

# Testing and Comparing Computational Approaches for Identifying the Language of Framing in Political News

Eric P. S. Baumer<sup>1,2</sup>, Elisha Elovic<sup>2</sup>, Ying “Crystal” Qin<sup>2</sup>, Francesca Polletta<sup>3</sup>, Geri K. Gay<sup>1,2</sup>

<sup>1</sup>Communication

<sup>2</sup>Information Science

<sup>3</sup>Sociology

Cornell University

University of California, Irvine

Ithaca, NY, USA

Irvine, CA, USA

{ericpsb, epe9, yq37, gkg1}@cornell.edu, polletta@uci.edu

## Abstract

The subconscious influence of framing on perceptions of political issues is well-documented in political science and communication research. A related line of work suggests that drawing attention to framing may help reduce such framing effects by enabling frame reflection, critical examination of the framing underlying an issue. However, definite guidance on how to identify framing does not exist. This paper presents a technique for identifying frame-invoking language. The paper first describes a human subjects pilot study that explores how individuals identify framing and informs the design of our technique. The paper then describes our data collection and annotation approach. Results show that the best performing classifiers achieve performance comparable to that of human annotators, and they indicate which aspects of language most pertain to framing. Both technical and theoretical implications are discussed.

## 1 Introduction

Contentious political issues are rarely understood *per se* but rather through the lens of framing. Terms such as “tax relief,” “death taxes,” “racial quotas,” “death panels,” and others have famously rallied citizens around fairly complex causes. More generally, research has shown that the way an issue is framed (Entman, 1993) – how a problem is defined, to what other problems and people it is linked, etc. – has a significant impact on both perceptions of the issue and prescriptions for action. A variety of work has shown that minor changes in language –

“global warming” vs. “climate change” (Schuldt et al., 2011), “gay civil unions” vs. “homosexual marriage” (Price et al., 2005), “not allow” vs. “forbid” (Rugg, 1941) – can significantly impact opinions.

A related but less explored line of research suggests that “frame reflection” (Schön and Rein, 1994), i.e., critical thinking about an issue’s framing, can play an important role in understanding issues and reconciling conflicts. Indeed, some recent work suggests that drawing attention to framing may help mitigate framing effects (Baumer et al., 2015). However, such reflection is no mean feat. “Various observers have noted how subtly and unconsciously [framing] operates” (Gamson and Modigliani, 1989, p. 7), making it difficult to acknowledge that an issue is being framed at all, let alone examine that framing critically or consider alternatives. Furthermore, “straightforward guidelines on how to identify [...] a frame in communication do not exist” (Chong and Druckman, 2007, p. 106).

To address this challenge, this paper compares different computational approaches for identifying the language of framing, specifically in political news coverage. The best performing classifiers achieve accuracy around 61% and F1 scores of 0.45 to 0.46, outperforming a dummy baseline and approaching or matching human performance of 73% accuracy and F1 score of 0.46. This work makes two key contributions. First, compares different techniques for identifying language invoking conceptual framing, a novel yet important task. Second, it offers evidence about what language is perceived as related to framing, helping to address the gap identified by (Chong and Druckman, 2007).

## 2 Related Work

### 2.1 Conceptual Framing

Generally speaking, in order to make sense of their interactions, people frame their experiences (Goffman, 1974). Facts “take on their meaning by being embedded in a frame [...] that organizes them and gives them coherence” (Gamson, 1989, p. 157). Frames help people “locate, perceive, identify, and label,” i.e., organize and give meaning to, information about experiences in the world. A frame consists of a variety of components, including “keywords, stock phrases, stereotype images, sources of information” (Entman, 1993, p. 52), “metaphors, exemplars, catchphrases, depictions, [...] visual images” (Gamson and Modigliani, 1989, p. 3), and other devices that provide an interpretive lens or “package.” Frames define what counts as a problem, diagnose what is causing the problem, make moral judgments about those involved, and suggest remedies for resolving the problem (Entman, 1993). Framing can significantly impact the perception of a variety of political issues. (Schuldt et al., 2011) found that belief in “global warming” was significantly lower than in “climate change,” specifically among Republicans. Gamson and Modigliani (1989) show how nuclear power is framed by such phrases as “atoms for peace,” “we have created a Frankenstein,” and “the war being waged against the environment and our health.”

Crucially, framing differs from subjectivity, sentiment, bias, and related constructs. Subjectivity detection may not effectively identify well-established, codified frames (e.g., “tax relief” or “racial quotas”). Sentiment analysis focuses on assessing the valence (e.g., positive, neutral, or negative) of an entity’s description. Bias involves a clear, often intentional, preference shown in writing for one position or opinion on an issue. In contrast, there does not exist a one-to-one mapping between framing and opinions (Gamson and Modigliani, 1989). For example, in late 2013, the international community was considering what actions should be taken against Syrian president Bashar al-Assad for using chemical weapons against rebelling citizens. Some viewed the situation as a humanitarian crisis and argued for military intervention. Others argued that al-Assad’s actions were a threat to regional security, also argu-

ing for military action. Here, different framings are used to support the same position on an issue.

In contrast, framing involves an ensemble of rhetorical elements to create an “interpretive package” (Gamson and Modigliani, 1989) that functions by altering the relative salience or importance of different aspects of an issue (Chong and Druckman, 2007). In the humanitarian crisis framing vs. regional security framing example above, a regional security framing does not negatively valence human suffering; rather, it shifts the emphasis to make other concerns more apparent. Thus, while we can draw on subjectivity, bias, and sentiment detection, identifying framing requires features and techniques that go beyond any one of these related concepts.

### 2.2 Related Computational Approaches

Some computational work explored concepts related to conceptual framing. For example, Choi et al. (2012) identify hedging in discussion of GMOs using an SVM trained on n-grams from annotated cue phrases. Greene and Resnik (2009) showed how examining grammatical construction (i.e., syntax) can reveal implicit sentiment; for example, passive and active voice imply different degrees of agency and causality. Recasens et al. (2013) used edits from Wikipedia intended to enforce a neutral point of view to identify biased sentences and the terms indicative of that bias.

Relatively little work has been done on identifying frames per se. Lind and Salo (2002) used co-occurrence frequencies to examine the framing of feminism in news media. Matthes and Kohring (2008) take a mixed methods approach, asking human coders to annotate the occurrence of certain features in a text problem definition, attribution of causation, moral evaluation, etc. then cluster the text using that coding to identify high-level frames. Boydston et al. (2013) suggest that approaches based on hierarchical topic modeling may be an effective means of identifying both issue-specific and generic frames.

Such techniques, while useful from an analytic standpoint, are not as directly relevant here. The work described in this paper does not aim to identify the framing in a text. Rather, it seeks to determine what language is perceived as being most related to framing, especially by lay-persons, as a means of

supporting reflection on that framing.

### 3 Exploratory Pilot Study

While the framing literature provides some guidance (Entman, 1993; Chong and Druckman, 2007), little work has explored how exactly non-elites go about identifying framing (Chong and Druckman, 2007). Thus, we conducted a pilot investigating to explore laypersons' understandings and identification of framing.

#### 3.1 Methods

We recruited undergraduate students at two major US universities. Participants were asked to read an article that framed the issue of health care in terms of equality (as opposed to cost). Previous work showed that reading this article was associated with stronger support for a national healthcare system (Druckman et al., 2012). Participants were then asked to re-read the article and highlight any words or phrases they believed were related to framing. Specifically, participants were given the following prompt and instructions:

*Political issues can often be complex, contentious, and difficult to understand. One way of making sense of these issues, and the different positions that one can take on an issue, is to think about the frames that structure debate about the issue. Frames help organize facts and information. They help define what counts as a problem, diagnose the problem's causes, and suggest remedies for solving the problem. These ways of thinking have lots of different parts, including stereotypes, metaphors, images, catchphrases, and other elements.*

*These different frames are often associated with a particular way of talking about or communicating about an issue. Certain words or phrases might suggest that one or another frame.*

*Please use the tool at the link below to highlight the words or phrases that help you identify the framings used in the article you just read.*

After a student finished highlighting the article, s/he also participated in a debriefing interview where a researcher asked her or him about what s/he highlighted and why.

We recruited total of 47 students; 20 completed the task in person, who completed the debriefing interview immediately, and 27 used an online annotation tool, who completed the debriefing over the phone within 24 hours.

#### 3.2 Results

First, we sought to determine the extent to which study participants' annotations agreed with one another. Do different people see the same words and phrases as being related to framing? An intraclass correlation (ICC) among participants' annotations of 0.757 indicated that the annotators demonstrated a moderately high degree of agreement as to which words and phrases were most related to framing.

As an example, one article about health care contained the sentence: "A good doctor might recognize the regenerative powers of the body politic and come up with a comprehensive treatment plan that also attacks root causes including the twin cancers of racism and poverty." Three students highlighted the entire sentence, another three highlighted only the phrase "twin cancers of racism and poverty," and one more highlighted "twinpoverty," "regenerative powers," and "body politic."

Several important insights were also derived from the debriefing interviews. Participants drew a distinction between facts and opinions, the later being more relevant to framing. For example, statistics were rarely seen as related to framing. Also, framing often dealt not only with a word itself but also with aspects of its context and its relationships with other terms in the article. For example, a participant might highlight just the word "but" because it indicates an important rhetorical shift and, implicitly, the article's take on the issue. Similarly, latent relationships between an individual word and the article's main argument also played an important role. For example, the article participants read emphasized "disparities" between healthcare available to the wealthy and to the working class. Many participants indicated that they would highlight any words or phrases that drew attention to such disparities. These insights, in conjunction with the theoretical literature on conceptual framing, were used to guide feature selection.

### 4 Data

We sought to develop a classifier that could automatically identify the language in a text that most related to framing. We chose to focus on news coverage rather than, say, opinion and editorial columns. Framing likely occurs in a more apparent, poten-

tially obvious fashion in opinion articles. In ostensibly "straight" news, though, framing may be harder to identify. Thus, this context would benefit more from a classifier that could automatically draw attention to frame-invoking language.

For training data, we wanted political texts where lay readers had indicated the words and phrases they perceived as most related to framing. Lay annotators were used instead of experts because the classifier's purpose is to support frame reflection among the lay public. Thus, the words and phrases the classifier highlights should align with that population's perception of framing. To our knowledge, no such data set exists. So, we used Mechanical Turk (Snow et al., 2008) and university students to build an annotated dataset that could be used for training and testing.

#### 4.1 Collection

We began by collecting political news articles from top 15 online sources of news, as determined by Alexa rankings (<http://www.alexa.com/topsites/category/Top/News>). We excluded sources outside the US (e.g., BBC), news aggregators (e.g., Yahoo News, Google News), blogs (Huffington Post), and sites without a dedicated politics feed (e.g., USA Today, weather.com). Doing so left eight sources: CNN, NYT, Fox, NBC, Washington Post, ABC, LA Times, Reuters.

For each source, we collected all items on their politics-specific RSS feed on two separate days roughly six months apart to provide content about diverse issues and events: Tuesday November 12, 2013, and Thursday May 15, 2014. We manually removed duplicate posts, "round-up" style posts that simply summarized and linked to other stories, video-only posts, and other non-textual content, resulting in a total of 205 documents. Of these, we randomly selected 75 to be annotated.

#### 4.2 Annotation

Each article was annotated by five to 13 annotators, who were either Mechanical Turk (MTurk) workers or students at one of two major US universities. The task first asked the annotator to read the article, then gave the same directions from the framing prompt described above.

To encourage MTurk workers to pay attention to

the task and complete high-quality work, we provided a scheme for bonus payments. Every word a worker annotated that was also annotated by at least two others (i.e., a majority of the 5 workers annotating each document) would earn the worker a \$0.02 bonus (two cents). Each word s/he annotated that was annotated by no other work would reduce the bonus by \$0.005 (half a cent). Workers were then linked to our web tool where they could complete and submit their annotations. Student annotators received no agreement-based incentive but were granted extra course credit.

#### 4.3 Quality Assurance

Crowd workers do not always provide reliable annotations (Snow et al., 2008). For example, we noted multiple instances where annotators had only annotated about a dozen words in an article of several hundred words. These seemed likely to be cases in which the annotator was completing the task as quickly as possible without paying much attention. By comparing the annotations with those collected during our pilot study, in which participants' justifications during the debriefing ensured higher attention and quality, we developed the following criteria for identifying questionable annotations. Those that did not pass at least three of these five requirements were removed from the analysis.

1. All Annotations Short — While some annotations of short words, such as conjunctions, could be meaningful (see example above), we encountered a number of annotations where every annotated phrase consisted of only one or two words at a time. Thus, we required that the average annotated contiguous segment be at least 3 words long.

2. Few Words Annotated — When very few words in a document are annotated, we suspect the annotator may have been completing the task as quickly as possible and paying little attention. Thus, we required that each annotator's work include enough annotated words to total more than 5% of the document's length.

3. Large Contiguous Passages without Annotation — While pilot study participants pointed out portions of a text that consisted of "facts and figures," these were generally relatively short. Thus, we require the longest block of text without any annotations to be no more than one third the length of

the entire document.

4. Large Contiguous Passages Entirely Annotated — Large, unbroken annotations rarely occurred in our pilot study. Such annotations may also have indicated that the annotator was not attending to the entire content being annotated. Thus, we require that the longest contiguous annotation be no more than 120 words long.

5. Annotating Solely Stop Words — A number of annotations include only very common words, such as articles, prepositions, forms of the verb “to be,” etc. Single words such as “but,” “all,” or “not” could arguably be related to framing. However, accounts from our pilot study participants made us less likely to believe that other single words, such as “an,” “of,” or “that,” instantiated framing. Thus, we require that no more than 3 annotated passages consistent entirely of such stop words<sup>1</sup>.

We also encountered two situations in which pairs of MTurk workers submitted virtually identical annotations for multiple documents. Following up with the workers, we discovered that in one situation a husband and wife team had actually been working together. Based on our pilot study, we would expect annotators to agree to some degree, but we would not want such agreement to arise because annotators were collaborating. Thus, we also exclude these annotations where we suspect the possibility of collaboration.

We only include in our dataset documents with at least three valid annotations. In total, the resulting data set includes 74 articles containing 53,878 total words ( $M=728.1$  words per article), each with three to 11 valid annotations ( $Mdn=6$ ). The data set includes 59,948 annotated words across 507 annotations from 372 annotators ( $M=122.8$  annotated words per article). The full data set is available at <http://hdl.handle.net/1813/39216>.

## 5 Classifier Design and Implementation

As argued above, drawing attention to framing-related language may both mitigate frame effects (Baumer et al., 2015) and facilitate frame reflection (Schön and Rein, 1994). This section describes the features used to train a classifier to identify framing along with justifications for each, different subsets of features that were tested, our classifier selection,

and the training and testing methods.

### 5.1 Features and Subsets

We treat each word as a data point to be classified as framing-related or not. Feature extraction began with splitting each article into sentences with NLTK (Bird et al., 2009) then using Stanford CoreNLP (Klein and Manning, 2003; De Marneffe et al., 2006) to parse each sentence, obtain POSs and lemmatized forms for each word, etc. We then construct a feature vector for each non-stop word. Table 1 lists all features used. The remainder of this subsection describes the justification for each feature, as well as several subsets of features.

Framing is often instantiated by specific “key-words, stock phrases,” (Entman, 1993, p. 52) or “catchphrases” (Gamson and Modigliani, 1989, p. 3). Therefore, we include lexical features that capture the specific words used. Furthermore, many of our pilot study participants pointed out that a given word might be seen as related to framing because of the other words near which it appears. Therefore, each of these features includes a contextual window of up to two words before and after the word being classified (Recasens et al., 2013).

Participants in our pilot study said physical locations, such as the names of states, were often not related to framing. However, they sometimes saw names of political figures or experts as indicating framing. Thus, we include the named entity type as a feature, both of the word itself and of its context.

Pilot study participants also mentioned that a word’s relationship to the remainder of a document and its overall thesis played important roles. A number of similar structural aspects of the document, both explicit and latent, may help draw out these relationships.

Several specific types of terms may be indicative of framing. For example, “depictions,” “visual images,” and figurative language such as metaphor (Gamson and Modigliani, 1989) often invoke frames. For imagery, we use previously established imagery ratings for 1818 common words (Paivio et al., 1968; van der Veur, 1975). Words that appear in this list are given an imagery rating, either low (first quantile), medium (second and third quantile), or high (fourth quantile) imagery. The context feature represents the average imagery of the

Feature	Description	Feature Subset(s)
Token $\pm 1, \pm 2$	The word token itself and the tokens in its context.	Lexical, Theoretical
Lemma $\pm 1, \pm 2$	The lemmatized word and the lemmas in its context.	Lexical, Theoretical
Bigrams and trigrams	All bigrams and trigrams of which the word is a part.	Lexical, Theoretical
POS $\pm 1, \pm 2$	The word's and its context's part(s) of speech.	Grammatical
Root	Whether the word is the sentence's parse tree's root.	Grammatical
Relation and Role	The grammatical relations in which the word is involved and its role in those relations, e.g., passive subject of a verb.	Grammatical
Named Entity $\pm 1, \pm 2$	Named entity type of the word and its context.	Grammatical
In Title	Whether the word appears in the article's title.	Document
Sentence Length	The number of words in the sentence.	Document
Sentence Position	Whether the word appears in the first third of the sentence, the middle third, or the last third.	Document
TFIDF	The word's tf-idf score, grouped into 8 bins of increasing size.	Document
Imagery & Context	Imagery rating of the word and average imagery rating of its context (Paivio et al., 1968; van der Veur, 1975).	Theoretical, Dictionaries
Figurativeness & Context	Figurativeness rating of the word and average figurativeness rating of its context (Gamson and Modigliani, 1989; Turney et al., 2011).	Theoretical, Dictionaries
Abstractness & Context	Abstractness rating of the word and average abstractness rating of its context (Gamson and Modigliani, 1989; Turney et al., 2011).	Theoretical, Dictionaries
Dictionaries & Context	One feature each for whether word or context is a subjective word (Riloff and Wiebe, 2003), a report verb (Recasens et al., 2013), a hedge (Hyland, 2005), a factive verb (Hooper, 1975), an entailment (Berant et al., 2012), an assertive word (Hooper, 1975), a bias word (Recasens et al., 2013), a negative word (Liu et al., 2005), or a positive word (Liu et al., 2005).	Theoretical, Dictionaries

Table 1: Name and description of all features for each word, as well as the set(s) to which each feature belongs.

word's context. For figurative language, we used Turney et al.'s (Turney et al., 2011) approach of measuring figurativeness based on the absolute difference in the concreteness of two terms that are grammatically related. The figurativeness score for a given word is the average absolute difference between its concreteness and the concreteness of every word with which it is in some grammatical relationship. We also used Turney et al.'s (Turney et al., 2011) approach to rate the individual abstractness of each word and its context. Lastly, subjective words, hedges, entailments, and other terms that perform specific psycholinguistic functions (Recasens et al., 2013) may also be useful in identifying fram-

ing. For each, we include one feature for the word itself and one feature for the two-word context around the word.

The features used here are informed by a combination of theoretical literature on framing, our own pilot studies, and prior work in computational linguistics. However, we have little means of knowing a priori which of these features will be most important or even necessary. Therefore, in the interest of developing the most perspicacious model, we test several feature subsets, as noted in Table 1. Lexical features involve only words that actually occur in the text. Grammatical features use only aspects of grammatical structure. Document features use various ex-

pllicit and latent aspects of the document’s structure. Theoretical features are those specifically mentioned in the theoretical literature on framing. Finally, the Dictionaries feature set tests whether there might be specific terms that invoke framing.

## 5.2 Training and Testing

We implemented and tested a variety of different classifiers, including Stochastic Gradient Descent (SGD), Multinomial Naïve Bayes, Perceptron, Nearest Neighbor, Logistic Regression, and Passive Aggressive classifiers. In several tests, the Naïve Bayes classifier performed best, so we used a Naïve Bayes classifier throughout.

As described above, our data include three to 11 annotators’ annotations for each word in the corpus. To create training data, we aggregated these annotations such that any word highlighted by at least one fourth of the annotators was treated as framing-related (i.e., true positive) for training and testing purposes, and the remaining words were treated as not frame-related (i.e., true negative).

These data, in the combinations of features described above, were used to train our ensemble classifier using scikit-learn (Pedregosa et al., 2011) using ten-fold random shuffle cross-validation, as well as a random dummy classifier based on class distributions in the training set. Performance in terms of f-score, accuracy, precision, and recall was averaged across the ten folds.

## 6 Results

This section summarizes results, highlighting some important aspects thereof, while the subsequent Discussion section considers interpretation and broader implications. Figure 1 presents a summary of results for each feature set, including comparison with the dummy baseline and the aggregate human performance. The Document Structure feature set performed very poorly, identifying 0 true positives, so we exclude it from the detailed results report.

Since overall accuracy does not vary drastically, we focus on other performance metrics. Except for the Dictionaries, all feature sets perform statistically significantly better than the dummy (ANOVA with Tukey’s posthoc  $p < 0.001$ ). F1 scores among the top three performers (Lexical, Theoretical, and All

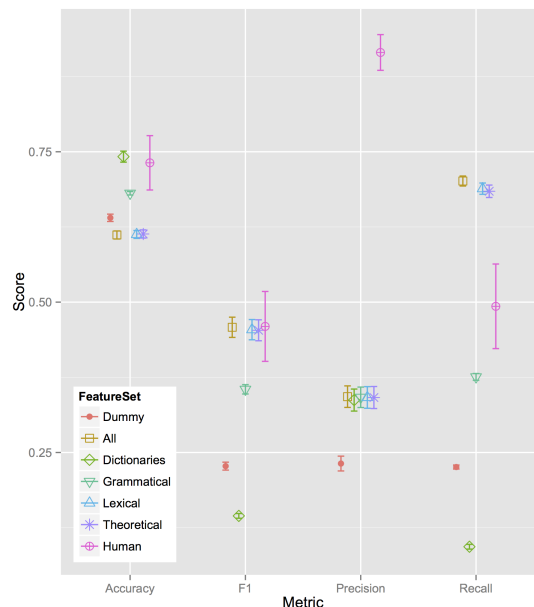


Figure 1: Performance of each feature set, as well as the dummy baseline and human annotators, on accuracy, F1, precision, and recall. Three feature sets (Lexical, Theoretical, and All features) match human annotators on F1 and outperform human recall, but humans demonstrate higher precision.

features) are statistically indistinguishable. Furthermore, each of these three top performers matches aggregate human annotator performance.

Looking at precision and recall, we can see that the classifiers and the human annotators make different trade offs. Precision for all feature sets is around 34% (all statistically significantly better than the dummy and significantly indistinguishable from one another), while human average precision is 91.5%. On the other hand, the three top performing feature sets (All, Lexical, and Theoretical) all achieve recall around 70%, while average recall for the human annotators is only 49.3%.

It is also important to note that human performance is calculated by comparing each individual annotator to the aggregate of all the annotators. Because each individual is part of that aggregate, precision scores for the humans are fairly high. We consider comparison with human performance further in the discussion below.

We also examine the most influential features for classifiers using each of the three top performing

feature sets. For the Lexical feature set, the most informative features for negative cases are bigrams that begin with quotation marks. Since we manually assign all punctuation, including quotation marks, as true negatives, this is perhaps unsurprising. For positive cases, many of the "offset" style features emerge as informative. For example, words that appear one or two words before a comma or period are more likely to involve framing. Similarly, words that occur just before or just after prepositions or conjunctions — to, and, in, of, etc. — are more likely classified as framing (i.e., positive). This result aligns with some of our pilot study findings about the relationships between such function terms the words surrounding them. Interestingly, though, these features do not resemble the catchphrases and keywords described in the framing literature (Entman, 1993; Gamson and Modigliani, 1989).

For the Theoretically Informed feature set, which adds in imagery, figurativeness, and other dictionary-based features, roughly the same kinds of lexical features are most important in identifying negative cases. For positive cases, constructs such as descriptiveness and abstractness play important roles, but mostly when words in the context do *not* appear in these dictionaries, calling into question the importance of imagery, metaphor, and other figurative language (Gamson and Modigliani, 1989). Other dictionary-based features, such as being an entailment word, having an entailment in context, having a bias term in context, or having a subjective term in context, become important in identifying positive cases. Some of the lexical features, such as proximity to a comma or period, also remain informative.

With All features, which adds features for document structure and grammatical structure, two major differences occur. First, some elements of document structure become important features for positive cases, including sentence length and TFIDF. Second, various part of speech tags, mostly NN and NNP, become important to identifying framing. These findings suggest that the choice of terms used to label concepts or entities can indicate framing. Indeed, entity type, both of the word and its context, also emerges as an informative feature.

## 7 Discussion

Chiefly important among the results, three of our feature sets were able to achieve F1 scores on par with those of human annotators. This result provides encouraging evidence that a machine classifier can effectively accomplish the task of identifying the language that invokes framing in political news coverage.

That said, examining the results more closely exposes that, in order to achieve this level of performance, the classifier makes a trade-off. Specifically, the classifier is far more aggressive than humans, resulting in significantly higher recall but lower precision. We did experiment with different post-hoc decision thresholds to make the classifier less aggressive. However, as the decision threshold rose, recall fell much more quickly than precision increased, resulting in lower overall F1 scores. Moreover, this precision-recall trade off may have different ramifications in different applications. For example, in terms of supporting frame reflection, would it be harmful or distracting for a machine classifier to mark too many words as frame-invoking, thereby potentially overwhelming a potential user? Or would it be worse if the classifier were too sparse, missing certain important key words or phrases? These are questions for later empirical work that incorporates the results of this classifier into interactive systems to support frame reflection.

Also, our results above identifying important features within the classifier contribute to and build on prior computational work. For example, Recasens et al. (Recasens et al., 2013) find entailment (Berant et al., 2012), implicature (Karttunen, 1971), subjectivity (Riloff and Wiebe, 2003), and other related constructs helpful in identifying bias. The results above suggest that, when included, such features also emerge as important for identifying the language of framing. However, the feature sets that include dictionaries of these terms do not perform statistically significantly better than those feature sets without them. This result supports our initial assertion that bias and framing, while conceptually related, are separate constructs that are each perceived and instantiated via different linguistic cues.

These findings also relate to recent efforts at identifying which frame(s) are operating in a text (e.g.,



Boydstun et al., 2013). On the one hand, the performance of the Lexical feature set suggests that topic modeling, especially including certain n-grams, may prove an effective approach. That said, Boydstun et al. are more interested in determining *which* frames are at work in a text as a whole, whereas this paper focuses more on determining *where* within a text frames are invoked. Thus, different computational approaches and features may prove effective for each task.

Interestingly, the grammatical structure feature set is not one of the top performers. Given our expectations, including the importance that pilot study participants placed on structural relationships within the sentence, the role of grammatical construction both in metaphors (Turney et al., 2011) in framing more broadly (Fairclough, 1999), and prior computational work on implicit sentiment (Greene and Resnik, 2009), this result appears fairly surprising. It could be that grammatical structure alone is not sufficient to identify framing, but even combining grammatical structure with other features does not significantly improve performance. Perhaps, then, grammatical construction matters less than the specific words chosen. On the other hand, lexical features may be topic-specific, such that even obtaining two samples six months apart still resulted in the same buzzwords invoking framing. Future work should examine more closely the role that structural features play, or perhaps do not play, in invoking framing.

This point also draws attention to some practical implications. Performing a full grammatical parse can be computationally intensive. If using other features that do not require a full parse can achieve comparable performance, then perhaps real time applications, such as analyzing live speeches as they happen, could employ only features that are relatively quicker and easier to extract.

Finally, we note that the annotated data analyzed here come from political news stories in US mainstream media. Since these sources ostensibly strive for impartiality, framing in these data may occur implicitly or unconsciously. Future work should compare these results with similar analyses of texts containing more explicit framing, such as opinion columns, campaign speeches, or political advertisements. Differences in how framing is identified may

give important clues to how framing operates in different contexts. Similarly informative insights could be gained by comparing lay-persons' annotations with framing experts'.

## 8 Conclusion

“Facts have no intrinsic meaning. They take on their meaning by being embedded in a frame” (Gamson, 1989, p. 157). Given framing's pervasive influence, this paper argues for the importance of computational techniques that can identify and draw attention to the language of framing. Doing so can help support frame reflection (Schön and Rein, 1994) and, thereby, deeper understanding of and engagement with political issues.

This paper both develops a computational approach to identifying framing and tests how well different linguistic features indicate frame-invoking language. Results suggest grammatical structure alone as the most important indicator of framing. However, other data less computationally demanding to extract, such as lexical features (e.g., tokens and n-grams), can prove almost as effective.

In sum, the paper makes two main contributions. First, it provides a technical contribution by identifying a task of importance and demonstrating a technique that performs close to as well as humans. Second, the paper makes a theoretical contribution, helping to provide “guidelines on how to identify [] a frame in communication” (Chong and Druckman, 2007, p. 106). The data set of annotations released with this paper may also prove a valuable resource for future analyses of framing.

## Acknowledgments

This material is based upon work supported by the NSF under Grant No. IIS-1110932. Thanks to the Turker and student annotators, to Andrea Lin for research assistance, and to Cristian Danescu-Niculescu-Mizil and to Peter Turney for sharing technical resources.

## References

- Eric P. S. Baumer, Francesca Polletta, Nicole Pierski, and Geri K. Gay. 2015. A Simple Intervention to Reduce Framing Effects in Perceptions of Global Climate Change. *Environmental Communication (to appear)*.

- Jonathan Berant, Ido Dagan, Meni Adler, and Jacob Goldberger. 2012. Efficient tree-based approximation for entailment graph learning. In *Proc ACL*, pages 117–125.
- Steven Bird, Edward Loper, and Ewan Klein. 2009. *Natural Language Processing with Python*. O’Reilly Media Inc.
- Amber E. Boydston, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2013. Identifying Media Frames and Frame Dynamics Within and Across Policy Issues. In *New Directions in Analyzing Text as Data Workshop*, London.
- Eunsol Choi, Chenhao Tan, Lillian Lee, Cristian Danescu-Niculescu-Mizil, and Jennifer Spindel. 2012. Hedge Detection as a Lens on Framing in the GMO Debates: A Position Paper. In *Proc ACL Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, number July, pages 70–79.
- Dennis Chong and James N. Druckman. 2007. Framing Theory. *Annual Review of Political Science*, 10(1):103–126, June.
- M.C. De Marneffe, Bill MacCartney, and C.D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proc LREC*, Genoa, Italy.
- James N. Druckman, Jordan Fein, and Thomas J. Leeper. 2012. A Source of Bias in Public Opinion Stability. *American Political Science Review*, 106(02):430–454, May.
- Robert M. Entman. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4):51–58.
- Norman Fairclough. 1999. Global Capitalism and Critical Awareness of Language. *Language Awareness*, 8(2):71–83.
- William A. Gamson and Andre Modigliani. 1989. Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach. *The American Journal of Sociology*, 95(1):1–37, July.
- William A. Gamson. 1989. News as Framing. *American Behavioral Scientist*, 33(2):157–161, November.
- Erving Goffman. 1974. *Frame Analysis*. Harvard University Press, Cambridge, MA.
- Stephan Greene and Philip Resnik. 2009. More than Words: Syntactic Packaging and Implicit Sentiment. In *Proc HLT*, number June, pages 503–511, Boulder, CO.
- Joan B. Hooper. 1975. On Assertive Predicates. In J. Kimball, editor, *Syntax and Semantics*, pages 91–124. Academic Press, New York, volume 4 edition.
- Ken Hyland. 2005. *Metadiscourse: Exploring Interaction in Writing*. Continuum, London and New York.
- Lauri Karttunen. 1971. Implicative Verbs. *Language*, 47(2):340–358.
- Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. Sapporo, Japan.
- Rebecca Ann Lind and Colleen Salo. 2002. The Framing of Feminists and Feminism in News and Public Affairs Programs in U.S. Electronic Media. *Journal of Communication*, 52(1):211–228.
- Bing Liu, Minqing Hu, and Junsheng Cheng. 2005. Opinion Observer: analyzing and comparing opinions on the Web. In *Proc WWW*, pages 342–351.
- Jörg Matthes and Matthias Kohring. 2008. The Content Analysis of Media Frames: Toward Improving Reliability and Validity. *Journal of Communication*, 58(2):258–279.
- Allan Paivio, John C Yuille, and Stephen A Madigan. 1968. Concreteness, Imagery, and Meaningfulness Values for 925 Nouns. *Journal of Experimental Psychology*, 76(1).
- Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. 2011. Scikit-learn : Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Vincent Price, Lilach Nir, and Joseph N. Cappella. 2005. Framing Public Discussion of Gay Civil Unions. *Public Opinion Quarterly*, 69(2):171–212.
- Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic Models for Analyzing and Detecting Biased Language. In *Proc ACL*, pages 1650–1659, Sofia, Bulgaria.
- Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proc EMNLP*, pages 105–112.
- Donald Rugg. 1941. Experiments in Wording Questions: II. *Public Opinion Quarterly*, 5(1):91–92.
- Donald A. Schön and Martin Rein. 1994. *Frame Reflection: Toward the Resolution of Intractable Policy Controversies*. Basic Books, New York.
- Jonathon P. Schuldt, S. H. Konrath, and N. Schwarz. 2011. ”Global warming” or ”climate change”? : Whether the planet is warming depends on question wording. *Public Opinion Quarterly*, 75(1):115–124, February.
- Rion Snow, Brendan O’Connor, Daniel Jurafsky, Andrew Y Ng, Dolores Labs, and Capp St. 2008. Cheap and Fast But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks. In *Proc EMNLP*, number October, pages 254–263, Honolulu, HI.
- Peter D. Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. Literal and Metaphorical Sense Identification through Concrete and Abstract Context. In *Proc*

*EMNLP*, volume 2, pages 680–690, Edinburgh, Scotland.

Barbara W. van der Veur. 1975. Imagery Rating of 1,000 Frequently Used Words. *Journal of Educational Psychology*, 67(1):44–56.