009)—what they believe to be true and the extent to which they commit to these beliefs.

Consider an example shown in Figure 1, where the author of the text (s1) quotes a speculation from the Congressional Quarterly (s2) about what Mitch McConnell (s3) said concerning Obama (s4). In this example, while the author of the text believes that the Congressional Quarterly hinted something about McConnell (thus, exhibiting a *certainly positive* (*CT+*) stance towards the event (e1), she remains *uncommitted* (*Uu*) about the quoted event (e3) that McConnell describes (edge omitted for visual clarity). Of course, this event is asserted as *certainly negative* (*CT-*) by McConnell, the speaker of the quote. The Congressional Quarterly suggests that Mitch McConnell made a statement (a *probably positive* (*PR+*) stance towards e2) while remaining *uncommitted* towards what he said. Finally, *Obama*'s own beliefs about whether he paid attention to Republican ideas are not expressed in this sentence; thus, s4 (Obama) has a *non-epistemic* label toward the listening event (e3).

To address this challenging problem of epistemological analysis, researchers within the NLP community have created several datasets and models

Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS), pages 89 - 104
November 7, 2022 ©2022 Association for Computational Linguistics

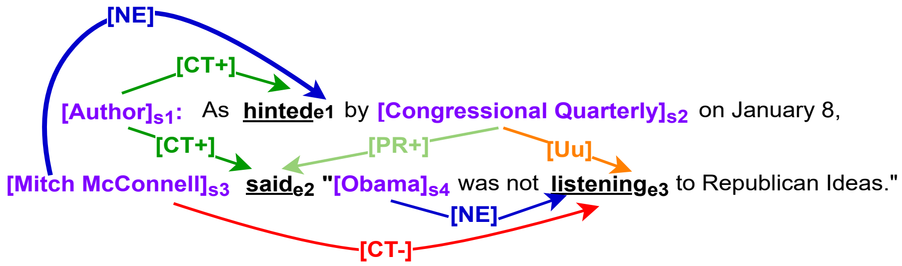

Figure 1: Illustrative example, simplified and adapted from a sentence in the Mass Market Manifestos corpus. There are four sources (s1–s4) and three events (e1–e3) with $4 \times 3 = 12$ labels between them; all epistemic stances are shown, but most non-epistemic (NE) labels are hidden for clarity. §1 and §3 describe the labels.

in various domains (Minard et al., 2016; Rambow et al., 2016; Rudinger et al., 2018b; Lee et al., 2015; Stanovsky et al., 2017; White et al., 2016; de Marneffe et al., 2012), often motivated directly by the interesting challenges of these linguistic semantic phenomena. However, there is a great potential to use an epistemic stance framework to analyze social relations (Soni et al., 2014; Prabhakaran et al., 2015; Swamy et al., 2017), motivating us to further advance this framework to support analysis of common rhetorical strategies and argumentation styles used in political discourse.

In this paper, we seek to further how *epistemic stance* analysis can help computationally investigate the use of *rhetorical strategies* employed in political discourse. In particular, we use the theory, structure and annotations of FactBank (Saurí and Pustejovsky, 2009), an expert-annotated corpus drawn from English news articles, which distinguishes different types of epistemic stances expressed in text. FactBank features annotations not just for the author, but also other sources (entities) mentioned in the text. Such multi-source annotations allow us to disambiguate the author's own beliefs from the beliefs they attribute to others.

Our main contributions in this work are:

- We conduct a literature review connecting ideas related to epistemic stance as studied across several disconnected scholarly areas of linguistics, NLP, and political science (§2).
- We develop a fine-tuned RoBERTa model (Liu et al., 2019) for multi-source epistemic stance prediction (§4), whose simplicity makes it accessible to social scientist users,[1] while performing at par with a more complex state-of-the-art model (Qian et al., 2018).
- We use our model to identify the most frequent *belief holders* which are epistemic sources whose views or statements are expressed by the author. Identifying belief holders is an essential first step in analyzing rhetorical strategies and arguments. We conduct this study on the Mass-Market Manifestos (MMM) Corpus, a collection of 370 contemporary English-language political books authored by an ideologically diverse group of U.S. political opinion leaders. We compare results to traditional named entity recognition. Finally, we analyze differences in what belief holders tend to be cited by left-wing versus right-wing authors, revealing interesting avenues for future work in the study of U.S. political opinion (§5).
- In the appendix, we additionally validate our model by replicating an existing manual case study comparing the commitment levels of different political leaders (§D, Jalilifar and Alavi, 2011), and give further analysis of the model's behavior with negative polarity items and different types of belief holders (§E).

2 Epistemic Stance from Different Perspectives

The notion of epistemic stances has been studied under several scholarly areas, including linguistics, political science and NLP. In this section, we discuss various notions of epistemic stances and how they have been utilized in these different areas.

2.1 Epistemic Stance in Linguistics

A speaker's *epistemic stance* is their positioning about their knowledge of, or veracity of, communicated events and assertions (Biber and Finegan, 1989; Palmer, 2001; Langacker, 2009). Epistemic stance relates to the concept of *modality*, which deals with the degree of certainty of situations in

[1] All resources accompanying this project are added to our project page: `https://github.com/slanglab/ExPRES`

the world, and has been extensively studied under linguistics (Kiefer, 1987; Palmer, 2001; Lyons, 1977; Chafe, 1986) and logic (Horn, 1972; Hintikka, 1962; Hoek, 1990; Holliday, 2018). From a cognitivist perspective, epistemic stance concerns the pragmatic relation between speakers and their knowledge regarding assertions (Biber and Finegan, 1989; Mushin, 2001; Martin and White, 2005).

2.2 Epistemic Stance in Political Science

The use of epistemic stances is widespread in political communication and persuasive language, to describe assertions when attempting to influence the reader's view (Chilton, 2004; Arrese, 2009). For instance, Chilton (2004) studies use of epistemic stances by speakers/writers for legitimisation and coercion; Arrese (2009) examines epistemic stances taken by speakers to reveal their ideologies. In these studies, a speaker's communicated stance may follow what they believe due to their experiences, inferences, and mental state (Anderson, 1986). From a psychological perspective, Shaffer (1981) employs balance theory (Heider, 1946)—the cognitive effect of knowing an entity's stance towards an issue—in explaining public perceptions of presidential candidates' issue positions.

2.3 Epistemic Stance in NLP

In the NLP literature, epistemic stances—typically of authors, and sometimes of mentioned textual entities—have been studied under the related concepts of *factuality* (Saurí and Pustejovsky, 2012; Rudinger et al., 2018a; Lee et al., 2015; Stanovsky et al., 2017; Minard et al., 2016; Soni et al., 2014) and *belief commitments* (Prabhakaran et al., 2015; Diab et al., 2009). de Marneffe et al. (2012) prefers the term *veridicality* to study the reader's, not author's, perspective.

We use the term *epistemic stance* to avoid confusion with at least two more recent subliteratures that use *factuality* differently from the above. In misinformation detection, factuality refers to a proposition's objective truth (Rashkin et al., 2017; Mihaylova et al., 2018; Thorne et al., 2018; Vlachos and Riedel, 2014). By contrast, we follow the epistemic stance approach in not assuming any objective reality—we simply model whatever subjective reality that agents assert. Furthermore, text generation work has studied whether text summaries conform to a source text's asserted propositions—termed the factuality or "factual correctness" of a summary (Maynez et al., 2020; Wiseman et al., 2017; Kryscinski et al., 2019; Dhingra et al., 2019).

Type	Dataset	Perspective	Genre	Label
Factuality	FactBank (Saurí and Pustejovsky, 2012)	Multi	News	Disc (8)
	Stanovsky et al., 2017	Author	News	Cont [-3, 3]
	MEANTIME (Minard et al., 2016)	Multi	News (Italian)	Disc (3)
	Lee et al., 2015	Author	News	Cont [-3, 3]
	UDS-IH2 (Rudinger et al., 2018b)	Author	Open	Disc (2) & Conf [0,4]
	Yao et al., 2021	Multi	News	Disc (6)
	Vigus et al., 2019	Multi	Open	Disc (6)
Indirect Reporting	Soni et al., 2014	Reader	Twitter	Likert (5)
Pragmatic Veridicality	PragBank (de Marneffe et al., 2012)	Reader	News	Disc (7)
Beliefs	Diab et al., 2009	Author	Open	Disc (3)
	Prabhakaran et al., 2015	Author	Forums	Disc (4)

Table 1: Summary of epistemic stance annotated datasets. *Perspective*: which sources are considered for annotation? Stance *Label* may be discrete with the given number of categories (where many or all are ordered), or continuous with a bounded range.[2] All datasets except MEANTIME consist of English text.

Several researchers in NLP have explored interesting social science applications in multiple settings such as organizational interactions (Prabhakaran et al., 2010), Supreme Court hearings (Danescu-Niculescu-Mizil et al., 2012), discussion (Bracewell et al., 2012; Swayamdipta and Rambow, 2012) and online forums (Biran et al., 2012; Rosenthal, 2014). In particular, Prabhakaran et al. (2010) use epistemic stances to analyse power relations in organizational interactions. These studies demonstrate the potential of using epistemic stance analysis for social science applications. Motivated by these advances, we use epistemic stance framework to analyze political rhetoric, a genre that has not been explored earlier.

Existing Datasets Several existing datasets (Rudinger et al., 2018b; Lee et al., 2015; Prabhakaran et al., 2015; Diab et al., 2009; Stanovsky et al., 2017) have successfully driven the progress of epistemic stance analysis in NLP, but have largely focused on author-only analysis. Soni et al. (2014) and de Marneffe et al. (2012) examine epistemic stances from the reader's (not author's) perspective. Table 1 summarizes these datasets.[2]

Political discourse is a particularly interesting because the multiple sources discussed can have diverse stances towards the same event. Among all existing datasets, FactBank (Saurí and Pustejovsky, 2012) and MEANTIME (Minard et al., 2016) explore multi-source analysis in the news domain.

[2]UDS-IH2 collects a binary category and a confidence score. Yao et al. (2021) and Vigus et al. (2019) extend multi-source annotations as dependency graphs with additional edge types.

Algorithm	Features/Model	Perspective	Systems
Rule-Based	Predicate Lexicons	Author	Nairn et al., 2006 Lotan et al., 2013 (TruthTeller)
		Multiple	Saurí and Pustejovsky, 2012 (DeFacto)
Feature-Based Supervised Machine Learning	Lexico-Syntactic	Author	Diab et al., 2009, Lee et al., 2015 Prabhakaran et al., 2015
		Reader	de Marneffe et al., 2012 Soni et al., 2014
		Multiple	Qian et al., 2015
	Output of Rule System	Author Multiple	Stanovsky et al., 2017 Saurí and Pustejovsky, 2012
Neural Networks (NN)	LSTM GAN Graph NN	Author Multiple Author	Rudinger et al., 2018b Qian et al., 2018 Pouran Ben Veyseh et al., 2019
Neural Pretrained	BERT	Author Multiple	Jiang and de Marneffe, 2021 This work

Table 2: Epistemic stance prediction models.

While MEANTIME has helped advance epistemic stance analysis in Italian, FactBank—built on English news text—is closest to our goal.

Existing Models Several computational models have been developed for epistemic stance prediction as explicated in Table 2. Early models proposed deterministic algorithms based on hand-engineered implicative signatures for predicate lexicons (Lotan et al., 2013; Nairn et al., 2006; Saurí and Pustejovsky, 2012). A number of systems used lexico-syntactic features with supervised machine learning models, such as SVMs or CRFs (Diab et al., 2009; Prabhakaran et al., 2010; Lee et al., 2015; Saurí and Pustejovsky, 2012; Stanovsky et al., 2017). Lately, there has been a growing interest in using neural models for epistemic stance prediction (Rudinger et al., 2018b; Pouran Ben Veyseh et al., 2019), though sometimes with complex, task-specific network architectures (e.g. GANs; Qian et al. (2018)), which raise questions about generalization and replicability for practical use by experts from other fields. Recently, Jiang and de Marneffe (2021) explore fine-tuning pre-trained language models (LM), such as BERT, for author-only epistemic stance prediction by adding a simple task-specific layer. We take this more robust approach, extending it to multiple sources.

General Stance Detection in NLP Recently, there has been a growing interest in analyzing stance, including a broad spectrum of stance-takers (speaker/writer), the objects of stances, and their relationship. While our work also examines the stance relationship between a source (stance-taker) and an event (object), we differ from the existing literature in several ways. For instance, unlike our work where a stance-taker is the author or a mentioned source in the text, Mohtarami et al. (2018), Pomerleau and Rao (2017) and Zubiaga et al. (2016) consider the entire document/message to be a stance-taker. Similarly, the object of the stance could be a target entity (such as a person, organization, movement, controversial topic, etc.) that may or may not be explicitly mentioned in the input document (Mohammad et al., 2016). On the contrary, in this work, event propositions (object) are always embedded within the text.

Finally, we can also analyze the kind of stance relationship exhibited by the stance-taker towards an object from two linguistic perspectives: affect and epistemic. Affect involves the expression of a broad range of personal attitudes, including emotions, feelings, moods, and general dispositions (Ochs and Schieffelin, 1989), and has been explored in Mohammad et al. (2016). On the other hand, epistemic—this work's focus—refers to the speaker's expressed attitudes towards knowledge of events and her degree of commitment to the validity of the communicated information (Chafe, 1986; Biber and Finegan, 1989; Palmer, 2001). The analysis explored in Mohtarami et al. (2018), Pomerleau and Rao (2017) and Zubiaga et al. (2016) seems to be epistemic as they implicitly incorporate the knowledge or claims expressed in the evidence document and hence their stances towards them, although such distinctions are not made explicitly in their work. While the stance literature discussed in this section has not been connected to epistemic stance literature in NLP, we think interesting future work can be done to establish this relationship.

3 An Epistemic Stance Framework for Analyzing Political Rhetoric

This section formally introduces the task of epistemic stance detection and describes the details of the FactBank dataset. We then explain how the epistemic stance framework relates to several rhetorical strategies often used in political discourse.

3.1 Epistemic Stances

We define an epistemic stance tuple as a triple of *(source, event, label)* within a sentence, where the label is the value of the source's epistemic stance (or a non-epistemic relation) toward the event. The triples can be viewed as a fully connected graph among all sources and events in the sentence (Figure 1). We use the structure and theory of FactBank (Saurí and Pustejovsky, 2012) to identify sources, events and the stance labels.

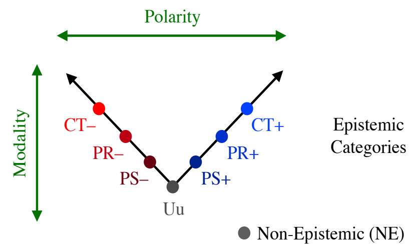

Figure 2: Stance labels used in this work, ordered along two linguistic dimensions, as well as a separate non-epistemic category.

Sources and Events A *source* is an entity—either the text's author, or an entity mentioned in the sentence—which can hold beliefs. FactBank contains annotations for sources that are subjects of source-introducing predicates (SIPs), a manually curated lexicon of verbs about report and belief such as *claim, doubt, feel, know, say, think*. Annotations of these embedded sources allow us to analyze the author's depiction of the embedded source's beliefs towards an event. The special *Author* source is additionally included to analyze the author's own beliefs. FactBank's definition of *events* includes a broad array of textually described eventualities, processes, states, situations, propositions, facts, and possibilities. FactBank identifies its event tokens as those marked in the highly precise, manually annotated TimeBank and AQUAINT TimeML[3] corpora.

Epistemic Stance Label FactBank characterizes epistemic stances along two axes, polarity and modality. The polarity is binary, with the values positive and negative—the event did (not) happen or the proposition is (not) true. The modality constitutes a continuum ranging from uncertain to absolutely certain, discretely categorized as *possible (PS)*, *probable (PR)* and *certain (CT)*. An additional *underspecified or uncommitted stance (Uu)* is added along both axes to account for cases such as attribution to another source (non-commitment of the source) or when the stance of the source is unknown. The epistemic stance is then characterized as a pair *(modality, polarity)* containing a modality and a polarity value (e.g., *CT+*) (Figure 2).

FactBank gives epistemic stance labels between certain pairs of sources and events only, based on structural syntactic relationships. However, for raw text we may not have reliable access to syntactic structures, and sources and events must be automatically identified, which may not be completely accurate. We use a simple solution by always assuming edges among the cross-product of all sources and events within a sentence, and to predict a separate *Non-Epistemic (NE)* category for the large majority of pairs. This accounts for any spurious event-source pairs, structurally invalid configurations such as an embedded source's stance towards an event outside their factive context (Figure 1: (s4,e2)), or a source that cannot be described as a belief holder (and thus, all its stances are *NE*).

Given that a variety of datasets have been collected for tasks related to epistemic stance (§2), Stanovsky et al. (2017) argues to combine them for modeling. However, some datasets address different epistemic questions (e.g., the reader's perspective), and they follow very different annotation guidelines and annotation strategies, risking ambiguity in labels' meaning. In preliminary work we attempted to crowdsource new annotations but found the resulting labels to be very different than FactBank, which was created by a small group of expert, highly trained annotators. Thus we decided to exclusively use FactBank for modeling.

3.2 Connections between Epistemic Stances and Rhetorical Strategies

Some epistemic stances in FactBank's framework can be mapped to a common political rhetorical strategy. For instance, a source utilizing *certainly positive/negative* (*CT+/CT-*) stances more frequently can be associated with displaying higher commitment levels. The *CT+/CT-* stances can also help analyze *political disagreements* by identifying two sources with opposite stances towards an event, i.e., a source asserting an event (*CT+*) and a source refuting the same event (*CT-*). A source may exhibit a *probable/possible* (*PR/PS*) stance to indicate that the event could have happened, abstaining from expressing strong commitments towards this event, which can be useful to analyze *hedging*. Finally, *underspecified/uncommitted* (*Uu*) stances can help identify the embedded sources whose beliefs are mentioned by the author while remaining uncommitted, a strategy related to *footing-shift* in political discourse. Use of *Uu* stances is also helpful to identify *belief holders*—entities described as having epistemic stances (§5)—since sometimes the author remains uncommitted while reporting the embedded source's stance.

4 Model

We present a simple and reproducible RoBERTa-based neural model for epistemic stance classifica-

[3] https://web.archive.org/web/20070721130754/http://www.timeml.org/site/publications/specs.html

Model	CT+	CT-	PR+	PS+	Uu	NE	Macro Avg (Non-NE)	Macro Avg (All)
DeFacto (Saurí and Pustejovsky, 2012)	85.0	75.0	46.0	59.0	75.0	-	70.0	-
SVM (Saurí and Pustejovsky, 2012; Prabhakaran et al., 2010)	90.0	61.0	29.0	39.0	66.0	-	59.0	-
BiLSTM (Qian et al., 2018)	85.2	74.0	58.2	61.3	73.3	-	70.4	-
AC-GAN (Qian et al., 2018)	85.5	74.1	63.1	65.4	75.1	-	72.6	-
BERT (Jiang and de Marneffe, 2021)	89.7	69.8	45.0	46.7	82.8	97.9	66.8	72.0
RoBERTa (this work)	90.7	78.4	51.4	62.7	84.8	97.8	73.6	77.6

Table 3: F1 scores for our RoBERTa based epistemic stance classifier and all baseline models.

tion using a standard fine-tuning approach.[4] BERT fine-tuning is effective for many NLP tasks (Devlin et al., 2019), and recent work on pre-trained language models such as BERT (Shi et al., 2016; Belinkov, 2018; Tenney et al., 2019a,b; Rogers et al., 2020) shows such models encode syntactic and semantic dependencies within a sentence, which is highly related to the epistemic stance task.

Recently, Jiang and de Marneffe (2021) use a fine-tuned BERT model for author-only epistemic stance prediction, obtaining strong performance on several datasets. We extend their approach, developing a BERT model (using the RoBERTa (Liu et al., 2019) pre-training variant) for the structurally more complex multi-source task, and give the first full comparison to the foundational multi-source system, DeFacto (Saurí and Pustejovsky, 2012). We leave the exploration of other advanced transformer-based models (Brown et al., 2020; Raffel et al., 2020) for further performance gains as future work.

To develop a model suitable for multi-source predictions, we follow Tenney et al. (2019b) and Rudinger et al. (2018a)'s architecture for semantic (proto-role) labeling, which they formulate as predicting labels for pairs of input embeddings. To predict the epistemic stance for an event-source pair (e, s) in a sentence, we first compute contextual embeddings for the sentence's tokens, $[h_1^L, h_1^L, ..., h_n^L]$, from a BERT encoder's last (L^{th}) layer. We concatenate the source (h_s^L) and event (h_e^L) token embeddings (each averaged over BERT's sub-token embeddings), and use a single linear layer to parameterize a final softmax prediction $\hat{f} \in [0, 1]^C$ over the $C = 6$ epistemic stance classes,[5] which is trained with cross entropy loss over all tuples in the training set. We apply inverse frequency class weighting to encourage accurate modeling for comparatively rare classes like the *CT-, PR+ and PS+* class. Finally, to cleanly analyze the author source in the same manner as other mentioned sources, we augment the sentence with the prefix "Author: " (following a dialogue-like formatting),[6] and use its index and embedding for inferences about the author source.

Table 3 shows the performance of our RoBERTa based epistemic stance classifier. We compare our model against several baselines, including rule-based methods (DeFacto; Saurí and Pustejovsky (2012)), machine learning classifiers (SVM Saurí and Pustejovsky (2012); Prabhakaran et al. (2010)), and neural network based methods (BiLSTM and AC-GAN by Qian et al. (2018)) as described in §2.3.[7] We also extend the author-only BERT model by Jiang and de Marneffe (2021) to support multi-source predictions in line with our modeling approach. The RoBERTa model performs the best obtaining a macro-averaged F1 score of 77.6±0.011 on all six epistemic labels and an F1 score of 73.6±0.031 on the original five epistemic labels (excluding the *Non-Epistemic* label). Although the RoBERTa model has a much simpler architecture, it performs the same or better than AC-GAN. All pairwise significance tests resulted in p-values $<$ 0.01. Details of implementations and statistical testing is provided in Appendix §A.1 and §A.2.

The above epistemic stance classifier, like most previous modeling approaches (Qian et al., 2015; Saurí and Pustejovsky, 2012), requires pre-identified sources and events, which do not exist in real-world text. We use Qian et al. (2018)'s two-step approach to first identify sources and events in the input text and then determine stances for every recognized (source, event) pair. Source and event identification is through two RoBERTa-based token classifiers, using a linear logistic layer for binary classification of whether a token is a source (or event), fine-tuned on the same training corpus.

Our source and event identification models

[4]We intentionally keep the modeling simple to make it more accessible to political scientists and users with less computational experience. We further simplify by augmenting BERT with a single task-specific layer, as opposed to a new task-specific model architecture proposed in Pouran Ben Veyseh et al. (2019); Qian et al. (2018); Rudinger et al. (2018b).

[5]CT+, CT-, PR+, PS+, Uu, NE; Saurí and Pustejovsky (2012) additionally define probably/possibly negative (PR-/PS-) stances. However, these stances are rare in the corpus, making modeling and evaluation problematic. Following Qian et al. (2015, 2018), we omit them in this study.

[6]With and without the trailing colon gave same results.

[7]Since the DeFacto implementation is not available, we compare our model's predictions on the FactBank test set against evaluation statistics derived from the test set confusion matrix reported by Saurí and Pustejovsky. We use implementation provided at `https://github.com/qz011/ef_ac_gan` for SVM, BiLSTM and AC-GAN baselines.

achieve a macro-averaged F1 score of 81.8 ± 0.019 and 85.78 ± 0.007, respectively, slightly improving upon the only existing prior work of Qian et al. (2018) by 1.85% and 1.29% respectively, with pairwise significance tests resulting in p-values < 0.01. We also experimented with a joint model to identify sources and events; however, individual classifiers gave us better performance (Appendix §B.1).

5 Case Study: Belief Holder Identification

Political discourse involves agreement and contention between the author and other belief-holding sources they cite. As a first step, we extract major belief holders mentioned in a text to allow analysis of ideological trends in U.S. political discourse.

5.1 Corpus Description

We conduct our case study on the new Mass-Market Manifestos (MMM) corpus, a curated collection of political nonfiction authored by U.S. politicians, media activists, and opinion elites in English, published from 1993-2020. It subsumes and more than triples the size of Contemporary American Ideological Books (Sim et al., 2013). The corpus contains 370 books (31.9 million tokens) spanning various U.S. political ideologies. Human coders identified 133 books as liberal or left-wing, 226 as conservative or right-wing, and 11 as explicitly centrist or independent. Since ideological opponents often draw from a shared set of concepts—sometimes stating perceived facts and sometimes dismissing others' claims—this presents us with a perfect challenge for detection of epistemic stance.

5.2 Belief Holder Identification

A *belief holder* is defined as a non-author source that holds at least one epistemic stance toward some event. We identify belief holders by using our best-performing model (fine-tuned RoBERTa, predictions averaged over 5 random restarts) to infer epistemic stances for all source-event pairs identified in the 370 books in the MMM corpus. For the problem of identifying sources that are belief holders as per this definition, we obtain 77.3 precision and 79.4 recall on FactBank's evaluation corpus.

For aggregate analysis (§5.4), especially for named entity sources, a longer span is more interpretable and less ambiguous. Thus, when a source token is recognized as part of a traditional named entity (via spaCy v3.0.6; Honnibal and Johnson (2015)), the belief holder is defined as the full NER span; otherwise, simply the source token is used.

5.3 Comparison to Named Entity Recognition

Instead of using epistemic stance-based belief holder identification, an alternative approach is to exclusively rely on named entity recognition (NER) from a set of predefined types. NER has been used in opinion holder identification (Kim and Hovy, 2004) and within belief evaluation in the TAC KBP Belief/Sentiment track (TAC-KBP, 2016). By contrast, our model can instead find *any* entity as a belief holder, as long as it holds epistemic stances, without a type restriction. To illustrate this, we compare our belief holder identifier to a standard NER implementation from spaCy v3.0.6 (Honnibal and Johnson, 2015),[8] trained on English web corpus of OntoNotes 5.0 (Hovy et al., 2006). We use entities identified as one of OntoNotes' 11 non-numeric named entity types.[9] Aggregating among all books in the corpus, the set of belief holders identified by our model has only a 0.198 Jaccard similarity with the set of NER-detected entities (Appendix §E.2 Table 9 provides qualitative examples from one conservative book).[10]

Is it reasonable to define a set of named entity types to identify belief holders? We calculate each named entity type's *belief score*, which is the average proportion of named entities of that type that are described as holding an epistemic stance.[11] As shown in Figure 3, while the Organization, NORP, Person and GPE types have significantly higher belief score than others, there is a wide range of variation, including non-obvious types such as Work of Art (e.g., The Bible), suggesting that a NER type whitelists undercover or overcover possible belief holders. We provide a further linguistic breakdown of identified belief holders in Appendix §E.3.

5.4 Political Analysis of Belief Holders

The MMM corpus, including both left and right-wing authors, gives an opportunity to study the belief holder citation practices for each U.S. political ideology. Using our epistemic stance and entity aggregation postprocessing (§5.2), we count the number of books each belief holder is mentioned in. There are 1269 sources mentioned as a belief

[8]CPU optimized version of `en_core_web_lg`. Stanza's (Qi et al., 2020) performance-optimized NER system gave broadly similar results.

[9]*Event, Facility, GPE, Language, Law, Location, NORP, Organization, Person, Product, Work_of_Art*

[10]An entity is defined as a belief holder if it is the source for at least one epistemic tuple; similarly, it is a named entity if at least one occurrence is identified as part of an NER span.

[11]For each source instance with same NER type, we find the proportion of epistemic (non-NE) stances among events in its sentence, then average these values across the corpus.

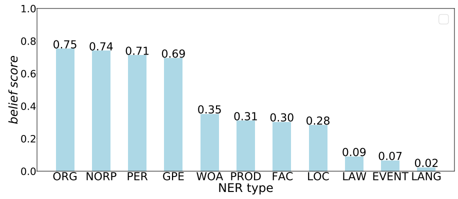

Figure 3: Imperfect correlation between belief scores and OntoNotes NER types. (WOA: Work of Art, PROD: Product, PER: Person, ORG: Organization, LOC: Location, NORP: Nationalities or Religious or Political Groups, FAC: Facility, LANG: Language, GPE: Geo-Political Entity)

Highly Cited by Left-wing Authors		Highly Cited by Right-wing Authors	
Belief Holder	View	Belief Holder	View
Tom Delay	Opposed	Paul Johnson	Respected
Martin Gilens	Respected	Marvin Olasky	Respected
Michelle Alexander	Respected	Saul Alinsky	Opposed
Grover Norquist	Opposed	Robert Rector	Respected
Jane Mayer	Respected	Thomas Sowell	Respected
Albert Camus	Respected	The Tax Foundation	Respected
Consumers	Respected	Soviets	Opposed
Thomas Edsall	Respected	George Soros	Opposed
Jacob Hacker	Respected	Pew Research	Respected
James Baldwin	Respected	John Edwards	Opposed
Jeffrey Sachs	Respected	George Stephanopoulos	Opposed
Michele Bachmann	Opposed	John Stossel	Respected
Ben Bernanke	Unclear	Thomas Sowell	Respected
Chris Hedges	Respected	Nicholas Eberstadt	Respected
Lobbyists	Opposed	James Wilson	Respected
Bill Moyers	Respected	Iran	Opposed
Daniel Bell	Respected	Hollywood	Opposed
David Cay Johnston	Respected	George Gilder	Respected
Instructor	Generic	Dennis Prager	Respected
Moderator	Generic	Arthur Brooks	Respected

Table 4: Top 20 most frequently mentioned belief holders per author ideology (left vs. right), among belief holders mentioned in ≥ 8 books in the MMM corpus.

holder in ≥ 8 books. For each belief holder, we calculate its left-right citation ratio: the proportion of left-wing books it is mentioned in, versus the proportion of right-wing books (proportions are calculated using a book pseudocount of 1 to avoid dividing by zero). Belief holders with a ratio ~ 1.0 include some generic (*team, organization, official*) and anaphoric (*anyone, many*) examples.

Table 4 shows the top 20 belief holders for both left and right, as ranked by this ratio, yielding a rich set of politicians (Delay, Edwards), journalists (Mayer, Stephanopoulos), activists (Norquist, Alinsky), and many social scientists and scholars (Gilens, Johnson). Most of these belief holders were recognized by an expert (political scientist coauthor) as being respected or opposed from the citing ideological perspective. Based on prior knowledge of U.S. politics it was straightforward to immediately give such judgments for most entries; for a few unclear ones, we checked individual sentences mentioning the belief holder. A common strategy is to describe an opponent's views or statements—the use of a rhetorical bogeyman.

Repeating the analysis for widely cited belief holders appearing in ≥ 100 books, yields more general, and again politically meaningful, entities (Table 5). Some well-known patterns are clearly visible, such as liberals' respect for technocratic authority (*economists, scientists, research*), and conservative respect for the semi-mythical *founders* alongside derision for the *media*. Both sides frequently cite the opposition (L: *Republicans*, R: *Democrats*), though interestingly the right cites both conservatives and liberals (relatively more frequently than the left). Figure 4 shows examples of *founders*, with the most skewed ratio ($0.308 \approx 3.2^{-1}$) among this set of entities. Overall, our automated belief holder identification yields a politically significant entity list, laying the groundwork for more systematic manual and computational analysis (e.g., network or targeted sentiment analysis).

Left-cited		Right-cited	
Economists	Studies	Founders	Democrats
Woman	Research	Media	Officials
Polls	Republicans	Poll	President
Scientists	Group	Obama	Conservatives
Groups	Friend	Government	Liberals

Table 5: Top 10 most frequently mentioned belief holders per author ideology, among belief holders mentioned in at least 100 books.

- We know that most of the **[Founders]**$_S$ regarded slavery as a wrong that would have to be addressed. *Chuck Norris, Black Belt Patriotism (R)*
- Sometimes, whether against gator or human predator, you're on your own, as the frontier-expanding **[Founders]**$_S$ well knew. *Charlie Kirk, The MAGA Doctrine (R)*
- This is not to say the **[founders]**$_S$ believed that only religious individuals could possess good character. *William Bennett, America the Strong (R)*
- The **[founders]**$_S$, however, had quite another idea, based on their experience in the colonies over the decades before, where actual religious freedom had existed. *Eric Metaxas, If You Can Keep It (R)*
- The **[Founders]**$_S$ recognized that there were seeds of anarchy in the idea of individual freedom [..], for if everybody is truly free, without the constraints of birth or rank or an inherited social order [..] then how can we ever hope to form a society that coheres? *Barack Obama, The Audacity of Hope (L)*

Figure 4: Examples of *founders* as a belief holder.

6 Conclusion

Semantic modeling has exciting potential to deepen the NLP analysis of political discourse. In this work, we analyze the epistemic stance of various sources toward events, by developing a RoBERTa-based model, and using it for identifying major belief holders mentioned by political authors. We conduct a large-scale analysis of the Mass Market Manifestos corpus of U.S. political opinion books, where we characterize trends in cited belief holders across U.S. political ideologies. In future, we hope to use this framework to help construct a database of beliefs, belief holders, and their patterns of agreement and disagreement in contentious domains.

Acknowledgements

We are thankful for the feedback and comments from the reviewers. We are grateful to Philip Resnik, Michael Colaresi, Arthur Spirling, Katherine Keith, Geraud Nangue Tasse, Kalpesh Krishna, Marzena Karpinska, and the UMass NLP group for valuable discussions during the course of the project. This work was supported by National Science Foundation grants 1814955 and 1845576. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Kareem Abdul-Jabbar and Raymond Obstfeld. 2016. *Writings on the wall: Searching for a new equality beyond Black and White*. Time Inc. Books.

Lloyd B. Anderson. 1986. Evidentials, paths of change, and mental maps: Typologically regular asymmetries. In Wallace L. Chafe and Johanna Nichols, editors, *Evidentiality: The Linguistic Coding of Epistemology*, pages 273–312. Ablex.

Juana I Marín Arrese. 2009. Effective vs. epistemic stance, and subjectivity/intersubjectivity in political discourse. a case study. *Studies on English modality in honour of Frank Palmer. Linguistic Insights*, 111:23–131.

Glenn Beck. 2018. *Addicted to Outrage: How Thinking Like a Recovering Addict Can Heal the Country*. Threshold Editions.

Yonatan Belinkov. 2018. *On internal language representations in Deep Learning: An analysis of Machine Translation and Speech Recognition*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, USA.

Taylor Berg-Kirkpatrick, David Burkett, and Dan Klein. 2012. An empirical investigation of statistical significance in NLP. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 995–1005, Jeju Island, Korea. Association for Computational Linguistics.

Steven Bethard, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou, and Dan Jurafsky. 2004. Automatic extraction of opinion propositions and their holders. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, volume 2224.

Douglas Biber and Edward Finegan. 1989. Styles of stance in english: Lexical and grammatical marking of evidentiality and affect. *Text-Interdisciplinary Journal for the study of Discourse*, 9(1):93–124.

Or Biran, Sara Rosenthal, Jacob Andreas, Kathleen McKeown, and Owen Rambow. 2012. Detecting influencers in written online conversations.

David B Bracewell, Marc Tomlinson, and Hui Wang. 2012. A motif approach for identifying pursuits of power in social discourse. In *2012 IEEE Sixth International Conference on Semantic Computing*, pages 1–8. IEEE.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Pat Buchanan. 1999. A republic, not an empire: Reclaiming america's destiny.

Wallace Chafe. 1986. Evidentiality in english conversation and academic writing. *W. Chafe, & J. Nichols (Eds.), Evidentiality: The Linguistic Coding of Epistemology (pp. 261-272). Norwood, NJ: Ablex Publishing Corp.*

Paul Chilton. 2004. *Analysing Political Discourse: Theory and Practice*. London: Routledge.

Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. Identifying sources of opinions with Conditional Random Fields and extraction patterns. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, page 355–362, USA. Association for Computational Linguistics.

David Christensen. 2009. Disagreement as evidence: The epistemology of controversy. *Philosophy Compass*, 4(5):756–767.

Steven E Clayman. 1992. Footing in the achievement of neutrality: The case of news interview discourse. *Talk at work: Interaction in institutional settings*, 163:198.

Ann H Coulter. 2009. Guilty: Liberal" victims" and their assault on america. Crown Forum.

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the 21st international conference on World Wide Web*, pages 699–708.

Marie-Catherine de Marneffe, Christopher D. Manning, and Christopher Potts. 2012. Did it happen? the pragmatic complexity of veridicality assessment. *Computational Linguistics*, 38(2):301–333.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Bhuwan Dhingra, Manaal Faruqui, Ankur Parikh, Ming-Wei Chang, Dipanjan Das, and William Cohen. 2019. Handling divergent reference texts when evaluating table-to-text generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4884–4895, Florence, Italy. Association for Computational Linguistics.

Mona Diab, Lori Levin, Teruko Mitamura, Owen Rambow, Vinodkumar Prabhakaran, and Weiwei Guo. 2009. Committed belief annotation and tagging. In *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*, pages 68–73, Suntec, Singapore. Association for Computational Linguistics.

Jesse Dodge, Gabriel Ilharco, Roy Schwartz, Ali Farhadi, Hannaneh Hajishirzi, and Noah A. Smith. 2020. Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping. *Computing Research Repository*, arXiv:2002.06305.

Rod Dreher. 2018. *The Benedict option: A strategy for Christians in a post-Christian nation*. Penguin.

Steve Forbes and Elizabeth Ames. 2012. *Freedom Manifesto: Why Free Markets Are Moral and Big Government Isn't*. Currency.

Bryan Frances. 2014. *Disagreement*. John Wiley & Sons.

Bruce Fraser. 2010. Hedging in political discourse. *OKULSKA, U., CAP, P., Perspectives in Politics and Discourse, Capitolo*, 8.

Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson F. Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer. 2018. AllenNLP: A deep semantic natural language processing platform. In *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, pages 1–6, Melbourne, Australia. Association for Computational Linguistics.

Erving Goffman. 1981. *Forms of talk*. University of Pennsylvania Press.

Fritz Heider. 1946. Attitudes and cognitive organization. *The Journal of psychology*, 21(1):107–112.

Kaarlo Jaakko Juhani Hintikka. 1962. Knowledge and belief: An introduction to the logic of the two notions.

Wiebe van der Hoek. 1990. Systems for knowledge and beliefs. In *European Workshop on Logics in Artificial Intelligence*, pages 267–281. Springer.

Wesley H Holliday. 2018. Epistemic logic and epistemology. In *Introduction to formal philosophy*, pages 351–369. Springer.

Matthew Honnibal and Mark Johnson. 2015. An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1373–1378, Lisbon, Portugal. Association for Computational Linguistics.

Laurence Robert Horn. 1972. *On the semantic properties of logical operators in English*. University of California, Los Angeles.

Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes: The 90% solution. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pages 57–60, New York City, USA. Association for Computational Linguistics.

Ken Hyland. 1996. Writing without conviction? hedging in science research articles. *Applied linguistics*, 17(4):433–454.

Ali Reza Jalilifar and Maryam Alavi. 2011. Power and politics of language use: A survey of hedging devices in political interviews. *Journal of Teaching Language Skills*, 30(3):43–66.

Seung-Jin Jang. 2009. Are diverse political networks always bad for participatory democracy? indifference, alienation, and political disagreements. *American Politics Research*, 37(5):879–898.

Nanjiang Jiang and Marie-Catherine de Marneffe. 2021. He thinks he knows better than the doctors: BERT for event factuality fails on pragmatics. *CoRR*, cs.CL/2107.00807v1.

Ferenc Kiefer. 1987. On defining modality.

Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, pages 1367–1373, Geneva, Switzerland. COLING.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Casey A Klofstad, Anand Edward Sokhey, and Scott D McClurg. 2013. Disagreeing about disagreement: How conflict in social networks affects political behavior. *American Journal of Political Science*, 57(1):120–134.

Wojciech Kryscinski, Nitish Shirish Keskar, Bryan McCann, Caiming Xiong, and Richard Socher. 2019. Neural text summarization: A critical evaluation. In *Proceedings of the 2019 Conference on Empirical*

Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 540–551, Hong Kong, China. Association for Computational Linguistics.

George Lakoff. 1975. Hedges: A study in meaning criteria and the logic of fuzzy concepts. In *Contemporary research in philosophical logic and linguistic semantics*, pages 221–271. Springer.

Ronald W. Langacker. 2009. *Investigations in Cognitive Grammar*. De Gruyter Mouton.

Kenton Lee, Yoav Artzi, Yejin Choi, and Luke Zettlemoyer. 2015. Event detection and factuality assessment with non-expert supervision. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1643–1648, Lisbon, Portugal. Association for Computational Linguistics.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on Human Language Technologies*, 5(1):1–167.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Amnon Lotan, Asher Stern, and Ido Dagan. 2013. TruthTeller: Annotating predicate truth. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 752–757, Atlanta, Georgia. Association for Computational Linguistics.

John Lyons. 1977. Semantics cambridge university press. *Cambridge, UK*, 1.

James G MacKinnon. 2009. Bootstrap hypothesis testing. *Handbook of computational econometrics*, 183:213.

James R Martin and Peter RR White. 2005. Appraisal in English. *The language of evaluation*.

Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1906–1919, Online. Association for Computational Linguistics.

Tsvetomila Mihaylova, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, Mitra Mohtarami, Georgi Karadzhov, and James Glass. 2018. Fact checking in community forums. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

Anne-Lyse Minard, Manuela Speranza, Ruben Urizar, Begoña Altuna, Marieke van Erp, Anneleen Schoen, and Chantal van Son. 2016. MEANTIME, the NewsReader multilingual event and time corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4417–4422, Portorož, Slovenia. European Language Resources Association (ELRA).

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California. Association for Computational Linguistics.

Mitra Mohtarami, Ramy Baly, James Glass, Preslav Nakov, Lluís Màrquez, and Alessandro Moschitti. 2018. Automatic stance detection using end-to-end memory networks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 767–776, New Orleans, Louisiana. Association for Computational Linguistics.

Marius Mosbach, Maksym Andriushchenko, and Dietrich Klakow. 2021. On the stability of fine-tuning BERT: Misconceptions, explanations, and strong baselines. In *International Conference on Learning Representations*.

Ilana Mushin. 2001. Evidentiality and epistemological stance. *Narrative retelling*.

Rowan Nairn, Cleo Condoravdi, and Lauri Karttunen. 2006. Computing relative polarity for textual inference. In *Proceedings of the Fifth International Workshop on Inference in Computational Semantics (ICoS-5)*.

Elinor Ochs and Bambi Schieffelin. 1989. Language has a heart:. *Text - Interdisciplinary Journal for the Study of Discourse*, 9(1):7–26.

Frank Robert Palmer. 2001. *Mood and modality*. Cambridge university press.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic Differentiation in PyTorch. In *NIPS 2017 Workshop on Autodiff*.

Susan U Philips. 1985. William m. o'barr, linguistic evidence: Language. power, and strategy in the courtroom.(studies on law and social control.) new york: Academic, 1982. pp. xv+ 192. *Language in Society*, 14(1):113–117.

Dean Pomerleau and Delip Rao. 2017. The Fake News Challenge: Exploring how Artificial Intelligence technologies could be leveraged to combat Fake News. *Fake News Challenge*.

Amir Pouran Ben Veyseh, Thien Huu Nguyen, and Dejing Dou. 2019. Graph based neural networks for event factuality prediction using syntactic and semantic structures. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4393–4399, Florence, Italy. Association for Computational Linguistics.

Vinodkumar Prabhakaran, Tomas By, Julia Hirschberg, Owen Rambow, Samira Shaikh, Tomek Strzalkowski, Jennifer Tracey, Michael Arrigo, Rupayan Basu, Micah Clark, Adam Dalton, Mona Diab, Louise Guthrie, Anna Prokofieva, Stephanie Strassel, Gregory Werner, Yorick Wilks, and Janyce Wiebe. 2015. A new dataset and evaluation for belief/factuality. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 82–91, Denver, Colorado. Association for Computational Linguistics.

Vinodkumar Prabhakaran, Owen Rambow, and Mona Diab. 2010. Automatic committed belief tagging. In *Coling 2010: Posters*, pages 1014–1022, Beijing, China. Coling 2010 Organizing Committee.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. 2020. Stanza: A Python Natural Language Processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.

Zhong Qian, Peifeng Li, Yue Zhang, Guodong Zhou, and Qiaoming Zhu. 2018. Event factuality identification via generative adversarial networks with auxiliary classification. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 4293–4300. International Joint Conferences on Artificial Intelligence Organization.

Zhong Qian, Peifeng Li, and Qiaoming Zhu. 2015. A two-step approach for event factuality identification. In *2015 International Conference on Asian Language Processing (IALP)*, pages 103–106.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Owen Rambow, Daniel Bauer, Axinia Radeva, Meenakshi Alagesan, Gregorios A. Katsios, Tomek Strzalkowski, Claire Cardie, Mona T. Diab, Michael Arrigo, Jennifer Tracey, Adam Dalton, and Greg Dubbin. 2016. The 2016 TAC KBP BeSt evaluation. In *Text Analysis Conference*. NIST.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2931–2937, Copenhagen, Denmark. Association for Computational Linguistics.

Robert B Reich. 2005. *Reason: Why liberals will win the battle for America*. Vintage.

Anna Rogers, Olga Kovaleva, and Anna Rumshisky. 2020. A primer in BERTology: What we know about how BERT works. *Transactions of the Association for Computational Linguistics*, 8:842–866.

Sara Rosenthal. 2014. Detecting influencers in social media discussions. *XRDS: Crossroads, The ACM Magazine for Students*, 21(1):40–45.

Rachel Rudinger, Adam Teichert, Ryan Culkin, Sheng Zhang, and Benjamin Van Durme. 2018a. Neural-Davidsonian Semantic Proto-Role labeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 944–955, Brussels, Belgium. Association for Computational Linguistics.

Rachel Rudinger, Aaron Steven White, and Benjamin Van Durme. 2018b. Neural models of factuality. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 731–744, New Orleans, Louisiana. Association for Computational Linguistics.

Roser Saurí and James Pustejovsky. 2009. FactBank: a corpus annotated with event factuality. *Language resources and evaluation*, 43(3).

Roser Saurí and James Pustejovsky. 2012. Are you sure that this happened? Assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261–299.

Stephen D Shaffer. 1981. Balance theory and political cognitions. *American Politics Quarterly*, 9(3):291–320.

Ben Shapiro. 2019. *Facts don't care about your feelings*. Creators Publishing.

Xing Shi, Inkit Padhi, and Kevin Knight. 2016. Does string-based neural MT learn source syntax? In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1526–1534, Austin, Texas. Association for Computational Linguistics.

Yanchuan Sim, Brice D. L. Acree, Justin H. Gross, and Noah A. Smith. 2013. Measuring ideological proportions in political speeches. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 91–101, Seattle, Washington, USA. Association for Computational Linguistics.

Sandeep Soni, Tanushree Mitra, Eric Gilbert, and Jacob Eisenstein. 2014. Modeling factuality judgments in social media text. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 415–420, Baltimore, Maryland. Association for Computational Linguistics.

Gabriel Stanovsky, Judith Eckle-Kohler, Yevgeniy Puzikov, Ido Dagan, and Iryna Gurevych. 2017. Integrating deep linguistic features in factuality prediction over unified datasets. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 352–357, Vancouver, Canada. Association for Computational Linguistics.

Sandesh Swamy, Alan Ritter, and Marie-Catherine de Marneffe. 2017. " i have a feeling trump will win..................": Forecasting winners and losers from user predictions on twitter. *arXiv preprint arXiv:1707.07212*.

Swabha Swayamdipta and Owen Rambow. 2012. The pursuit of power and its manifestation in written dialog. In *2012 IEEE Sixth International Conference on Semantic Computing*, pages 22–29. IEEE.

TAC-KBP. 2016. Task description for source/target belief and sentiment evaluation (best) at tac 2016. `http://www.cs.columbia.edu/~rambow/best-eval-2016/task-spec-v2.4.pdf`.

Ian Tenney, Dipanjan Das, and Ellie Pavlick. 2019a. BERT rediscovers the classical NLP pipeline. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4593–4601, Florence, Italy. Association for Computational Linguistics.

Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R Thomas McCoy, Najoung Kim, Benjamin Van Durme, Sam Bowman, Dipanjan Das, and Ellie Pavlick. 2019b. What do you learn from context? Probing for sentence structure in contextualized word representations. In *International Conference on Learning Representations*.

James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. FEVER: a large-scale dataset for fact extraction and verification. *CoRR*, abs/1803.05355.

Stephen E. Toulmin. 1958. *The Uses of Argument*. Cambridge University Press.

Dietrich Trautmann, Johannes Daxenberger, Christian Stab, Hinrich Schütze, and Iryna Gurevych. 2020. Fine-grained argument unit recognition and classification. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 9048–9056. AAAI Press.

Meagan Vigus, Jens E. L. Van Gysel, and William Croft. 2019. A dependency structure annotation for modality. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 182–198, Florence, Italy. Association for Computational Linguistics.

Andreas Vlachos and Sebastian Riedel. 2014. Fact Checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 18–22, Baltimore, MD, USA. Association for Computational Linguistics.

Hanna Wallach. 2014. Big data, machine learning, and the social sciences: Fairness, accountability, and transparency. Talk at NIPS 2014 Workshop on Fairness, Accountability, and Transparency in Machine Learning.

Douglas Walton, Christopher Reed, and Fabrizio Macagno. 2008. *Argumentation schemes*. Cambridge University Press.

Douglas N Walton. 1996. *Argumentation schemes for presumptive reasoning*. Psychology Press.

Larry Wasserman. 2004. *All of statistics: a concise course in statistical inference*, volume 26. Springer.

Aaron Steven White, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. 2016. Universal decompositional semantics on Universal Dependencies. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1713–1723, Austin, Texas. Association for Computational Linguistics.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2):165–210.

Sam Wiseman, Stuart Shieber, and Alexander Rush. 2017. Challenges in data-to-document generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2253–2263, Copenhagen, Denmark. Association for Computational Linguistics.

Jiarui Yao, Haoling Qiu, Jin Zhao, Bonan Min, and Nianwen Xue. 2021. Factuality assessment as modal dependency parsing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 1540–1550. Association for Computational Linguistics.

Tianyi Zhang, Felix Wu, Arzoo Katiyar, Kilian Q Weinberger, and Yoav Artzi. 2021. Revisiting few-sample BERT fine-tuning. In *International Conference on Learning Representations*.

Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, and Michal Lukasik. 2016. Stance classification in rumours as a sequential task exploiting the tree structure of social media conversations. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2438–2448, Osaka, Japan. The COLING 2016 Organizing Committee.

Appendices

A Experimental Details

A.1 Implementation Details

All our models are implemented with PyTorch 1.9, using `roberta-large` (with 1024-dimensional embeddings) accessed from AllenNLP 2.5.1 (Paszke et al., 2017; Gardner et al., 2018). We train the models with the Adam optimizer (Kingma and Ba, 2015), using at most 20 epochs, batch size 16, and learning rate 5×10^{-6}, following Zhang et al. (2021) and Mosbach et al. (2021)'s training guidelines. We use an early stopping rule if the validation loss does not reduce for more than two epochs; this typically ends training in $5-6$ epochs. We report macro-averaged precision, recall, and F1 over the original train-test set splits of FactBank. Since fine-tuning BERT (and its variants) can be unstable on small datasets (Dodge et al., 2020), we report average performance over five random restarts for each model. To fine-tune BERT and RoBERTa models, we start with the pre-trained language model, updating both the task-specific layer and all parameters of the language model.

A.2 Significance Testing

We use a nonparametric bootstrap (Wasserman, 2004, ch. 8) to infer confidence intervals for an individual model's performance metric (precision, recall, F1) and hypothesis testing between pairs of models. We utilize 10^4 bootstrap samples of sentences for source and event identification models and 10^4 bootstrap samples of epistemic stance tuples for stance classifier in FactBank's test set to report 95% bootstrap confidence intervals (CI), via the normal interval method (Wasserman, 2004, ch. 8.3), and compare models with a bootstrap two-sided hypothesis test to calculate a p-value for the null hypothesis of two models having an equal macro-averaged F1 score (MacKinnon, 2009).[12]

B Performance of Source and Event Identification Models

B.1 Source and Event Identification

Table 6 mentions performance scores of the source and event identification models.

[12]MacKinnon's bootstrap hypothesis test has subtle differences from Berg-Kirkpatrick et al. (2012)'s in the NLP liteature; we find MacKinnon's theoretical justification clearer.

Model	Event			Source		
	Prec	Recall	F1	Prec	Recall	F1
CNN (Qian et al., 2018)	86.6	82.8	84.6	80.7	77.4	78.9
RoBERTa (Joint)	84.4	87.6	86.0	81.4	62.7	70.8
RoBERTa (Individual)	84.1	87.2	85.6	79.7	81.2	80.5

Table 6: Performance of the source and event identification models. Individual classifiers perform better than a combined classifier.

B.2 Error Analysis: Correlation with the events denoted by verb *"say"*

We conducted an error analysis of our source identification model. We tested the model to examine whether the model understands the notion of source or merely associates the notion of source with presence of vents denoted by verb "say" in a given sentence. Table 7 demonstrates that the model does not merely rely on presence or absence of such events.

"Say"	F1	Precision	Recall	#sentences
Present	84.6	86.4	82.9	147
Absent	65.2	58.4	73.8	269

Table 7: Source Error Analysis

C Performance of Epistemic Stance Classifier

C.1 Error Analysis: Negative Polarity Items

The *CT-* class is the most rare in FactBank, and it is useful to identify for a possible future use case of finding disagreements in text. For corpus exploration, an alternative to our model could be to simply view sentences with explicit negative polarity items (NPIs); such sentences[13] indeed contain a large majority (88.2%) of FactBank's gold standard *CT-* tuples. They are still uncommon within NPI-containing sentences (13.5% of such tuples are *CT-*), and quite rare within sentences without NPIs (0.33% of such tuples are *CT-*). For this challenging CT- class, the model attains a F1 score of 78.4%. To examine the model performance on CT- class in political domain, we qualitatively analyzed correct classifications. We observe that the model exhibits ability to deal with complex connections between negation-bearing constructions like *Unable to*, *refuse*, etc. (Table 8).

D External Validity: A Case Study on Hedging and power

Jalilifar and Alavi (2011) examine the relationship between an author's perceived political power and their expressed commitment to their beliefs. While

[13]Using an NPI list of: *no*, *not*, *n't*, *never*, *nobody*, *none*

• **[Author]**$_s$: Unable to **reach**$_e$ Russo in the era before cell phones, the House Speaker, Jim Wright, kept the vote open for some twenty minutes while an aide coaxed a member to change his vote to yes.
• Author: **[John Boehner]**$_s$, the Speaker of the House, refused to **address**$_e$ immigration reform in 2013.
• Author: **[People]**$_s$ are beginning to move worlds apart and find it increasingly difficult to **establish**$_e$ common ground.
• **[Author]**$_s$: Although still incapable of actually **cutting**$_e$ spending, except for needed defense, conservative leaders imply our national crisis is merely some budgeting blunder remediable through a balanced budget amendment to the Constitution.

Table 8: Examples of *CT-* epistemic stances, in sentences without explicit NPIs in PoliBelief, that BERT correctly predicts; sources are highlighted in bold, and events are underlined.

hedging and hesitations have been utilized to measure lack of commitment (Philips, 1985), political discourse can feature many more strategies beyond a simple lexicon of hedge words, such as indirect speech acts, hypothetical if-then clauses, or framing claims as questions (Fraser, 2010). Thus, analyzing hedging requires understanding of syntactic contexts within which claims are expressed, which our model can tackle. We establish the external validity of our proposed epistemic stance framework by computationally replicating the findings of Jalilifar and Alavi (2011)'s manual content analysis. To ensure the external validity of our proposed epistemic stance framework, we computationally replicate the findings of Jalilifar and Alavi (2011)'s manual content analysis.

The study examines transcripts of topically similar television interviews of three political figures, George W. Bush (at the time, incumbent U.S. president), Jimmy Carter (former U.S. president), and David Coltart (founding member of Zimbabwe's main opposition party).[14] For each interview transcript, we employ our epistemic stance classifier to predict the stance of the political figure (author source) towards all extracted events, and calculate each author's uncertainty level as the fraction of events with a *PR+* or *PS+* epistemic stance.

We find the same ordering of commitment as the previous work: Bush using the fewest uncertain *PR+/PS+* stances (5.41%), with progressively more for Carter (8.32%) and Coltart (12.2%). This follows Jalilifar and Alavi's interpretation of commitment being correlated to power (Bush being the highest status, for example).

E Case Study: Belief Holder Identification

E.1 Details of MMM Corpus

The MMM, maintained by one of the authors (*anon. for review*), is an example of a researcher-curated "artisanal data" (Wallach, 2014) collection, common in political science and communication research. Books were chosen according to a number of selection criteria and not as a representative sample of any presumed population of publications. Nominees for consideration include books appearing on best-seller lists from a number of politically-oriented Amazon book categories, mostly under the heading "Politics & Government—Ideologies & Doctrines." Additionally, all presidential primary candidates authoring a book during this period were considered, as were other officials (e.g. governors, sheriffs, senators) and ideologues attaining public prominence. Over the course of several years, scholars of American ideology have been invited to nominate additional authors for consideration, as the long-term goal is to maintain as comprehensive as possible a corpus of mass-marketed ideologically-oriented manuscripts. Among nominees, books that were more memoir than manifesto were eliminated, as were books too narrowly focused on a particular policy area.

Books in the MMM were published from 1993 through 2020, with a majority during the Obama presidential administration (233 in 2009-2016), as well as 57 from the George W. Bush presidency (2001-2008) and 80 during the Trump presidency (2017-2020).

E.2 Comparison with NER: Qualitative Examples

Table 9 describes whether the book's belief holders are recognized as named entities—three of ten are not.

Belief Holder	Detected by NER?	Belief Holder	Detected by NER?
Media	Yes	Bernie Sanders	No
Democrats	Yes	Right	Yes
Donald Trump	Yes	Republicans	No
Left	No	Courts	Yes
Conservatives	Yes	Joe Biden	Yes

Table 9: Top 10 sources detected as belief holders in Ben Shapiro's *Facts Don't Care About Your Feelings*.

E.3 Linguistic Analysis of Belief Holders

We identify two interesting linguistic phenomena among belief holders mentions.

[14] Authors also analyzed interviews by U.S. politician Sarah Palin, but we these transcripts were not available at the provided URL.

Common and Collective Nouns Many belief holders can also be described by common nouns, such as a plural form referring to classes of people (or other agents), or collective nouns denoting aggregate entities, including informally defined ones. We show several examples, along with an event toward which they have an epistemic stance.

(1) A recent survey of studies published in peer-reviewed scientific journals found that 97 percent of actively publishing climate **[scientists]**$_s$ agree that global warming has been **caused**$_e$ by human activity. (Abdul-Jabbar and Obstfeld, 2016)

(2) The **[Left]**$_s$ properly pointed out the widespread problems of racism and sexism in American society in the 1950s — and their diagnosis was to **destroy**$_e$ the system utterly. (Shapiro, 2019)

(3) The agents seized rosewood and ebony that the **[government]**$_s$ believed was illegally **imported**$_e$. (Forbes and Ames, 2012)

(4) The **[media]**$_s$ simply asserted that Clinton was **beloved**$_e$ across the land — despite never being able to get 50 percent of the country to vote for him, even before the country knew about Monica Lewinsky. (Coulter, 2009)

(5) Maybe American **[society]**$_s$ concluded, at some deep level of collective unconsciousness, that it had to **reject**$_e$ the previous generation 's model of strict fathering in favor of nurturing mothering. (Reich, 2005)

Word Sense Disambiguation If an entity is described as a belief holder, that can help disambiguate its word sense or entity type. Our model distinguishes agentive versus non-agentive versions of a geographical locations. In the following two examples, the locations or ideas "Europe" and "Silicon Valley" are belief holders with opinions toward various future scenarios (all with uncommitted *Uu* stances, which FactBank uses for all conditionals and hypotheticals). These location entities are treated as agents with political desires and intentions, perhaps more like an organizational or geopolitical NER type, despite the fact that these instances do not represent formally defined or even universally agreed-upon entities.

(6) **[Europe]**$_s$ sees it [NATO expansion] as a scheme for permanent U.S. hegemony and has decided that if the Americans want to play Romans, let Americans **pay**$_e$ the costs and **take**$_e$ the risks. (Buchanan, 1999)

(7) "Currently **[Silicon Valley]**$_s$ is in the midst of a love affair with BMI, arguing that when robots **come**$_e$ to **take**$_e$ all of our jobs, we're going to **need**$_e$ stronger redistributive policies to **help**$_e$ **keep**$_e$ families afloat," Annie Lowrey, who has a book on the subject coming July 10, wrote in New York magazine. (Beck, 2018)

By contrast, "Europe" and "Iowa" below have no epistemic stances (all edges toward sentence events are *NE*), and the entities are used simply to describe geographic locations.

(8) Napoleon was the dictator of a French state so anticlerical that many in **[Europe]**$_s$ speculated that he was the Antichrist. (Dreher, 2018)

(9) While reporters waited outside in the **[Iowa]**$_s$ cold amid a mix-up at one of Trump's rallies [...] (Abdul-Jabbar and Obstfeld, 2016)