The Media Frames Corpus: Annotations of Frames Across Issues

Dallas Card[1]   **Amber E. Boydstun**[2]   **Justin H. Gross**[3]   **Philip Resnik**[4]   **Noah A. Smith**[1]

[1]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[2]Department of Political Science, University of California, Davis, CA 95616, USA
[3]Department of Political Science, University of Massachusetts, Amherst, MA 01003, USA
[4]UMIACS, University of Maryland, College Park, MD 20742, USA

dcard@cmu.edu   aboydstun@ucdavis.edu   jhgross@polsci.umass.edu
resnik@umiacs.umd.edu   nasmith@cs.cmu.edu

Abstract

We describe the first version of the Media Frames Corpus: several thousand news articles on three policy issues, annotated in terms of media *framing*. We motivate framing as a phenomenon of study for computational linguistics and describe our annotation process.

1 Introduction

An important part of what determines how information will be interpreted by an audience is how that information is *framed*. Framing is a phenomenon largely studied and debated in the social sciences, where, for example, researchers explore how news media shape debate around policy issues by deciding what aspects of an issue to emphasize, and what to exclude. Theories of framing posit that these decisions give rise to thematic sets of interrelated ideas, imagery, and arguments, which tend to cohere and persist over time.

Past work on framing includes many examples of issue-specific studies based on manual content analysis (Baumgartner et al., 2008; Berinsky and Kinder, 2006). While such studies reveal much about the range of opinions on an issue, they do not characterize framing at a level of abstraction that allows comparison *across* social issues.

More recently, there have also been a handful of papers on the computational analysis of framing (Nguyen et al., 2015; Tsur et al., 2015; Baumer et al., 2015). While these papers represent impressive advances, they are still focused on the problem of automating the analysis of framing along a single dimension, or within a particular domain.

We propose that framing can be understood as a general aspect of linguistic communication about facts and opinions on any issue. Empirical assessment of this hypothesis requires analyzing framing in real-world media coverage. To this end, we contribute an initial dataset of annotated news articles, the Media Frames Corpus (version 1). These annotations are based on 15 general-purpose meta-frames (here called "framing dimensions") outlined below, which are intended to subsume all specific frames that might be encountered on any issue of public concern.

Several features of this annotation project distinguish it from linguistic annotation projects familiar to computational linguists:

- A degree of subjectivity in framing analysis is unavoidable. While some variation in annotations is due to mistakes and misunderstandings by annotators (and is to be minimized), much variation is due to valid differences in interpretation (and is therefore properly preserved in the coding process).
- Annotator skill appears to improve with practice; our confidence in the quality of the annotations has grown in later phases of the project, and this attribute is not suppressed in our data release.

All of the annotations and metadata in this corpus are publicly available, along with tools to acquire the original news articles usable by those who have an appropriate license to the texts from their source (Lexis-Nexis).[1] This dataset and planned future extensions will enable computational linguists and others to develop and empirically test models of framing.

2 What is Framing?

Consider a politically contested issue such as same-sex marriage. Conflicting perspectives on this issue compete to attract our attention and influence our opinions; any communications about

---

[1]https://github.com/dallascard/media_frames_corpus

the issue—whether emanating from political parties, activist organizations, or media providers—will be fraught with decisions about how the issue should be defined and presented.

In a widely cited definition, Entman (1993) argues that "to frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described." Further elaborations have emphasized how various elements of framing tend to align and cohere, eventually being deployed "packages" which can be evoked through particular phrases, images, or other synecdoches (Gameson and Modigliani, 1989; Benford and Snow, 2000; Chong and Druckman, 2007). These may take the form of simple slogans, such as *the war on terror*, or more complex, perhaps unstated, assumptions, such as the rights of individuals, or the responsibilities of government. The patterns that emerge from these decisions and assumptions are, in essence, what we refer to as framing.[2]

Traditionally, in the social sciences, framing is studied by developing an extensive codebook of frames specific to an issue, reading large numbers of documents, and manually annotating them for the presence of the frames in the codebook (e.g., Baumgartner et al., 2008; Terkildsen and Schnell, 1997). Computational linguists therefore have much to offer in formalizing and automating the analysis of framing, enabling greater scale and breadth of application across issues.

3 Annotation Scheme

The goal of our annotation process was to produce a corpus of examples demonstrating how the choice of language in a document relates to framing in a non-issue-specific way. To accomplish this task, we annotated news articles with a set of 15 cross-cutting framing dimensions, such as economics, morality, and politics, developed by Boydstun et al. (2014). These dimensions, summarized in Figure 1, were informed by the framing literature and developed to be general enough to be applied to any policy issue.

For each article, annotators were asked to identify any of the 15 framing dimensions present in

---

Economic: costs, benefits, or other financial implications
Capacity and resources: availability of physical, human or financial resources, and capacity of current systems
Morality: religious or ethical implications
Fairness and equality: balance or distribution of rights, responsibilities, and resources
Legality, constitutionality and jurisprudence: rights, freedoms, and authority of individuals, corporations, and government
Policy prescription and evaluation: discussion of specific policies aimed at addressing problems
Crime and punishment: effectiveness and implications of laws and their enforcement
Security and defense: threats to welfare of the individual, community, or nation
Health and safety: health care, sanitation, public safety
Quality of life: threats and opportunities for the individual's wealth, happiness, and well-being
Cultural identity: traditions, customs, or values of a social group in relation to a policy issue
Public opinion: attitudes and opinions of the general public, including polling and demographics
Political: considerations related to politics and politicians, including lobbying, elections, and attempts to sway voters
External regulation and reputation: international reputation or foreign policy of the U.S.
Other: any coherent group of frames not covered by the above categories

Figure 1: Framing dimensions from Boydstun et al. (2014).

the article and to label spans of text which cued them. Annotators also identified the dominant framing of the article headline (if present), as well as for the entire article, which we refer to as the "primary frame." Finally, newspaper corrections, articles shorter than four lines of text, and articles about foreign countries were marked as irrelevant. There were no constraints on the length or composition of annotated text spans, and annotations were allowed to overlap. The last framing dimension ("Other") was used to categorize any articles that didn't conform to any of the other options (used in $< 10\%$ of cases). An example of two independent annotations of the same article is shown in Figure 2.

For the initial version of this corpus, three policy issues were chosen for their expected diversity of framing and their contemporary political relevance: immigration, smoking, and same-sex marriage. Lexis-Nexis was used to obtain all articles matching a set of keywords published by a set of 13 national U.S. newspapers between the years 1990 and 2012.[3] Duplicate and near-duplicate articles were removed and randomly selected articles were chosen for annotation for each issue (see supplementary material for additional details).

---

[2]A distinct though related usage, known as "equivalence framing" in psychology, refers to different phrasings of semantically equivalent expressions (e.g., is an 8-ounce glass containing 4 ounces of water *half empty* or *half full*?).

[3]The immigration articles extend back to 1969, though there are few before 1980.

Annotation guidelines for the project are documented in a codebook, which was used for training the annotators. The codebook for these issues was refined in an ongoing manner to include examples from each issue, and more carefully delineate the boundaries between the framing dimensions.

4 Annotation Process

Our annotation process reflects the less-than-ideal circumstances faced by academics requiring content analysis: relatively untrained annotators, high turnover, and evolving guidelines. Our process is delineated into three stages, summarized in Table 1 and discussed in detail below. Each stage involved 14–20-week-long rounds of coding; in each round, annotators were given approximately 100 articles to annotate, and the combinations of annotators assigned the same articles were rotated between rounds. Our annotators were undergraduates students at a U.S. research university, and a total of 19 worked on this project, with 8 being involved in more than one stage. The average number of frames identified in an article varied from 2.0 to 3.7 across annotators, whereas the average number of spans highlighted per article varied from 3.4 to 10.0. Additional detail is given in Table 1 in the supplementary material.

Stage	Issue	Articles	Av. annotators per article
1	Immigration	4,113	1.2
1	Smoking	4,077	1.2
2	Same-sex marriage	6,298	2.2
3	Immigration	5,549	2.2

Table 1: Summary of the number of articles annotated and average number of annotators per article

4.1 Stage 1

During the first stage, approximately 4,000 articles on each of immigration and smoking were annotated, with approximately 500 articles in each group annotated by multiple annotators to measure inter-annotator agreement. Our goals here were high coverage and ensuring that the guidelines were not too narrowly adapted to any single issue. Annotators received only minimal feedback on their agreement levels during this stage.

4.2 Stage 2

In the second stage, annotations shifted to same-sex marriage articles, again emphasizing general fit across issues. Beginning in stage 2, each article

[WHERE THE JOBS ARE]Economic **[Critics of illegal immigration can make many cogent arguments to support the position that the U.S. Congress and the Colorado legislature must develop effective and well-enforced immigration policies that will restrict the number of people who migrate here legally and illegally.]**Policy prescription **[It's true that all forms of [immigration exert influence over our economic and cultural make-up.]**Cultural identity In some ways, immigration improves our economy by adding laborers, taxpayers and consumers, and in other ways immigration detracts from our economy by increasing the number of students, health care recipients and other beneficiaries of public services.]**Economic **[Some economists say that immigrants, legal and illegal, produce a net economic gain, while others say that they create a net loss]**Economic. There are rational arguments to support both sides of this debate, and it's useful and educational to hear the varying positions.

[WHERE THE JOBS ARE]Economic **[Critics of illegal immigration can make many cogent arguments to support the position that the U.S. Congress and the Colorado legislature must develop effective and well-enforced immigration policies that will restrict the number of people who migrate here legally and illegally.]**Public opinion **[It's true that all forms of immigration exert influence over our economic and [cultural make-up.]**Cultural identity In some ways, immigration improves our economy by adding laborers, taxpayers and consumers, and in other ways [immigration detracts from our economy by increasing the number of students, health care recipients and other beneficiaries of public services.]**Capacity **]**Economic **[Some economists say that immigrants, legal and illegal, produce a net economic gain, while others say that they create a net loss.]**Economic There are rational arguments to support both sides of this debate, and it's useful and educational to hear the varying positions.

Figure 2: Two independent annotations of a 2006 editorial in the *Denver Post*. The annotators agree perfectly about which parts of the article make use of economic framing, but disagree about the first paragraph. Moreover, the second annotator identifies an additional dimension (capacity and resources). Although they both identify a reference to cultural identity, they annotated slightly different spans of text.

was assigned to at least two annotators, in order to track inter-annotator agreement more carefully, and to better capture the subjectivity inherent in this task. Since the guidelines had become more stable by this stage, we also focused on identifying good practices for annotator training. Annotators were informed of their agreement levels with each other, and pairs of framing dimensions on which annotators frequently disagreed were emphasized. This information was presented to annotators in weekly meetings.

4.3 Stage 3

The third stage revisited the immigration articles from stage 1 (plus an additional group of articles), with the now well-developed annotation guidelines. As in the second stage, almost all articles were annotated by two annotators, working independently. More detailed feedback was provided, including inter-annotator agreement for the use of each framing dimension anywhere in articles.

During stage 3, for each article where two annotators independently disagreed on the primary frame, the pair met to discuss the disagreement and attempt to come to a consensus.[4] Disagreements continue to arise, however, reflecting the reality that the same article can cue different frames more strongly for different annotators. We view these disagreements not as a weakness, but as a source of useful information about the diversity of ways in which the same text can be interpreted by different audiences (Pan and Kosicki, 1993; Rees et al., 2001).

The proportion of articles annotated with each framing dimension (averaged across annotators) is shown in Figure 3.

5 Inter-annotator Agreement

Because our annotation task is complex (selecting potentially overlapping text spans and labeling them), there is no single comprehensive measure of inter-annotator agreement. The simplest aspect of the annotations to compare is the choice of primary frame, which we measure using Krippendorff's α (Krippendorff, 2012).[5]
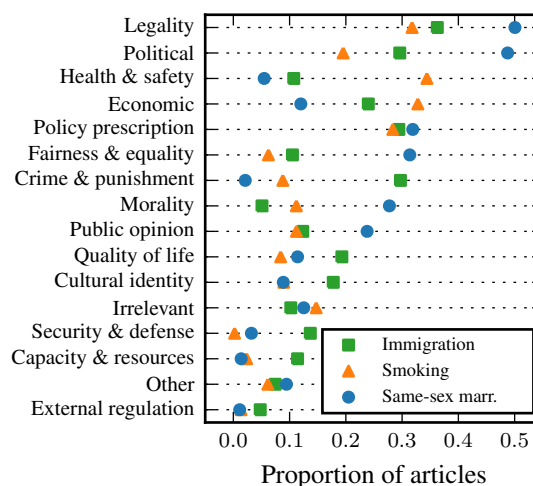

Figure 3: Proportion of articles annotated with each of the framing dimensions (averaging across annotators for each article).

Figure 4 shows the inter-annotator agreement on the primary frame per round. We observe first that difficulty varies by issue, with same-sex marriage the most difficult. Annotators do appear to improve with experience. Agreement on immigration articles in stage 3 are significantly higher ($p < 0.05$, permutation test) than agreement on the same articles in stage 1, even though only one annotator worked on both stages.[6]

These results demonstrate that consistent performance can be obtained from different groups of annotators, given sufficient training. Although we never obtain perfect agreement, this is not surprising, given that the same sentences can and do cue multiple types of framing, as illustrated by the example in Figure 2.

Inter-annotator agreement at the level of individually selected spans of text can be assessed using an extension of Krippendorff's α (α_U) which measures disagreement between two spans as the sum of the squares of the lengths of the parts which do not overlap.[7] As with the more common α statistic, α_U is a chance-corrected agreement metric scaled such that 1 represents perfect agreement and 0 represents the level of chance. This met-

---

[4]A small secondary experiment, described in supplementary material, was used to test the reliability of this process.

[5]Krippendorff's α is similar to Cohen's κ, but calculates expected agreement between annotators based on the combined pool of labels provided by all annotators, rather than considering each annotators's frequency of use separately. Moreover, it can be used for more than two annotators and

accommodates missing values. See Passonneau and Carpenter (2014) for additional details.

[6]Note that this is not a controlled experiment on annotation procedures, but rather a difference observed between two stages of an evolving process.

[7]For example, in the example shown in Figure 2, the amount of disagreement on the two Cultural identity annotations would be the square of the length (in characters) of the non-overlapping part of the annotations ("immigration exert influence over our economic and") which is $50^2 = 2500$.
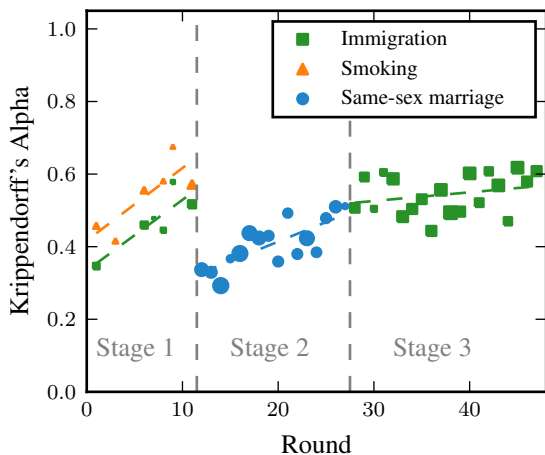
Figure 4: Chance-corrected inter-annotator agreement on the primary frame. Marker size indicates the number of annotations being compared; $\alpha = 1$ indicates perfect agreement.

ric has been previously recommended for tasks in computational linguistics that involve *unitizing* (Artstein and Poesio, 2008). For a more complete explanation, see Krippendorff (2004).

The pattern of α_U values across rounds is very similar to that shown in Figure 4, but not surprisingly, average levels of agreement are much lower. Arguably, this agreement statistic is overly harsh for our purposes. We do not necessarily expect annotators to agree perfectly about where to start and end each annotated span, or how many spans to annotate per article, and our codebook and guidelines offer relatively little guidance on these low-level decisions. Nevertheless, it is encouraging that in all cases, average agreement is greater than chance. The α_U values for all annotated spans of text (averaged across articles) are 0.16 for immigration (stage 1), 0.23 for tobacco, 0.08 for same-sex marriage, and 0.20 for immigration (stage 3).

6 Prior Work

Several previous papers in the computer science literature deal with framing, though usually in a more restricted sense. Perhaps the most common approach is to treat the computational analysis of framing as a variation on sentiment analysis, though this often involves reducing framing to a binary variable. Various models have been applied to news and social media datasets with the goal of identifying political ideology, or "perspective" (typically on a liberal to conservative scale) (Ahmed and Xing, 2010; Gentzkow and Shapiro, 2010; Lin et al., 2006; Hardisty et al., 2010; Kle-

banov et al., 2010; Sim et al., 2013; Iyyer et al., 2014), or "stance" (position for or against an issue) (Walker et al., 2012; Hasan and Ng, 2013). A related line of work is the analysis of subjective language or "scientific" language, which has also been posed in terms of framing (Wiebe et al., 2004; Choi et al., 2012). While the study of ideology, sentiment, and subjectivity are interesting in their own right, we believe that they fail to capture the more nuanced nature of framing, which is often more complex than positive or negative sentiment. In discussions of same-sex marriage, for example, both advocates and opponents may attempt to control whether the issue is perceived as primarily about politics, legality, or ethics. Moreover, we emphasize that framing is an important feature of even seemingly neutral or objective language.

A different but equally relevant line of work has focused on text re-use. Leskovec et al. (2009) perform clustering of quotations and their variations, uncovering patterns in the temporal dynamics of how memes spread through the media. On a smaller scale, others have examined text reuse in the development of legislation and the culture of reprinting in nineteenth-century newspapers (Smith et al., 2013; Smith et al., 2014). While not the same as framing, identifying this sort of text reuse is an important step towards analyzing the "media packages" that social scientists associate with framing.

7 Conclusion

Framing is a complex and difficult aspect of language to study, but as with so many aspects of modern NLP, there is great potential for progress through the use of statistical methods and public datasets, both labelled and unlabeled. By releasing the Media Frames Corpus, we seek to bring the phenomenon to the attention of the computational linguistics community, and provide a framework that others can use to analyze framing for additional policy issues. As technology progresses towards ever more nuanced understanding of natural language, it is important to analyze not just what is being said, but how, and with what effects. The Media Frames Corpus enables the next step in that direction.

References

Amr Ahmed and Eric P. Xing. 2010. Staying informed: Supervised and semi-supervised multi-view topical analysis of ideological perspective. In *Proc. of EMNLP*.

Ron Artstein and Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.

Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *Proc. of NAACL*.

Frank R. Baumgartner, Suzanna L. De Boef, and Amber E. Boydstun. 2008. *The decline of the death penalty and the discovery of innocence*. Cambridge University Press.

Robert D. Benford and David A. Snow. 2000. Framing processes and social movements: An overview and assessment. *Annual Review of Sociology*, 26:611–639.

Adam J. Berinsky and Donald R. Kinder. 2006. Making sense of issues through media frames: Understanding the Kosovo crisis. *Journal of Politics*, 68(3):640–656.

Amber E. Boydstun, Dallas Card, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2014. Tracking the development of media frames within and across policy issues. *APSA 2014 Annual Meeting Paper*.

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 of. ACL Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, pages 70–79.

Dennis Chong and James N. Druckman. 2007. Framing theory. *Annual Review of Political Science*, 10(1):103–126.

Robert M. Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4):51–58.

William A. Gameson and Andre Modigliani. 1989. Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology*, 95(1):1–37.

Matthew Gentzkow and Jesse M. Shapiro. 2010. What drives media slant? Evidence from U.S. daily newspapers. *Econometrica*, 78(1):35–71.

Eric Hardisty, Jordan L. Boyd-Graber, and Philip Resnik. 2010. Modeling perspective using adaptor grammars. In *Proc. of EMNLP*.

Kazi Saidul Hasan and Vincent Ng. 2013. Stance classification of ideological debates: Data, models, features, and constraints. In *Proc. of IJCNLP*.

Mohit Iyyer, Peter Enns, Jordan L. Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proc. of ACL*.

Beata Beigman Klebanov, Eyal Beigman, and Daniel Diermeier. 2010. Vocabulary choice as an indicator of perspective. In *Proc. of ACL*.

Klaus Krippendorff. 2004. Measuring the reliability of qualitative text analysis data. *Quality and Quantity*, 38(6):787–800.

Klaus Krippendorff. 2012. *Content Analysis: An Introduction to its Methodology*. SAGE Publications.

Jure Leskovec, Lars Backstrom, and Jon Kleinberg. 2009. Meme-tracking and the dynamics of the news cycle. In *Proc. of KDD*.

Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, and Alexander Hauptmann. 2006. Which side are you on? Identifying perspectives at the document and sentence levels. In *Proc. of CoNNL*.

Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, and Kristina Miler. 2015. Tea party in the house: A hierarchical ideal point topic model and its application to Republican legislators in the 112th congress. In *Proc. of ACL*.

Zhongdang Pan and Gerald M. Kosicki. 1993. Framing analysis: An approach to news discourse. *Political communication*, 10(1):55–75.

Rebecca J. Passonneau and Bob Carpenter. 2014. The benefits of a model of annotation. In *Proc. of ACL*.

Stephen D. Rees, Oscar H. Gandy Jr., , and August E. Grant, editors. 2001. *Framing public life: Perspectives on media and our understanding of the social world*. Routledge.

Yanchuan Sim, Brice D. L. Acree, Justin H. Gross, and Noah A. Smith. 2013. Measuring ideological proportions in political speeches. In *Proc. of EMNLP*.

David A. Smith, Ryan Cordell, and Elizabeth Maddock Dillon. 2013. Infectious texts: modeling text reuse in nineteenth-century newspapers. In *IEEE International Conference on Big Data*.

David A. Smith, Ryan Cordel, Elizabeth Maddock Dillon, Nick Stramp, and John Wilkerson. 2014. Detecting and modeling local text reuse. In *IEEE/ACM Joint Conference on Digital Libraries*.

Nayda Terkildsen and Frauke Schnell. 1997. How media frames move public opinion: An analysis of the women's movement. *Political research quarterly*, 50(4):879–900.

Oren Tsur, Dan Calacci, and David Lazer. 2015. Frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proc. of ACL*.

Marilyn A. Walker, Pranav Anand, Robert Abbott, and Ricky Grant. 2012. Stance classification using dialogic properties of persuasion. In *Proc. of NAACL*.

Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Computational Linguistics*, 30(3):277–308.