Author/s:
Khanehzar, S;Cohn, T;Mikolajczak, G;Turpin, A;Frermann, L

Title:
Framing Unpacked: A Semi-Supervised Interpretable Multi-View Model of Media Frames

Date:
2021-01-01

Citation:
Khanehzar, S., Cohn, T., Mikolajczak, G., Turpin, A. & Frermann, L. (2021). Framing Unpacked: A Semi-Supervised Interpretable Multi-View Model of Media Frames. Naacl Hlt 2021 2021 Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Proceedings of the Conference, pp.2154-2166. ASSOC COMPUTATIONAL LINGUISTICS-ACL. https://doi.org/10.18653/v1/2021.naacl-main.174.

Persistent Link:
https://hdl.handle.net/11343/281261

Framing Unpacked: A Semi-Supervised Interpretable Multi-View Model of Media Frames

Shima Khanehzar[†] **Trevor Cohn**[†] **Gosia Mikolajczak**[⋆] **Andrew Turpin**[†] **Lea Frermann**[†]
[†] School of Computing and Information Systems
[⋆]School of Social and Political Sciences
The University of Melbourne
skhanehzar@student.unimelb.edu.au
{tcohn,mmikolajcza,aturpin,lfrermann}@unimelb.edu.au

Abstract

Understanding how news media frame political issues is important due to its impact on public attitudes, yet hard to automate. Computational approaches have largely focused on classifying the frame of a full news article while framing signals are often subtle and local. Furthermore, automatic news analysis is a sensitive domain, and existing classifiers lack transparency in their predictions. This paper addresses both issues with a novel semi-supervised model, which jointly learns to embed local information about the events and related actors in a news article through an auto-encoding framework, and to leverage this signal for document-level frame classification. Our experiments show that: our model outperforms previous models of frame prediction; we can further improve performance with unlabeled training data leveraging the semi-supervised nature of our model; and the learnt event and actor embeddings intuitively corroborate the document-level predictions, providing a nuanced and interpretable article frame representation.

1 Introduction

Journalists often aim to package complex real-world events into comprehensive narratives, following a logical sequence of events involving a limited set of actors. Constrained by word limits, they necessarily select some facts over others, and make certain perspectives more salient. This phenomenon of *framing*, be it purposeful or unconscious, has been thoroughly studied in the social and political sciences (Chong and Druckman, 2007). More recently, the natural language processing community has taken an interest in automatically predicting the frames of news articles (Card et al., 2016; Field et al., 2018; Akyürek et al., 2020; Khanehzar et al., 2019; Liu et al., 2019a; Huguet Cabot et al., 2020).

Definitions of framing vary widely including: expressing the same semantics in different forms (equivalence framing); presenting selective facts and aspects (emphasis framing); and using established syntactic and narrative structures to convey information (story framing) (Hallahan, 1999). The model presented in this work builds on the concepts of emphasis framing and story framing, predicting the global (aka. primary) frame of a news article on the basis of the events and participants it features.

Primary frame prediction has attracted substantial interest recently with the most accurate models being supervised classifiers built on top of large pre-trained language models (Khanehzar et al., 2019; Huguet Cabot et al., 2020). This work advances prior work in two ways. First, we explicitly incorporate a formalization of *story framing* into our frame prediction models. By explicitly modeling news stories as latent representations over events and related actors, we obtain interpretable, latent representations lending transparency to our frame prediction models. We argue that transparent machine learning is imperative in a potentially sensitive domain like automatic news analysis, and show that the local, latent labels inferred by our model lend explanatory power to its frame predictions.

Secondly, the latent representations are induced without frame-level supervision, requiring only a pre-trained, off-the-shelf semantic role labeling (SRL) model (Shi and Lin, 2019). This renders our frame prediction models semi-supervised, allowing us to use large unlabeled news corpora.

More technically, we adopt a dictionary learning framework with deep autoencoders through which we learn to map events and their agents and patients[1] independently into their respective structured latent space. Our model thus learns a latent multi-view representation of news stories, with each view contributing evidence to the primary

[1]We experiment with three types of semantic roles: predicates and associated arguments (ARG0 and ARG1), however, our framework is agnostic to the types of semantic roles, and can further incorporate other types of semantic roles or labels.

Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 2154–2166
June 6–11, 2021.

frame prediction from its own perspective. We incorporate the latent multi-view representation into a transformer-based document-level frame classification model to form a semi-supervised model, in which the latent representations are jointly learnt with the classifier.

We demonstrate empirically that our semi-supervised model outperforms current state-of-the-art models in frame prediction. More importantly, through detailed qualitative analysis, we show how our latent features mapped to events and related actors allow for a nuanced analysis and add interpretability to the model predictions[2]. In summary, our contributions are:

- Based on the concepts of story- and emphasis framing, we develop a novel semi-supervised framework which incorporates local information about core events and actors in news articles into a frame classification model.
- We empirically show that our model, which incorporates the latent multi-view semantic role representations, outperforms existing frame classification models, with only labeled articles. By harnessing large sets of unlabeled in-domain data, our model can further improve its performance and achieves new state-of-the-art performance on the frame prediction task.
- Through qualitative analysis, we demonstrate that the latent, multi-view representations aid interpretability of the predicted frames.

2 Background and Related Work

A widely accepted definition of frames describes them as a selection of aspects of perceived reality, which are made salient in a communicating context to promote a particular problem definition, causal interpretation, moral evaluation and the treatment recommendation for the described issue (Entman, 1993). While detecting media frames has attracted much attention and spawned a variety of methods, it poses several challenges for automatic prediction due to its vagueness and complexity.

Two common approaches in the study of frames focus either on the detailed issue-specific elements of a frame or, somewhat less nuanced, on generic framing themes prevalent across issues. Within the first approach, Matthes and Kohring (2008) developed a manual coding scheme, relying on Entman's definition (Entman, 1993). While the scheme assumes that each frame is composed of common elements, categories within those elements are often specific to the particular issue being discussed (e.g., "same sex marriage" or "gun control"), making comparison across different issues, and detecting them automatically difficult. Similarly, earlier studies focusing specifically on unsupervised models to extract frames, usually employed topic modeling (Boydstun et al., 2013; Nguyen, 2015; Tsur et al., 2015) to find the issue-specific frames, limiting across-issue comparisons.

Studies employing generic frames address this shortcoming by proposing common categories applicable to different issues. For example, Boydstun et al. (2013) proposed a list of 15 broad frame categories commonly used when discussing different policy issues, and in different communication contexts. The Media Frames Corpus (MFC; Card et al. (2015)) includes about 12,000 news articles from 13 U.S. newspapers covering five different policy issues, annotated with the dominant frame from Boydstun et al. (2013). Table 5 in the Appendix lists all 15 frame types present in the MFC. The MFC has been previously used for training and testing frame classification models. Card et al. (2016) provide an unsupervised model that clusters articles with similar collections of "personas" (i.e., characterisations of entities) and demonstrate that these personas can help predict the coarse-grained frames annotated in the MFC. While conceptually related to our approach, their work adopts the Bayesian modelling paradigm, and does not leverage the power of deep learning. Ji and Smith (2017) proposed a supervised neural approach incorporating discourse structure. The current best result for predicting the dominant frame of each article in the MFC comes from Khanehzar et al. (2019), who investigated the effectiveness of a variety of pre-trained language models (XLNet, Bert and Roberta).

Recent methods have been expanded to multilingual frame detection. Field et al. (2018) used the MFC to investigate framing in Russian news. They introduced embedding-based methods for projecting frames of one language into another (i.e., English to Russian). Akyürek et al. (2020) studied multilingual transfer learning to detect multiple frames in target languages with few or no annotations. Recently, Huguet Cabot et al. (2020) investigated joint models incorporating metaphor,

[2]Source code of our model is available at https://github.com/shinyemimalef/FRISS

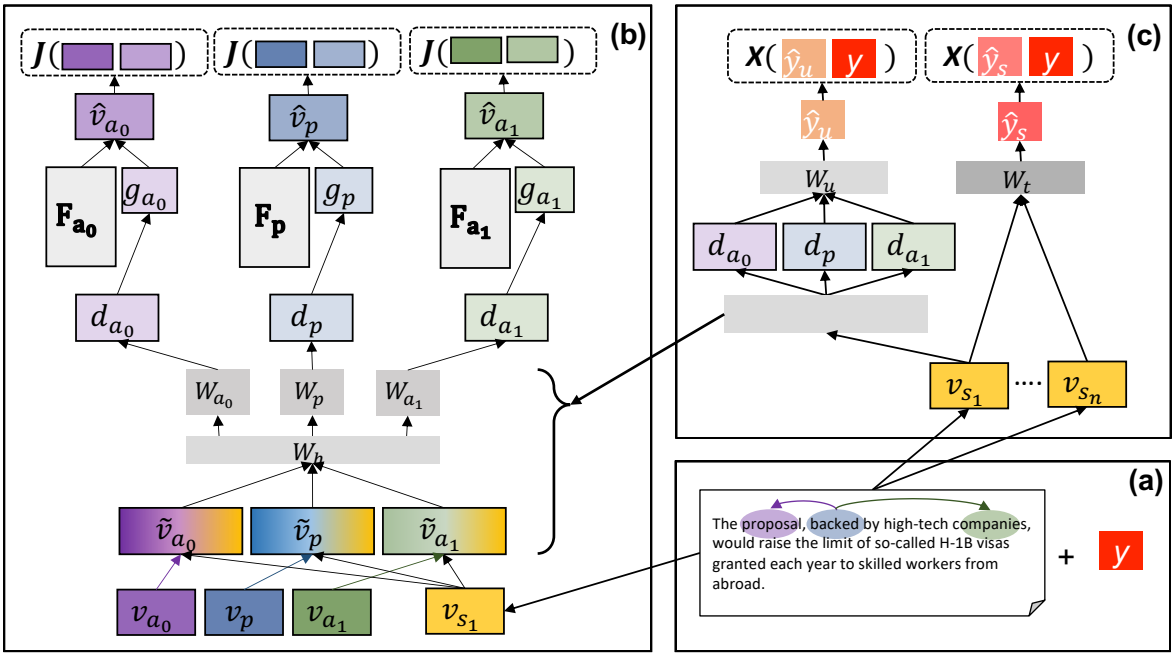

Figure 1: Our FRISS model. **(a)** Given an input document with a frame label y , we perform semantic role labeling (SRL) to obtain the predicate (blue), ARG0 (purple) and ARG1 (green) for input sentences $s_1 \dots s_n$. **(b)** The unsupervised module takes as input semantic role embeddings (v_{a_0}, v_p, v_{a_1}) and a sentence embedding (v_{s_i}), and learns latent, role-specific embedding matrices (**F**) in an auto-encoding framework. The latent representations are incorporated into the overall frame classification module. **(c)** We predict document-level frames based on transformer-based document embeddings ($\hat{y}_s$) and the view-specific latent representations ($\hat{y}_u$), using cross-entropy loss with the true frame label y.

emotion and political rhetoric within multi-task learning to predict framing of policy issues.

Our modelling approach is inspired by recent advances in learning interpretable latent representations of the participants and relationships in fiction stories. Iyyer et al. (2016) present Relationship Modelling Networks (RMNs), which induce latent descriptors of types of *relationships* between characters in fiction stories, in an unsupervised way. RMNs combine dictionary learning with deep autoencoders, and are trained to effectively encode text passages as linear combinations over latent descriptors, each of which corresponds to a distinct relationship (not unlike topics in a topic model). Frermann and Szarvas (2017) extend the idea to a multi-view setup, jointly learning multiple dictionaries, which capture *properties* of individual characters in addition to relationships. We adopt this methodology for modeling news articles through three latent views: capturing their events (predicates), and participants (ARG0, ARG1). We combine the unsupervised autoencoder with a frame classifier into an interpretable, semi-supervised framework for article-level frame prediction.

3 Semi-supervised Interpretable Frame Classification

In this section, we present our `Frame` classifier, which is `I`nterpretable and `S`emi-`s`upervised (FRISS). The full model is visualized in Figure 1. Given a corpus of news articles, some of which have a label indicating their primary frame y (Figure 1(a)), FRISS learns to predict $\hat{y}$ for each document by combining a supervised classification module (Figure 1(c)) and an unsupervised[3] autoencoding module (Figure 1(b)), which are jointly trained. The unsupervised module (i) can be trained with additional unlabeled training data, which improves performance (Section 5.2); and (ii) learns interpretable latent representations which improve the interpretability of the model (Section 5.3).

Intuitively, FRISS predicts frames based on aggregated sentence representation (supervised module; Section 3.2) as well as aggregated fine-grained latent representations capturing actors and events in the article (unsupervised module; 3.1). The un-

[3]The autoencoder is unsupervised wrt. frame-level information, but it relies on an off-the-shelf semantic role labeler.

supervised module combines an auto-encoding objective with a multi-view dictionary learning framework (Iyyer et al., 2016; Frermann and Szarvas, 2017). We treat predicates, their ARG0 and ARG1 as three separate views, and learn to map each view to an individual latent space representative of their relation to the overall framing objective. Below, we will sometimes refer to views collectively as $z \in \{p, \mathrm{a}_0, \mathrm{a}_1\}$. We finally aggregate the view-level representations and sentence representations to predict a document-level frame. The following sections describe FRISS in technical detail.

3.1 Unsupervised Module

3.1.1 Input

Each input document is sentence-segmented and automatically annotated by an off-the shelf transformer-based semantic role labeling model (Shi and Lin, 2019; Pradhan et al., 2013) to indicate spans over the three semantic roles: predicates, ARG0s and ARG1s.

We compute a contextualized vector representation for each semantic role span (s_p, s_{a_0}, s_{a_1}). We describe the process for obtaining predicate input representations $\boldsymbol{v}_p$ here for illustration. Contextualized representations for views a_0 ($\boldsymbol{v}_{\mathrm{a}_0}$) and a_1 ($\boldsymbol{v}_{\mathrm{a}_1}$) are obtained analogously. First, we pass each sentence through a sentence encoder, and obtain the predicate embedding by averaging all contextualized token representations $\boldsymbol{v}_w$ (of dimension D_w) in its span s_p of length $|s_p|$:

$$\boldsymbol{v}_\mathrm{p} = \frac{1}{|s_p|} \sum_{w \in s_p} \boldsymbol{v}_w. \tag{1}$$

We concatenate $\boldsymbol{v}_\mathrm{p}$ with an overall sentence representation $\boldsymbol{v}_s$, which is computed by averaging all contextualized token embeddings of the sentence s of length $|s|$,[4]

$$\boldsymbol{v}_s = \frac{1}{|s|} \sum_{w \in s} \boldsymbol{v}_w \tag{2}$$

$$\tilde{\boldsymbol{v}}_\mathrm{p} = [\boldsymbol{v}_\mathrm{p}; \boldsymbol{v}_s], \tag{3}$$

where [;] denotes vector concatenation. If a sentence has more than one predicate, a separate representation is computed for each of them.

3.1.2 Multi-view Frame representations

We combine ideas from auto-encoding (AE) and dictionary learning, as previously used to capture the content of fictitious stories (Iyyer et al., 2016), and its multi-view extension (Frermann and Szarvas, 2017). We posit a latent space as three view-specific dictionaries (Figure 1 (b)) capturing events (predicates; $\boldsymbol{F}_p$), their first (ARG0; $\boldsymbol{F}_{\mathrm{a}_0}$) and second (ARG1; $\boldsymbol{F}_{\mathrm{a}_1}$) arguments, respectively. Given a view-specific input as described above, the autoencoder maps it to a low-dimensional distribution over "dictionary terms" (henceforth descriptors), which are learnt during training. The descriptors are vector-valued latent variables that live in word embedding space, and are hence interpretable through their nearest neighbors (Table 3 shows examples of descriptors inferred by our model). By jointly learning the descriptors with the supervised classification objective, each descriptor will capture coherent information corresponding to a frame label in our supervised data set. We hence set the number of descriptors for each dictionary to $K = 15$, the number of frames in our data set. For each view $z \in \{p, \mathrm{a}_0, \mathrm{a}_1\}$, we define a dictionary $\boldsymbol{F}_z$ of dimensions $K \times D_w$.

More technically, our model follows two steps. First, we *encode* the input $\tilde{\boldsymbol{v}}_z$ of a known view z by passing it through a feed forward layer $\boldsymbol{W}_h$ of dimensions $2D_w \times D_h$, shared across all the views, followed by a ReLU non-linearity, and then another feed forward layer $\boldsymbol{W}_z$ of dimensions $D_h \times K$, specific to each view z. This results in a K-dimensional vector over the view-specific descriptors,

$$\boldsymbol{l}_z = \mathbf{W}_z \, \mathrm{ReLU}(\boldsymbol{W}_h \tilde{\boldsymbol{v}}_z), \tag{4}$$

Second, we reconstruct the original view embedding $\boldsymbol{v}_z$ as a linear combination of descriptors. While previous work used $\boldsymbol{l}_z$ directly as weight vector, we hypothesize that on our fine-grained semantic role level, only one or a few descriptors will be relevant to any specific span. We enforce this intuition using Gumbel-Softmax differentiable sampling with temperature annealing (Jang et al., 2017). This allows us to gradually constrain the number of relevant descriptors used for reconstruction. We first normalize $\boldsymbol{l}_z$,

$$\boldsymbol{d}_z = \mathrm{Softmax}(\boldsymbol{l}_z), \tag{5}$$

and then draw $\boldsymbol{g}$ from the Gumbel distribution, and add it to our normalized logits $\boldsymbol{d_z}$ scaled by temperature τ , which is gradually annealed over the

[4] We also experimented with representing the sentence as the $[CLS]$ token embedding, but found it to perform worse empirically.

training phase:

$$\boldsymbol{g} \sim \text{Gumbel}(0, 1)$$
$$\boldsymbol{g}_z = \frac{\exp(\log(\boldsymbol{d}_z) + \frac{\boldsymbol{g}}{\tau})}{\sum_f \exp(\log(\boldsymbol{d}_z) + \frac{\boldsymbol{g}}{\tau})}. \quad (6)$$

We finally reconstruct the view-specific span embedding as

$$\hat{\boldsymbol{v}}_z = \boldsymbol{F}_z^T \boldsymbol{g}_z. \quad (7)$$

3.1.3 Unsupervised Objective

Contrastive Loss We use the contrastive max-margin objective function following previous works in dictionary learning (Iyyer et al., 2016; Frermann and Szarvas, 2017; Han et al., 2019). We randomly sample a set of negative samples (N_-) with the same view as the current input from the mini-batch. The unregularized objective J_z^u (Eq. 8) is a hinge loss that minimizes the L2 norm[5] between the reconstructed embedding $\hat{\boldsymbol{v}}_z$ and the true input's view-specific embedding $\boldsymbol{v}_z$, while simultaneously maximizing the L2 norm between $\hat{\boldsymbol{v}}_z$ and negative samples $\boldsymbol{v}_z^n$:

$$J_z^u(\theta) = \frac{1}{|N_-|} \sum_{\boldsymbol{v}_z^n \in N_-} \max(0, 1 + l_2(\hat{\boldsymbol{v}}_z, \boldsymbol{v}_z) - l_2(\hat{\boldsymbol{v}}_z, \boldsymbol{v}_z^n)), \quad (8)$$

where θ represents the model parameters, $|N_-|$ is the number of negative samples, and the margin value is set to 1.

Focal Triplet Loss Preliminary studies (Section 5) suggested that some descriptors (aka frames) are more similar to each other than others. We incorporate this intuition through a novel mechanism to move the descriptors that are least involved in the reconstruction proportionally further away from the most involved descriptor.

Concretely, we select t descriptors in $\mathbf{F}_z$ with smallest weights in $\boldsymbol{g}_z$ as additional negative samples. We denote the indices of the selected t smallest components in $\boldsymbol{g}_z$

$$I = [i_1, i_2, \ldots, i_t]. \quad (9)$$

We use $\boldsymbol{F}_z^t$ to denote the matrix $(t \times D_w)$ with only those t descriptors. We re-normalize the weights of the selected t descriptors, and denote the renormalized weights vector as $\boldsymbol{g}_z^t = [g_z^{i_1}, g_z^{i_2}, \ldots, g_z^{i_t}]$. For each element in $\boldsymbol{g}_z^t$, we compute an individual margin based on its magnitude. Intuitively, the smaller the weight is, the larger its required margin from a given total margin budget $|M|$,

$$m_z^{i_t} = |M| * (1 - g_z^{i_t})^2. \quad (10)$$

We compute the standard margin-based hinge loss over the additional negative samples with sample-specific margins:

$$J_z^t(\theta) = \frac{1}{|T|} \sum_{i_t \in I} \max(0, m_z^{i_t} + l_2(\hat{\boldsymbol{v}}_z, \boldsymbol{v}_z) - l_2(\hat{\boldsymbol{v}}_z, \boldsymbol{v}_z^{i_t})). \quad (11)$$

We sum the focal triplet objective J_z^t with J_z^u, and then sum over all specific spans $s \in S_z$, while adding an additional orthogonality encouraging regularization term.

$$J_z(\theta) = \sum_{s \in S_z} (J_z^u + J_z^t) + \lambda ||\mathbf{F_z} \mathbf{F_z^T} - \mathbf{I}||_{\mathbf{F}}^{\mathbf{2}}, \quad (12)$$

where λ is a hyper-parameter that can be tuned. We finally aggregate the loss from all the views:

$$J(\theta) = \sum_{z \in \{\text{p}, \text{a}_0, \text{a}_1\}} J_z(\theta). \quad (13)$$

3.2 Supervised Document-level Frame Classification

We incorporate the semantic role level predictions as described above into a document-level frame classifier consisting of two parts, which are jointly learnt with the unsupervised model described above: (i) a classifier based on aggregated span-level representations computed as described in Sec 3.1 (Fig. 1 (c; left); Sec. 3.2.1) and (ii) a classifier based on an aggregated sentence representations (Fig. 1 (c; right); Sec. 3.2.2).

3.2.1 Span-based Classifier

The unsupervised module makes predictions on the semantic role span level, however, our goal is to predict document-level frame labels. We aggregate span-level representations $\boldsymbol{d}_z$ (Eq. 5) by averaging across spans and then views:[6]

$$\boldsymbol{w}_u = \frac{1}{Z} \sum_{z \in \{\text{p}, \text{a}_0, \text{a}_1\}} \frac{1}{|S_z|} \sum_{s \in S_z} \boldsymbol{d}_z^s$$
$$\hat{y}_u = \text{Softmax}(\boldsymbol{w}_u), \quad (14)$$

[5] We empirically found that L2 norm outperforms the dot product, and cosine similarity.

[6] We empirically found these representations to outperform the sparser $\boldsymbol{g}_z$.

where Z is the number of the views, and S_z are the set of view-specific spans in the current document. We finally pass the logits through a softmax layer to predict a distribution over frames.

3.2.2 Sentence-based Classifier

We separately predict a document-level frame based on the aggregate sentence level representations computed in Eq. (2). We first pass each sentence embedding through a feed forward layer W_r of dimensions $D_w \times D_w$, followed by a ReLU non-linearity, and another feed forward layer W_t to map the resulting representation to K dimensions. Then average across sentences of the current document S_d and pass the result through a softmax layer,

$$\begin{aligned} \boldsymbol{w}_s &= \text{ReLU}(\boldsymbol{W}_r \boldsymbol{v}_s) \\ \hat{y}_s &= \text{Softmax}\left(\frac{1}{|S_d|}\sum_{s \in S_d} \boldsymbol{W}_t \boldsymbol{w}_s\right). \end{aligned} \tag{15}$$

3.3 Full Loss

We jointly train the supervised and unsupervised model components. The supervised loss $X(\theta)$ consists of two parts, one for the sentence-based classification and one for the aggregated span-based classification:

$$X(\theta) = X(\hat{y}_u, y) + X(\hat{y}_s, y). \tag{16}$$

The full loss balances the supervised and unsupervised components with a hyper-parameter α:

$$L(\theta) = \alpha \times X(\theta) + (1 - \alpha) \times J(\theta). \tag{17}$$

4 Experimental Settings

Dataset We follow prior work on automatic prediction of a single, primary frame of a news article as annotated in the Media Frames Corpus (MFC; Card et al. (2015)). The MFC contains a large number of news articles on five contentious policy issues (immigration, smoking, gun control, death penalty, and same-sex marriage), manually annotated with document- and span-level frames labels from a set of 15 general frames (listed in Table 5 in the Appendix). Articles were selected from 13 major U.S. newspapers, published between 1980 and 2012. Following previous work, we focus on the immigration portion of MFC, which comprises 5,933 annotated articles, as well as an additional 41,286 unlabeled articles. The resulting dataset contains all 15 frames. Table 5 (Appendix) lists the corresponding frame distribution. We partition the labeled dataset into 10 folds, preserving the overall frame distribution for each fold.

Pre-processing and Semantic Role labeling We apply state-of-art BERT-based SRL model (Shi and Lin, 2019) to obtain SRL spans for each sentence. The off-the-shelf model from AllenNLP is trained on OntoNotes5.0 (close to 50% news text). While a domain-adapted model may lead to a small performance gain, the off-the-shelf model enhances generalizability and reproducibility. Qualitative examples of detected SRL spans are shown in Table 4, which confirm that SRL predictions are overall accurate.

We extract semantic role spans for predicates, their associated first (ARG0) and second (ARG1) arguments for each sentence in a document. For the unsupervised component, we disregard sentences with no predicate, and sentences missing both ARG0 and ARG1.

Sentence Encoder In all our experiments, we use RoBERTa (Liu et al., 2019b) as our sentence encoder, as previous work (Khanehzar et al., 2019) has shown that it outperforms BERT (Devlin et al., 2019) and XLNet (Yang et al., 2019). We pass each sentence through RoBERTa and retrieve the token-level embeddings. To obtain the sentence embedding, we average the RoBERTa embeddings of all words (Eq. 2). To obtain SRL span embeddings, we average the token embeddings of all words in a predicted span (Eq. 1). Following Gururangan et al. (2020), we pre-train RoBERTa with immigration articles using the masked language model (MLM) objective. Only the labeled data is used for pre-training for fair comparison between FRISS and previous models.

Parameter Settings We set the maximum sequence length to RoBERTa 64 tokens, the maximum number of sentences per document to 32, and the maximum number of predicates per sentence to 10.[7] We set the number of dictionary terms $K = 15$, i.e., the number of frame classes in the MFC corpus. Each dictionary term is of dimension $D_w = 768$, equal to the RoBERTa token embedding dimension. We also fix the dimensions of hidden vector $\boldsymbol{w}_s$ (Eqn. 15) and D_h to this value. We set the number of descriptors in Focal Triplet Loss

[7] 96% of the sentences are under 64 tokens; 95% of the documents have less than 32 sentences, and $\gg$ 99% of the sentences have less than 10 predicates.

Model	Acc.	Macro-F1
Card et al. (2016)	56.8	-
Field et al. (2018)	57.3	-
Ji and Smith (2017)	58.4	-
Khanehzar et al. (2019)	65.8	-
RoBERTa-S	66.4 (0.008)	58.1 (0.013)
RoBERTa-S+MLM	67.1 (0.008)	58.8 (0.013)
FRISS (labeled only)	**68.8** (0.009)	**60.1** (0.011)
FRISS (labeled+unlabeled)	**69.7** (0.011)	**60.5** (0.015)

Table 1: Primary Frame prediction on the MFC immigration data set with mean (standard deviation) over 10-fold cross-validation. We compare our full model FRISS against recent work, and the supervised RoBERTa component with and without pre-training. The results are statistically significant ($p < 0.05$; paired sample t-test). FRISS (labeled only) vs RoBERTa-S+MLM: p=0.009; FRISS (labeled + all unlabeled) vs FRISS (labeled only): p=0.012; FRISS (labeled + all unlabeled) vs RoBERTa-S+MLM: p=0.003.

Model	Acc.	Macro-F1
FRISS	**68.83**	**60.05**
- focal	68.73	59.98
- Gumbel	68.51	59.70
- focal, Gumbel	68.47	59.72
p only	68.35	59.50
ARG0 only	68.47	59.68
ARG1 only	68.53	59.71

Table 2: Ablation results for FRISS on primary frame prediction. Top removing Focal Triplet Loss (focal; 3.1.3), and/or Gumbel distribution (Gumbel; 3.1.2). Bottom FRISS trained with only a single view (predicate, ARG0 and ARG1) on primary frame prediction.

$t = 8$ and the margin pool $|M| = t$. We set the balancing hyper-parameter between the supervised and unsupervised loss $\alpha = 0.5$, and $\lambda = 10^{-3}$. The dropout rate is set to 0.3.

We perform stochastic gradient descent with mini-batches of 8 documents. We use the Adam optimizer (Kingma and Ba, 2015) with the default parameters, except for the learning rate, which we set to 2×10^{-5} (for the RoBERTa parameters) and 5×10^{-4} (for all other parameters). We use a linear scheduler for learning rate decay. The weight decay is applied to all parameters except for bias and batch normalization. We update the Gumbel softmax temperature with the schedule: $\tau = \max(0.5, \exp(-5 \times 10^{-4} \times \text{iteration})$, updating the temperature every 50 iterations. For all our experiments, we run a maximum of 10 epochs, evaluate every 50 iterations, and apply early-stopping if the accuracy does not improve for 20 consecutive evaluations.

5 Evaluation

In this section, we evaluate the performance of FRISS on primary frame prediction for issue-specific news articles against prior work (Sec 5.1), demonstrate the benefit of adding additional unlabeled data to our semi-supervised model (Sec 5.2), and present a qualitative analysis of our model output corroborating its interpretability (Sec 5.3).

Preliminary Studies Preliminary analyses revealed that human annotators disagree on frame labels non-uniformly, suggesting that some pairs of frames are perceived to be more similar than others. This observation motivated the Focal Triplet Loss and Gumbel regularization components of our model. In particular, the following groups of frame labels are confused most frequently {"Policy Prescription and Evaluation", "Public Sentiment", "Political"}, {"Fairness", "Legality"}, {"Crime and Punishment", "Security and Defense"}, and {"Morality", "Quality of Life", "Cultural Identity"}. This ovservation is also corroborated through the empirical gain through the focal triplet loss (Table 2).

5.1 Experiment 1: Frame Prediction

For the supervised model, we report accuracy, as has been done in previous work, as well as Macro-F1, which is oblivious to class sizes, shedding light on performance across all frames. Table 1 compares FRISS against related work. Card et al. (2016) incorporate latent personas learnt with a Bayesian model; Field et al. (2018) derive frame-specific lexicons based on pointwise-mutual information; Ji and Smith (2017) incorporate a supervised discourse classifier, and Khanehzar et al. (2019) train frame classifiers on top of RoBERTa-based document embeddings. RoBERTa-S corresponds to the sentence-embedding based component of FRISS (Fig 1(b); left) without and with (+MLM) unsupervised pre-training. Overall, we can see that all our model variants outperform previous work in terms of both accuracy and macro-F1. Experiments were run 5 times with 10-fold cross-validation. The results in Table 1 are statistically significant ($p<0.05$; paired sample t-test).

To better understand the impact of our various

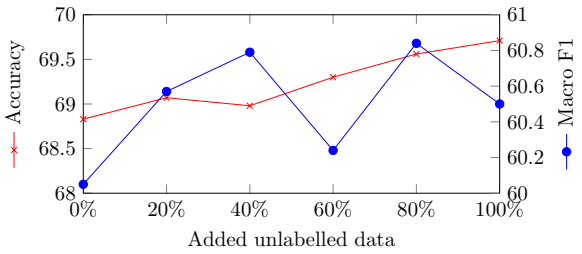

Figure 2: FRISS frame prediction performance with different portions of the 41K unlabeled documents.

ARG0	USCIS, state department, agency, federal official
	Trump, house republican, Obama, democrat, senate
	supreme court, justice, federal judge, court
	organizer, activist, protester, demonstrator, marcher
PRED	process, handle, swamp, accommodate, wait, exceed
	veto, defeat, vote, win, introduce, endorse, elect
	sue, uphold, entitle, appeal, shall, violate, file
	chant, march, protest, rally, wave, gather, organize
ARG1	application, foreign worker, visa, applicant
	amendment, reform, legislation, voter, senate bill
	political asylum, asylum, lawsuit, suit, status, case
	rally, marcher, march, protest, movement, crowd

Table 3: Spans inferred as most highly associated with the Capacity & Resources (■), Political (■), Legality (■), and Public Sentiment (■) frames, for each view (ARG0, PRED, ARG1).

model components, we performed an ablation study on Focal Triplet Loss, the Gumbel regularization, and the impact of individual views. Table 2 shows that both the focal loss and the Gumbel regularization contribute to model performance. Training FRISS with any single view individually leads to a performance drop, which is most drastic if the two arguments are omitted, suggesting that the model relies on both predicate and argument information, with arguments playing a slightly more important role.

5.2 Experiment 2: Benefit of Unlabeled Data

Our semi-supervised model can leverage news articles without a frame label, in addition to a labeled training set. We investigated the impact of training FRISS with different amounts of additional news articles, taken from the unlabeled immigration portion of the MFC. Figure 2 shows the impact of additional unlabeled data on accuracy and F1: Models with access to more unlabelled data tend to result in higher accuracy and Macro F1 scores. Given the abundance of online news articles, this motivates future work on minimally supervised frame prediction, minimizing the reliance on manual labels and maximizing generalizability to new issues, news outlets or languages.

5.3 Experiment 3: Qualitative Evaluation

In this experiment, we explore the added interpretability contributed by the local latent frame representations. Table 4 contains two MFC documents, highlighted with the most highly associated frame for each identified span for p, a_0 or a_1. We can observe that the frame associations (a) are intuitively meaningful; and (b) provide a detailed account of the predicted primary frame. For both documents the gold primary frame is 'Political', the bottom document is classified correctly, whereas the top document is mis-classified as 'Capacity & Resources'. The detailed span-level predictions help to explain the model prediction, and in fact add support for the the mis-prediction, suggesting that predicting a single primary document frame may be inappropriate. In the bottom document "a letter" serves as both a_1 of "Republicans sent a letter", where it is predicted as 'Political', and as a_0 of the clause "a letter [...] describing the legal challenges", where it is classified as 'Legality', another example of the nuance of our model predictions, which can support further in-depth study of issue-specific framing.

The potential of our model for fine-grained frame analysis is illustrated in Table 4, which shows how each particular SRL span contributes differently towards various frame categories. It adds a finer-grained framing picture, and estimate of the trustworthiness of model predictions. It allows to assess the main actors wrt. a particular frame (within and across articles), as well as the secondary frames in each article. Also, using SRL makes our model independent of human annotation, and more generalizable. Going beyond "highlighting indicative phrases", our model can distinguish their roles (e.g., the "ICE" as an actor vs. participant in a particular frame).

Table 3 shows the semantic role spans, which are most closely related to Capacity & Resources (blue), Political (red), Legality (purple) and Public Sentiment (green) descriptors in the latent space. We can observe that all associated spans are intuitively relevant to the {frame, view}. Furthermore, ARG0 spans tend to correspond to active participants (agents) in the policy process (including politicians and government bodies), whereas ARG1 spans illustrate the affected participants (patients such as foreign workers, applicants), pro-

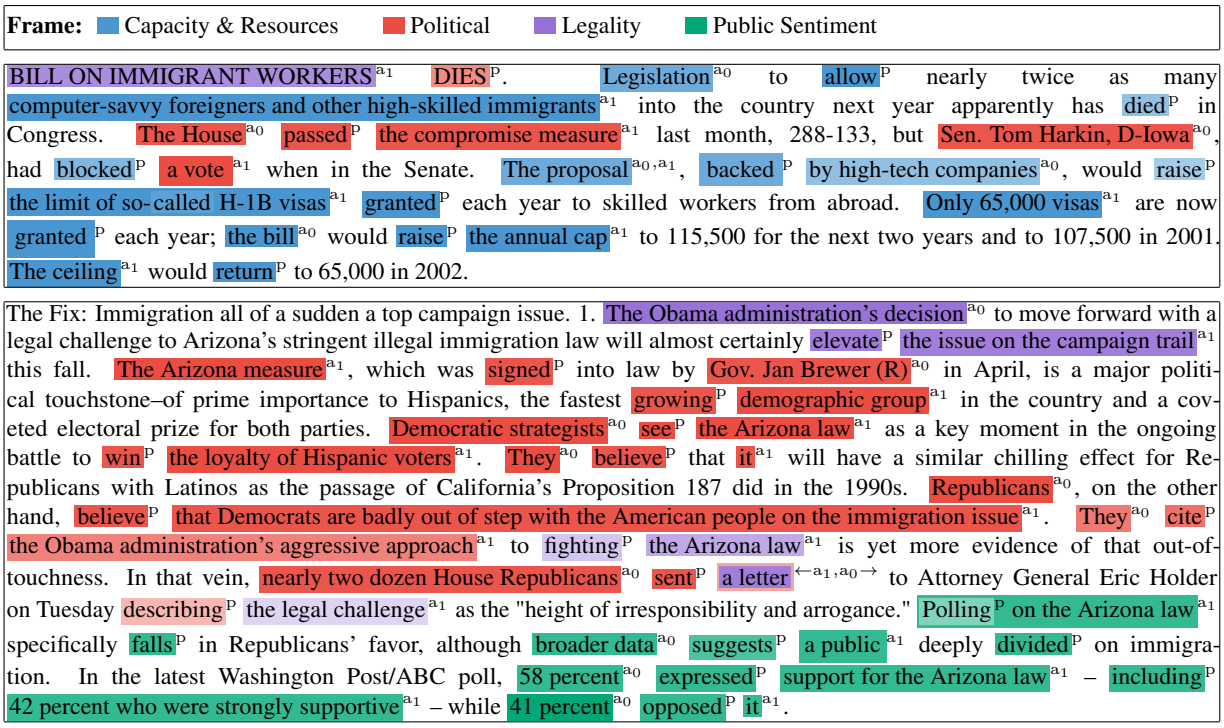

Table 4: Two articles from the MFC, annotated with SRL span-level frame predictions generated by FRISS. The true frame label of both articles is Political (red). Each detected span (p, a_0 or a_1) has been highlighted with its most closely associated frame. Darker shades indicate higher confidence. The top document is mis-classified as "Capacity & Resources" (blue), the bottom document is classified correctly.

cesses (reforms, cases, movements), or concepts under debate (political asylum). In future work, we aim to leverage these representations in scalable, in-depth analyses of issue-specific media framing. A full table illustrating the learnt descriptors for all 15 frames in the MFC and all three views is included in Table 6 in Appendix.

6 Conclusion

We presented FRISS, an interpretable model of media frame prediction, incorporating notions of emphasis framing (selective highlighting of issue aspects) and story framing (drawing on the events and actors described in an article). Our semi-supervised model predicts article-level frame of news articles, leveraging local predicate and argument level embeddings. We demonstrated its three-fold advantage: first, our model empirically outperforms existing models for frame classification; second, it can effectively leverage additional unlabeled data further improving performance; and, finally, its latent representations add transparency to classifier predictions and provide a nuanced article representation. The analyses provided by our model can support downstream applications such as automatic, yet transparent, highlighting of reporting patterns across countries or news outlets; or frame-guided summarization which can support both frame-balanced or frame-specific news summaries. In future work, we plan to extend our work to more diverse news outlets and policy issues, and explore richer latent models of article content, including graph representations over all involved events and actors.

Acknowledgments

We thank the anonymous reviewers for their helpful feedback and suggestions. This article was written with the support from the graduate research scholarship from the Melbourne School of Engineering, University of Melbourne provided to the first author. The original news articles used in this work were obtained from Lexis Nexis under the institutional licence held by the University of Melbourne. This research was undertaken using the LIEF HPC-GPGPU Facility hosted at the University of Melbourne, established with the assistance of LIEF Grant LE170100200.

References

Afra Feyza Akyürek, Lei Guo, Randa Elanwar, Prakash Ishwar, Margrit Betke, and Derry Tanti Wijaya. 2020. Multi-label and multilingual news framing analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8614–8624, Online. Association for Computational Linguistics.

Amber E Boydstun, Justin H Gross, Philip Resnik, and Noah A Smith. 2013. Identifying media frames and frame dynamics within and across policy issues.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 2: Short Papers*, pages 438–444. The Association for Computer Linguistics.

Dallas Card, Justin H. Gross, Amber E. Boydstun, and Noah A. Smith. 2016. Analyzing framing through the casts of characters in the news. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 1410–1420. The Association for Computational Linguistics.

Dennis Chong and James N Druckman. 2007. Framing theory. *Annu. Rev. Polit. Sci.*, 10:103–126.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT (1)*.

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

Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and agenda-setting in russian news: a computational analysis of intricate political strategies. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 3570–3580.

Lea Frermann and György Szarvas. 2017. Inducing semantic micro-clusters from deep multi-view representations of novels. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 1873–1883. Association for Computational Linguistics.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360, Online. Association for Computational Linguistics.

Kirk Hallahan. 1999. Seven models of framing: Implications for public relations. *Journal of public relations research*, 11(3):205–242.

Xiaochuang Han, Eunsol Choi, and Chenhao Tan. 2019. No permanent Friends or enemies: Tracking relationships between nations from news. 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 1660–1676, Minneapolis, Minnesota. Association for Computational Linguistics.

Pere-Lluís Huguet Cabot, Verna Dankers, David Abadi, Agneta Fischer, and Ekaterina Shutova. 2020. The Pragmatics behind Politics: Modelling Metaphor, Framing and Emotion in Political Discourse. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4479–4488, Online. Association for Computational Linguistics.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III. 2016. Feuding families and former Friends: Unsupervised learning for dynamic fictional relationships. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1534–1544, San Diego, California. Association for Computational Linguistics.

Eric Jang, Shixiang Gu, and Ben Poole. 2017. Categorical reparameterization with gumbel-softmax. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

Yangfeng Ji and Noah A. Smith. 2017. Neural discourse structure for text categorization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 996–1005.

Shima Khanehzar, Andrew Turpin, and Gosia Mikolajczak. 2019. Modeling political framing across policy issues and contexts. In *Proceedings of the The 17th Annual Workshop of the Australasian Language Technology Association*, pages 61–66, Sydney, Australia. Australasian Language Technology Association.

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*.

Siyi Liu, Lei Guo, Kate K. Mays, Margrit Betke, and Derry Tanti Wijaya. 2019a. Detecting frames in news headlines and its application to analyzing news framing trends surrounding U.S. gun violence. In *Proceedings of the 23rd Conference on Computational Natural Language Learning, CoNLL 2019, Hong Kong, China, November 3-4, 2019*, pages 504–514. Association for Computational Linguistics.

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

Jörg Matthes and Matthias Kohring. 2008. The content analysis of media frames: Toward improving reliability and validity. *Journal of communication*, 58(2):258–279.

Viet An Nguyen. 2015. *Guided Probabilistic Topic Models for Agenda-Setting and Framing*. Ph.D. thesis.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Hwee Tou Ng, Anders Björkelund, Olga Uryupina, Yuchen Zhang, and Zhi Zhong. 2013. Towards robust linguistic analysis using ontonotes. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning, CoNLL 2013, Sofia, Bulgaria, August 8-9, 2013*, pages 143–152.

Peng Shi and Jimmy Lin. 2019. Simple BERT models for relation extraction and semantic role labeling. *CoRR*, abs/1904.05255.

Oren Tsur, Dan Calacci, and David Lazer. 2015. A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pages 1629–1638.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in neural information processing systems*, pages 5753–5763.

A Frame Descriptors Extractions

We denote the set of spans that belong to a specific {frame, view} combination as C_z^f, for all possible combinations of frame categories and views. We obtain the high confidence set $\hat{C}_z^f$ for each specific {frame, view} combination by only preserving the spans whose $g_z^f > 0.8$. To obtain a more general sets of spans, we remove the stopwords and lemmatize the spans. To normalize the sets of high-confidence spans, we calculate the inverse-document frequency for each span C_z^f and sort the spans accordingly by the inverse-document frequency to obtain the representative descriptors. Table 6 illustrating the learnt descriptors for all 15 frames in the MFC and all three views.

Frame	Frame description	% IMM
Economic:	costs, benefits, or other financial implications	7.0%
Capacity and Resources:	availability of physical, human or financial resources, and capacity of current systems Morality: religious or ethical implications	3.5%
Morality:	religious or ethical implications	1.3%
Fairness and Equality:	balance or distribution of rights, responsibilities, and resources	2.6%
Legality, Constitutionality, Jurisdiction:	rights, freedoms, and authority of individuals, corporations, and government	16.1%
Policy Prescription and Evaluation:	discussion of specific policies aimed at addressing problems	8.0%
Crime and Punishment:	effectiveness and implications of laws and their enforcement	13.6%
Security and Defence:	threats to welfare of the individual, community, or nation	4.9%
Health and Safety:	health care, sanitation, public safety	4.0%
Quality of Life:	threats and opportunities for the individual's wealth, happiness, and well-being	7.0%
Cultural Identity:	traditions, customs, or values of a social group in relation to a policy issue	9.%
Public Sentiment:	attitudes and opinions of the general public, including polling and demographics	4.1%
Political:	considerations related to politics and politicians, including lobbying, elections, and attempts to sway voters	16.3%
External Regulation and Reputation:	international reputation or foreign policy of the U.S.	2.2%
Other:	any coherent group of frames not covered by the above categories	0.2%

Table 5: Framing dimensions from (Boydstun et al., 2013). The final column (% IMM) denotes the frame prevalence in the Immigration portion of the MFC used in the experiments reported in this paper.

ARG0	USCIS, state department, agency, federal official, IN , immigration service, ASA international, visitor Trump, house republican, Obama, democrat, senate, rubio, tancredo, Mr. romney, Mr. bush, clinton, gop supreme court, justice, federal judge, court, board immigration appeal, high court, judge, justice department organizer, activist, protester, demonstrator, marcher, immigrant advocate, mayor, poll, group, coalition grower, farmer, taxpayer, company, employer, native, foreign, business, high-tech company, federal government church, god, bishop, bible, course action, christ, parishioner, archbishop, organization, local jewish leader Mr.Crocker's letter, deputization, immigrants' rights group, American, critic, student, bigotry, commission executive order, Trump administration, legislation, legislator, measure, Obama administration, bill, provision federal agent, prosecutor, investigator, federal authority, federal immigration agent, Tyson, federal prosecutor FBI, ashcroft, coast guard, border patrol, Mr. Ashcroft, terrorist, homeland security official, border patrol agent health official, federal health official, doctor, migrant, hospital, disease, smuggler, patient, virus, mother, novice new office, teacher, perez, parent, mother, student, danilda, honduran immigrant christino castro, child, family Ziegler, census bureau, foreign, child immigrant, people want citizen country, new immigrant, museum Israel, Cuba, Mexican government, bahamian government, Castro, Clinton administration, Mexico, Cuban
predicate	process, handle, swamp, accommodate, wait, exceed, fill, reduce, flood, crowd, clear, jam, overwhelm, rush veto, defeat, vote, win, introduce, endorse, elect, oppose, overhaul, stall, criticize, derail, unveil, push, split, do sue, uphold, entitle, appeal, shall, violate, file, rule, challenge, qualify, dismiss, prove, expire, pending, block chant, march, protest, rally, wave, gather, organize, stag, shout, cheer, oppose, demand, denounce, favor, draw cost, import, invest, afford, contribute, save, employ, earn, cut, fill, fund, compete, depend, attract, educate, buy pray, worship, forgive, love, welcome, thank, bless, stand, honor, urge, hope, speak, join, recognize, offer discriminate, treat, single, persecute, complain, deserve, harass, punish, offend, tolerate, target, ignore verify, prohibit, streamline, crack, aim, implement, propose, tighten, fix, require, overhaul, bar, introduce, ban indict, sentence, plead, convict, fine, conspire, smuggle, harbor, sell, acquit, raid, shoot, commit, nab, arrest patrol, beef, track, apprehend, secure, tighten, overstay, intercept, investigate, link, cross, pose, deploy, fly infect, injure, hospitalize, die, drown, suffer, rescue, kill, cross, fell, crash, treat, flip, scorch, test, cause, hit cry, graduate, felt, imagine, feel, sleep, reunite, enrol, worry, enroll, sit, remember, escape, miss, love, learn assimilate, settle, celebrate, immigrate, account, teach, welcome, learn, found, preserve, spangle, publish, melt discuss, resume, press, ease, accept, visit, defect, meet, express, elect, legalize, urge, agree, refuse, talk, promote
ARG1	application, foreign worker, applicant, cap, number, application process, time, appointment, green card, staff amendment, reform, legislation, voter, senate bill, immigration bill, Latino voter, election, immigration law asylum, lawsuit, suit, status, case, Elian, legal status, permanent residency, license, decision, ruling, hearing rally, marcher, march, protest, movement, crowd, event, attention, message, poll, protester, immigration reform wage, economy, tax, state tuition, money, job, cost, income, worker, service, business, fee, foreign, budget church, sanctuary, god, refuge, yoga, faith, campaign, politics, home, better life, violence, pray, change heart racial profiling, discrimination, due process, right, Latino, every legal immigrant, Hispanic, advantage, woman path, system, legal immigration, legal status, immigration status, program, policy, require legislative approval guilty, crime, investigation, deportation proceeding, criminal, illegal alien, bribe, drug, trinket, death penalty border security, security, terrorist, national security, terrorism, fence, wall, border, information, troop medical care, disease, health insurance, health care, medical treatment, body, treatment, coverage, prenatal care food stamp, English, high school, goodbye, better life, everything, life, school, family, poverty, kid, father population, immigrant population, black, America, Asian, resident, Spanish, home, book, American dream Cuba, agreement, meeting, Mexico, Cuban, negotiation, office, Haiti, Mexican government, island, migrant

Table 6: Spans inferred as most highly associated with: Capacity & Resources (), Political (), Legality (), Public Sentiment (), Economic (), Morality (), Fairness & Equality (), Policy Prescription & Evaluation (), Crime & Punishment (), Security & Defense (), Health & Safety (), Quality of Life (), Cultural Identity(), External Regulation & Reputation () frames.