Do "Undocumented Workers" == "Illegal Aliens"? Differentiating Denotation and Connotation in Vector Spaces

Albert Webson[1,2], Zhizhong Chen[3], Carsten Eickhoff[1], and Ellie Pavlick[1]

{albert_webson, zhizhong_chen, carsten, ellie_pavlick}@brown.edu
[1]Department of Computer Science, Brown University
[2]Department of Philosophy, Brown University
[3]Department of Physics, Brown University

Abstract

In politics, neologisms are frequently invented for partisan objectives. For example, "undocumented workers" and "illegal aliens" refer to the same group of people (i.e., they have the same denotation), but they carry clearly different connotations. Examples like these have traditionally posed a challenge to reference-based semantic theories and led to increasing acceptance of alternative theories (e.g., Two-Factor Semantics) among philosophers and cognitive scientists. In NLP, however, popular pretrained models encode both denotation and connotation as one entangled representation. In this study, we propose an adversarial neural network that decomposes a pretrained representation as independent denotation and connotation representations. For intrinsic interpretability, we show that words with the same denotation but different connotations (e.g., "immigrants" vs. "aliens", "estate tax" vs. "death tax") move closer to each other in denotation space while moving further apart in connotation space. For extrinsic application, we train an information retrieval system with our disentangled representations and show that the denotation vectors improve the viewpoint diversity of document rankings.

1 Introduction

Language carries information through both denotation and connotation. For example, a reporter writing an article about the leftmost wing of the Democratic party can choose to refer to the group as "progressives" or as "radicals". The word choice does not change the individuals referred to, but it does communicate significantly different sentiments about the policy positions discussed. This type of linguistic nuance presents a significant challenge for natural language processing systems, most of which fundamentally assume words to have similar meanings if they are surrounded in similar word contexts. Such assumption risks confusing differences in connotation for differences in denotation or vice versa. For example, using a common skip-gram model (Mikolov et al., 2013) trained on a news corpus (described in §3.2), Figure 1 shows nearest neighbors of "government-run healthcare" and "economic stimulus". The resulting t-SNE clusters are influenced as much by policy denotation (shapes) as they are by partisan connotation (colors[1]). Using these entangled representations in applications such as information retrieval could have pernicious consequences such as reinforcing ideological echo chambers and political polarization. For example, a right-leaning query like "taxpayer-funded healthcare" could make one equally (if not more) likely to see articles about "totalitarian" and "horror stories" than about "affordable healthcare".

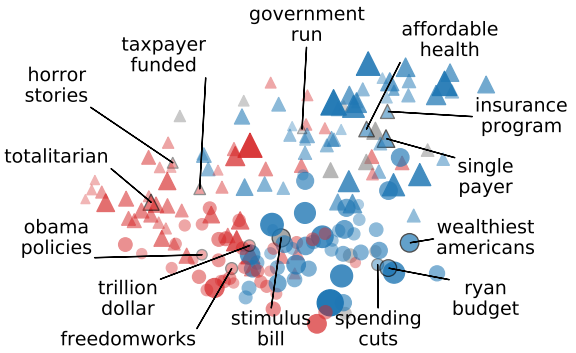

Figure 1: Nearest neighbors of government-run healthcare (triangles) and economic stimulus (circles). Note that words cluster as strongly by policy denotation (shapes) as by partisan connotation (colors); namely, pretrained representations conflate denotation with connotation. Plotted by t-SNE with perplexity = 10.

To address this, we propose classifier probes that

[1]Throughout this paper, blue reflects partisan leaning toward the Democratic Party and red reflects partisan leaning toward the Republican Party in the United States.

Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pages 4090–4105,
November 16–20, 2020.

measure denotation and connotation information in a given pretrained representation, and we arrange the probe losses in an adversarial setup in order to decompose the entangled pretrained meaning into distinct denotation and connotation representations (§4). We evaluate our model intrinsically and show that the decomposed representations effectively disentangle these two dimensions of semantics (§5). We then apply the decomposed vectors to an information retrieval task and demonstrate that our method improves the viewpoint diversity of the retrieved documents (§6). All data, code, preprocessing procedures, and hyperparameters are included in the appendix and our GitHub repository.[2]

2 Philosophical Motivation

Consider the following two sentences: "Undocumented workers are undocumented workers" vs. "Undocumented workers are illegal aliens". Frege (1892) famously used sentence pairs like these, which have the same truth conditions but clearly different meanings, in order to argue that meaning is composed of two components: "reference", which is some set of entities or state of affairs, and "sense", which accounts for how the reference is presented, encompassing a large range of aspects such as speaker belief and social convention.

In contemporary philosophy of language, the sense and reference argument has evolved into debates of semantic externalism vs. internalism and referential vs. conceptual role semantics. Externalists and referentialists[3] continue the truth-conditional tradition and emphasize meaning as some entity to which one is causally linked, invariant of one's psychological encoding of the referent (Putnam, 1975; Kripke, 1972). On the other hand, conceptual role semanticists emphasize meaning as what inferences one can draw from a lexical concept, deemphasizing the exact entities which the concept includes (Greenberg and Harman, 2005). Naturally, a popular position takes the Cartesian product of both schools of meaning (Block, 1986; Carey, 2009). This view is known as Two-Factor Semantics, and it forms the inspiration for our work. To avoid confusion with definitions from existing literature, we use the terms "denotation" and "connotation" rather than "reference" and "concept" when discussing our models in this paper.

[2]https://github.com/awebson/congressional_adversary

[3]Technically, one can be a referentialist while also being a semantic internalist. See Gasparri and Marconi (2019) for a nuanced overview as well as related theories in linguistics and cognitive science.

3 Data

We assume that it is possible to disentangle the two factors of semantics by grounding language to different components of the non-linguistic context. In particular, our approach assumes access to a set of training sentences, each of which grounds to a denotation d (which approximates reference) or a connotation c (which approximates conceptual inferences). We require at least one of d or c to be observed, but we do not require both (elaborated in §4.3). In this work, d and c are discrete symbols. However, our model could be extended to settings in which d and c are feature vectors.

While we are interested in learning lexical-level denotation and connotation, we train on sentence- and document-level speaker and reference labels. We argue that this emulates a more realistic form of supervision. For example, we often have metadata about a politician (e.g., party and home state) when reading or listening to what they say, and we are able to aggregate this to make lexical-level judgements about denotation and connotation.

We experiment on two corpora: the Congressional Record (CR) and the Partisan News Corpus (PN), which differ in linguistic style, partisanship distribution (Figure 2), and the available labels for grounding denotation and connotation.

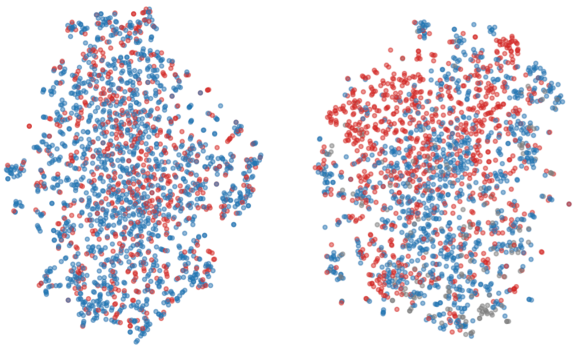

Figure 2: Vector spaces that result from training vanilla word2vec on the Congressional Record (left) and Partisan News (right). We evaluate on both corpora, but note that the Partisan News corpus better exemplifies the problem we target where words cluster strongly according to ideological stance.

3.1 Congressional Record

The Congressional Record (CR) is the official transcript of the floor speeches and debates of

Name	Corpus	Vocab.	Num. Sent.	Denotation Grounding	Connotation Grounding
CR BILL	Congr. Record	21,170	381,847	legislation title (1,029-class)	speaker party (2-class)
CR TOPIC	Congr. Record	21,170	381,847	policy topic (41-class)	speaker party (2-class)
CR PROXY	Congr. Record	111,215	5,686,864	none (LM proxy)	speaker party (2-class)
PN PROXY	Partisan News	138,439	3,209,933	none (LM proxy)	publisher partisan leaning (3-class)

Table 1: Summary of model variants experimented.

the United States Congress dating back to 1873. Gentzkow et al. (2019) digitized and identified approximately 70% of these speeches with a unique speaker, where each speaker is labeled with their gender, party, chamber, state, and district. To constrain the political and linguistic change over time, we use a subset of the corpus from 1981 to 2011.[4]

In order to assign labels that can be used as proxies of denotation, we weakly label each sentence with both its legislative topic and the specific bill being debated.[5] To do this, we collected a list of congressional bills from the U.S. Government Publishing Office.[6] For our purposes, this data provides the congressional session, policy topic, and an informal short title for each bill. We perform a regular expression search for each bill's short title among the speeches in its corresponding congressional session. For bills that are mentioned at least 3 times, we assume that the speech in which the bill was mentioned as well as 3 subsequent speeches are referring to that bill, and we label each speech with the title and the policy topic of that bill. Speeches that are not labeled by this process are discarded. Additional details and examples are given in Appendix D.

3.2 Partisan News Corpus

Hyperpartisan News is a set of web articles collected for a 2019 SemEval Task (Kiesel et al., 2019). It consists of articles scraped from the political sections of 383 news outlets in English. Each article is associated with a publisher which, in turn, has been manually labeled with a partisan leaning on a five-point scale: "left, center-left, center, center-right, right". Upon manual inspection, we find that the distinctions between right vs. center-right and left vs. center-left are prone to annotation artifacts. Therefore, we collapse these labels into a three-point scale, and we refer to this 3-class corpus as the Partisan News (PN) corpus throughout. No denotation label is available for this corpus.

[4] 2011 is the latest session available for the Bound Edition of CR; 1981 is chosen because the Reagan Administration marks the last party realignment and thus we can expect connotation signals to remain reasonably consistent over this period.

[5] We also experimented with collecting more precise reference labels using the entity linkers of both Google Cloud and Facebook Research on a variety of corpora. However, the results of entity linking were too poor to justify pursuing this direction further. We would love to see future works that devise creative ways to include better denotation grounding.

[6] https://www.govinfo.gov/bulkdata/BILLSTATUS

4 Model

Section 4.1 describes our model architecture. Sections 4.2 and 4.3 then describe specific instantiations that we use in our experiments. These variants are summarized in Table 1.

4.1 Overall Architecture

Let $V_{\text{deno}}, V_{\text{conno}}, V_{\text{pretrained}}$ be the vector spaces of denotation, connotation, and pretrained spaces respectively. Our model consists of two adversarial decomposers:

$$D : V_{\text{pretrained}} \to V_{\text{deno}}$$
$$C : V_{\text{pretrained}} \to V_{\text{conno}}$$

The goal is to train D to *preserve* as much denotation information as possible while *removing* as much connotation information as possible from the pretrained representation. Symmetrically, C will preserve as much *connotation* as possible while removing as much *denotation* as possible from the pretrained representation.

For clarity, let us focus on D for now. To measure how much denotation or connotation structure is encoded in V_{deno}, we use two classifiers probes trained to predict the denotation label d or connotation label c, which yield two cross-entropy losses $\ell_{\text{deno. probe}}$ and $\ell_{\text{cono. probe}}$ respectively. In order to encourage the decomposer D to preserve denotation and remove connotation, we define its loss function as

$$L_D = \sigma(\ell_{\text{deno. probe}}) + \sigma(\ell_{\text{conno. adversary}})$$

where σ is the sigmoid function and

$$\ell_{\text{conno. adversary}} = \text{KL Div}\,(\text{conno. probe predicted dist.}, \text{uniform dist.})$$

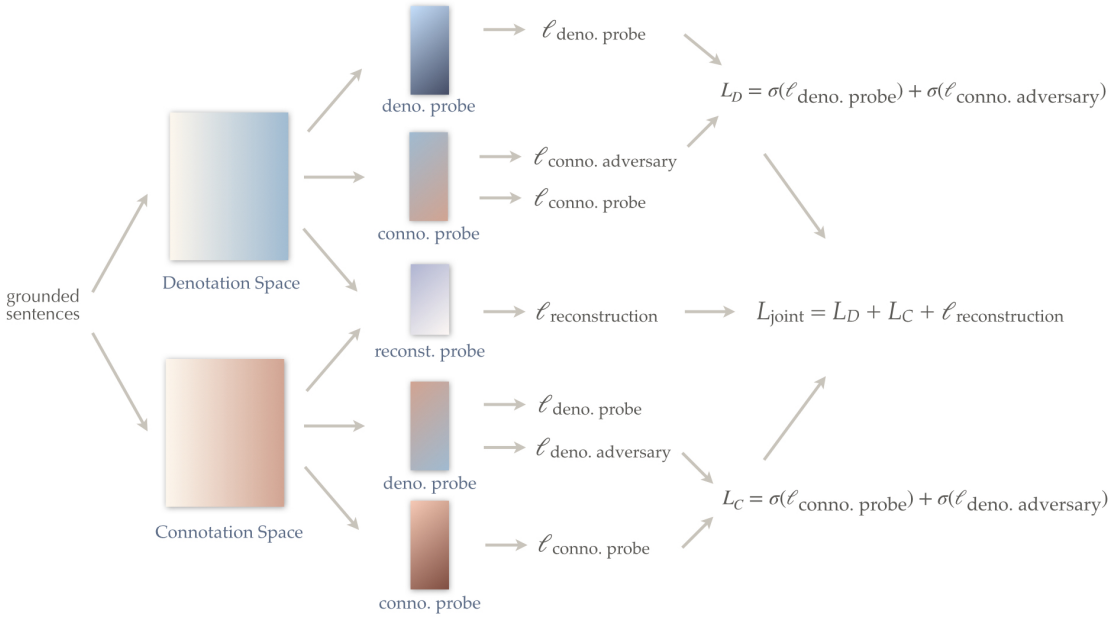

Figure 3: Overall model and composition of losses

The adversarial loss $\ell_{\text{conno. adversary}}$ rewards D to remove connotation structure such that the probe prediction is random. Meanwhile, the probes themselves are still only gradient updated with the usual cross-entropy losses—extracting and measuring as much denotation or connotation information as possible—independent of the decomposer D.

As shown in Figure 3, C is set up symmetrically, so it is trained with the usual classification loss from its connotation probe and a KL divergence adversarial loss from its denotation probe.

Finally, we impose a reconstruction probe R with the loss function:

$$\ell_{\text{recon.}} = 1 - \cos\text{sim}(R(v_{\text{deno}}, v_{\text{conno}}), v_{\text{pretrained}})$$

which enforces that the combination of denotation and connotation subspaces preserves all the semantic meaning of the original pretrained space, as opposed to merely encoding predictive features that maximize probe accuracies. (We verified in ablation experiments that this is in fact what happens without R.) Assembling everything together, the decomposers D and C are jointly trained with $L_{\text{Joint}} = L_D + L_C + \ell_{\text{recon.}}$.

In principle, D and C can be a variety of sentence encoders. In this work, we implement them as simple mean bags of static embedding for two reasons: First, it is difficult to interpret contextualized embedding for an individual word (especially for the type of analysis we present in §5). Second, many of the interesting heavily connotative expressions consist of multiple words (e.g., "socialized medicine", "universal healthcare") and compositionality is still far from being solved. Therefore, we conjoin multiword expressions with underscores so that we can model them in the same way as atomic words.[7]

4.2 Connotation Probes

We exploit the fact that much of the debate in American politics today is (sadly) reducible to partisan division (Lee, 2009; Klein, 2020), thus it is safe to define the connotation label of every document to be simply the partisanship of the speaker. Of course, connotation in the general domain can encompass much more than liberals vs. conservatives, and in future work, we hope to extend this to multifaceted connotations that are more true to the semantic theories as discussed in §2. For now, in CR, connotation is the speaker's party, and in PN, connotation is the partisan leaning of the publisher.

Again, in principle, the probes can be a variety of neural modules. In this work, we implement the connotation probes as 4-layer MLPs. We experimented with the more popular 1-layer MLP and 1-linear-layer probes. However, when the probes are shallow, the model converges before most of the information that should be removed is in fact removed. For example, when we use a 4-layer MLP

[7] Appendix B documents this preprocessing step in detail. Throughout this paper, "words" refers to both individual words and underscored short phrases.

probe on a decomposed representation trained with a 1-layer probe, the 4-layer probe accuracies are just as good as if the representation has not been decomposed at all. That is, our experiments suggest that the probes have to be sufficiently complex in order to truly measure what denotation/connotation structure is removed or preserved in a decomposed representation.

4.3 Denotation Probes

For the CR corpus, we experimented with two types of denotation labels: The specific piece of legislation under discussion and the general policy topic under discussion. In CR BILL, the label is one of the 1,029 short titles of bills. In CR TOPIC, the label is one of 41 policy topics. Both types of labels are annotated as described in §3.1. For the same reason as discussed in the previous paragraph, we implement the denotation probes as 4-layer MLPs.

Additionally, as mentioned in Footnote 5, precise denotation labels are difficult to collect, so we also experimented with more realistic settings (CR PROXY and PN PROXY) which do not use any denotation labels. In this case, we return to the theories discussed in §2 and note that, because semantic meaning can be partitioned into two components, we may assume pretrained representations encode the overall meaning and any aspects of meaning that are not explained by our connotation labels must belong to denotation.[8] Thus, we may continue to use the pretraining objective (in this implementation, skip-gram-style context word prediction) as a proxy probe for denotation information and rely on the adversarial connotation probe to remove connotation structure in the denotation space.

5 Intrinsic Evaluation

We confirm that our decomposed denotation and connotation spaces reflect their intended purposes by measuring their structures with homogeneity metrics (§5.1) on three sets of evaluation words (§5.2) as well as inspecting their t-SNE clusters.

5.1 Homogeneity Metrics

To quantify how much denotation or connotation structure is encoded in a vector space, we define the homogeneity (h_{deno}, h_{conno}) of a given space to be the average proportion of a query word's top-k nearest neighbors[9] which share the same denotation/connotation label as the query's own denotation/connotation label.[10] In particular, we are interested in comparing the delta of V_{deno} and V_{conno} against $V_{\text{pretrained}}$. For V_{deno}, we hope to see h_{deno} *increase* relative to the pretrained space and see h_{conno} *decrease* relative to the pretrained space. For V_{conno}, we hope to observe movement in the opposite direction.

As motivated in §3, our model is trained with labels at the sentence-level, while homogeneities are evaluated at the word-level. We assign a word's connotation label to simply be the party that uses the word most often. For CR BILL and CR TOPIC, we assign the word-level denotation label as either the bill or the topic that uses the word most often. For the PN corpus, no ground truth denotation label is available, so we cannot directly measure h_{deno}, but we show alternative evaluation in §5.3. Table 3 shows the baseline h_{deno} and h_{conno} scores for embeddings pretrained on each corpus and evaluating over two test sets of words (described in the next section).

5.2 Test Sets

We evaluate on words sampled in three different ways: **Random** is a random sample of 500 words drawn from each corpus' vocabulary that occur at least 100 times in order to filter out web scraping artifacts, e.g., URLs and author bylines. **High Partisan** is a sample of around 300 words from each corpus's vocabulary that occur at least 100 times and have high partisan skew; namely, words that are uttered by a single party more than 70% of the time. This threshold is chosen based on manual inspection, but we have evaluated on other thresholds as well with no significant difference in results. This High Partisan set is then bisected into two disjoint sets as dev and test data for model selection. All word sets sampled at different ratios are included in our released data. Finally, **Luntzesque** is a small set of manually-vetted pairs of words that are known to have the same denotation but different connotations. Most of them are drawn

[8] We acknowledge that this feels a bit backward: Ideally, in a Fregean sense, everything not explained by reference is left over to sense, rather than the converse. However, we are constrained by the available grounding. In a different setting, if we had explicit referential labels but no speaker information, we could use skip-gram as the proxy for connotation instead.

[9] We set $k = 10$, but we found that evaluation results remain robust across different choices of k.

[10] We also ran `sklearn.homogeneity_score` but saw no difference in trends, so we report our homogeneity metric for its simple interpretability.

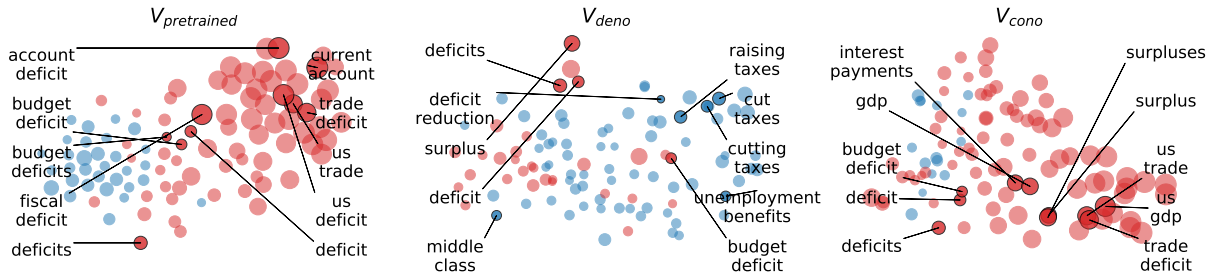

Figure 4: Neighborhood of "deficit" in $V_{\text{pretained}}$, V_{deno}, and V_{conno} of PN PROXY. Arrows point to the top-10 nearest neighbors. Colors reflect partisan leaning, where more opaque dots are more heavily partisan words. Note that in $V_{\text{pretained}}$ and in V_{conno}, the nearest neighbors are all Republican-leaning words, whereas they are balanced in V_{deno}.

		V_{deno} (and Δ with V_{pre})				V_{conno} (and Δ with V_{pre})			
Test Set	Model	h_{deno}	Δ ($\uparrow$)	h_{conno}	Δ ($\downarrow$)	h_{deno}	Δ ($\downarrow$)	h_{conno}	Δ ($\uparrow$)
High Partisan	CR BILL	0.28	+0.09	0.65	−0.11	0.02	−0.17	0.89	+0.13
	CR TOPIC	0.53	+0.18	0.59	−0.17	0.07	−0.28	0.98	+0.21
	CR PROXY	0.07	+0.00	0.71	−0.00	0.04	−0.03	0.99	+0.28
	PN PROXY	–	–	0.40	−0.26	–	–	0.76	+0.10
Random	CR BILL	0.14	+0.05	0.69	−0.01	0.04	−0.06	0.77	+0.07
	CR TOPIC	0.31	+0.02	0.63	−0.07	0.14	−0.15	0.81	+0.11
	CR PROXY	0.04	+0.00	0.64	−0.00	0.02	−0.03	0.85	+0.21
	PN PROXY	–	–	0.39	−0.21	–	–	0.69	+0.09

Table 2: Intrinsic evaluation results across models and test sets. Δ is change relative to $V_{\text{pretrained}}$ (Table 3). Arrows in parentheses mark the desired directions of change. Note that because denotation labels have far more classes than connotation labels, the magnitude of h_{deno} and h_{conno} are not directly comparable with each other.

	High Partisan		Random	
	h_{deno}	h_{conno}	h_{deno}	h_{conno}
CR BILLS	0.19	0.76	0.09	0.70
CR TOPIC	0.35	0.76	0.29	0.70
CR PROXY	0.07	0.71	0.05	0.64
PN PROXY	–	0.66	–	0.60

Table 3: Baseline homogeneity scores of embeddings pretrained on each corpus.

from *The New American Lexicon* (Luntz 2006[11]), a famous report from focus group research which explicitly prescribes word choices that are empirically favorable to the Republican party line.

5.3 Results

Overall, we see that our V_{deno} and V_{conno} spaces demonstrate the desired shift in homogeneities and structures, which is intuitively illustrated by Figure 4. Quantitatively, Table 2 enumerates the homogeneity scores of both decomposed spaces as well as their directions of change relative to the pretrained space. For V_{deno}, we see that denotation homogeneity h_{deno} consistently increases and connotation homogeneity h_{conno} consistently decreases as desired. Conversely, for V_{conno}, we see h_{conno} increases and h_{deno} decreases as desired. Further, we see that the magnitude of change is greater across the board for the highly partisan words than for random words, which is expected as the highly partisan words are usually loaded with more denotation or connotation information that can be manipulated. The only exception is CR PROXY's V_{deno}, which sees no significant movement in either direction. This is understandable because CR PROXY is not trained with ground truth denotation labels. (We evaluate it with the labels from CR BILL).

As means of closer inspection, we compute the cosine similarities of words in our Luntz-esque analysis set. Because these pairs of words are known to be political euphemisms (e.g. "estate tax" and "death tax", which refer to the same tax policy but imply opposite partisanship), we expect these pairs to become more cosine similar in V_{deno} and less cosine similar in V_{conno}. As shown in Table 4, even without ground truth denotation labels, the V_{deno} of CR PROXY and PN PROXY still preserve the pretrained denotation structure reasonably well. For pairs that do see decrease in V_{deno} similarity, the errors are far smaller relative to their correct

[11] This is a leaked report circulated via a Google Drive link which has been taken offline since. A copy is included in our released data.

	CR BILL			CR TOPIC			CR PROXY			PN PROXY		
	V_{pre}	$\Delta V_d(\uparrow)$	$\Delta V_c(\downarrow)$	V_{pre}	$\Delta V_d(\uparrow)$	$\Delta V_c(\downarrow)$	V_{pre}	$\Delta V_d(\uparrow)$	$\Delta V_c(\downarrow)$	V_{pre}	$\Delta V_d(\uparrow)$	$\Delta V_c(\downarrow)$
undocumented workers/illegal aliens	0.81	+0.03	−0.01	0.81	−0.09	+0.14	0.95	+0.03	−1.28	0.96	+0.01	−0.20
estate tax/death tax	0.89	+0.05	−0.76	0.89	+0.08	−0.84	0.96	+0.00	−0.98	0.93	+0.01	−0.06
capitalism/free market	0.79	+0.11	+0.03	0.79	+0.14	+0.16	0.85	−0.07	−0.20	0.96	−0.01	−0.02
foreign trade/international trade	0.90	−0.05	+0.02	0.90	+0.02	−0.01	0.86	+0.05	−0.40	0.93	+0.03	+0.00
public option/government-run	0.67	+0.06	−0.57	0.67	+0.24	−0.84	0.92	+0.02	−1.08	0.97	+0.00	−0.01
trickle-down/cut taxes	–	–	–	–	–	–	0.87	+0.02	−0.51	0.95	+0.02	−0.12
voodoo economics/supply-side	–	–	–	–	–	–	0.95	−0.04	−0.07	0.91	+0.05	−0.05
tax expenditures/spending programs	–	–	–	–	–	–	0.93	−0.17	−1.03	0.99	+0.00	−0.16
waterboarding/interrogation	–	–	–	–	–	–	0.90	−0.04	−0.22	0.97	+0.01	−0.01
socialized medicine/single-payer	–	–	–	–	–	–	0.88	−0.11	−0.56	0.89	+0.02	−0.03
political speech/campaign spending	–	–	–	–	–	–	0.86	−0.02	−0.81	0.99	+0.00	−0.05
star wars/strategic defense initiative	–	–	–	–	–	–	0.91	−0.16	−0.69	–	–	–
nuclear option/constitutional option	–	–	–	–	–	–	0.97	−0.14	−1.30	–	–	–
Changes in the Correct Direction		4/5	3/5		4/5	3/5		5/13	13/13		10/11	10/11

Table 4: Changes in cosine similarity (relative to $V_{\text{pretrained}}$) for known political euphemism' pairs, i.e. words with the same denotation but opposite partisan connotation. Omitted entries are out of vocabulary.

reduction in V_{conno} similarity. For example, "political speech" and "campaign spending" experience a small (-0.02) decrease in denotation similarity; in exchange, the model correctly recognizes that the two words have opposite ideologies (-0.81 in connotation similarity) on the issue of whether unlimited campaign donation is shielded by the First Amendment as "political speech".

6 Extrinsic Evaluation

Ultimately, our work aims to be more than just a theoretical exercise, but also to enable greater control over how sensitive NLP systems are to denotation vs. connotation in downstream tasks. To this end, we construct an ad hoc information retrieval task. We compare a system built on top of $V_{\text{pretrained}}$ to systems built on top of V_{deno} and V_{conno} in terms of both the quality of the ranking and the ideological diversity represented among the top results.

6.1 Setup

We focus only on PN PROXY for this evaluation since it best matches the setting where we would expect to apply these techniques in practice: (1) We cannot always assume access to discrete denotation labels. (2) Language in the PN corpus is strongly influenced by ideology (as shown in Figure 2).

To generate a realistic set of queries, we start with 12 seed words from our vocabulary, chosen based on a list of the most important election issues for Democrat and Republican voters according to a recent Gallup Poll[12]. This results in the following list: "economy, healthcare, immigration, women's rights, taxes, wealth, guns, climate change, foreign policy, supreme court, tariffs, special counsel". Then, for each seed word, we take 5 left-leaning seeds to be the 5 nearest neighbors according to $V_{\text{pretrained}}$, filtered to words which occur at least 100 times and for which at least 70% of occurrences appeared in left-leaning articles. We similarly chose 5 right-leaning seeds. We then submit each partisan seed to the Bing Autosuggest API and retrieve 10 suggestions each. We manually filter the list of queries to remove those that do not reflect the intended word sense (e.g., "VA" leading to queries about Virginia rather than the Veterans Administration) and those which are not well matched to our document collection (e.g., queries seeking dictionary definitions, job openings, or specific websites such as Facebook). Our final list contains 410 queries, 216 left-leaning and 194 right-leaning. Table 5 shows several examples, the full list is included in the supplementary material.

Wealth: globalist agenda ∘ globalist leaders ∘ extreme poverty rates ∘ romneys ties to burisma
Women's Rights: title ix impact ∘ safe spaces and snowflakes ∘ anti-choice zealots ∘ marriage equality court case
Immigration: illegal immigrants at southern border ∘ illegals caught voting 2016 ∘ drug policy fbi ∘ opioid crisis afghanistan

Table 5: Example right- and left-leaning queries generated using the procedure described.

6.2 Models

We generate a ranked list of documents for each query in a two-step manner: (1) We pre-select the 5,000 most relevant documents according to a tra-

[12] https://news.gallup.com/poll/244367/top-issues-voters-healthcare-economy-immigration.aspx

ditional BM25 model (Robertson et al., 1995) with default parameters. (2) This initial set of documents is then ranked using DRMM (Guo et al., 2016), a neural relevance matching model for ad-hoc retrieval. We train our retrieval model on the MS MARCO collection (Bajaj et al., 2016) of 550,000 queries and 8.8 million documents from Bing. To highlight the effect of pretrained vs. decomposed word embeddings, we freeze our word embeddings during retrieval model training. While (1) is purely based on tf-idf style statistics and remains static for all compared conditions, (2) is repeated for every proposed word embedding. This results in a ranked list of the top 100 most relevant documents for each query and word embedding.

6.3 Results

We compare the results of the DRMM retrieval model using different word embeddings in terms of quality and diversity of viewpoints reflected in the ranked results. To measure diversity, we report the overall distribution of political leanings among the top 100 documents and the rank-weighted α-nDCG (Clarke et al., 2008) diversity score. For α-nDCG, higher values indicate a more diverse list of results whose political leanings are evenly distributed across result list ranks. To measure ranking quality, we take a sample of 10 queries and collect top 10 results returned by each model variant, for a total of 300 query/document pairs. We shuffle the list of pairs to avoid biasing ourselves, and manually label each pair for whether or not the document is relevant to the query. We report Precision@10 estimated based on these 10 queries.

Figure 5 shows the overall party distributions. Table 6 reports the α-nDCG and P@10 metrics. We can see that models which use V_{deno} produce more diverse rankings than do models that use $V_{\text{pertained}}$, with V_{deno} producing an α-nDCG@100 of 0.94 vs. 0.92 for pretrained. This trend is especially apparent in the rankings returned for right-leaning queries: Under the pretrained model, 57% of the documents returned came from right-leaning news sources, whereas under the V_{deno}-based model, the results are nearly perfectly balanced between news sources. However, we do see a drop in precision when using V_{deno}. This is not surprising given the limitations observed in §5. If we had access to ground-truth denotation labels when training V_{deno}, we might expect to see these numbers improve. This is a promising direction for future work.

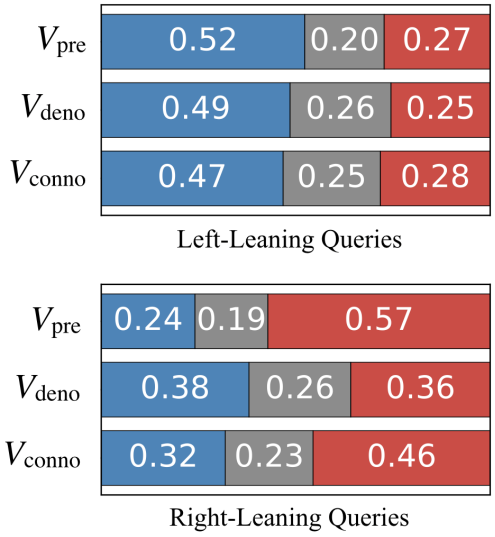

Figure 5: Distribution of partisanship of news source for top 100 documents for right-leaning and left-leaning queries. Red = right-leaning news sources; blue = left-leaning; gray = nonpartisan or apolitical.

	α-nDCG		Gini		
	@10	@100	L	R	P@10
$V_{\text{pretrained}}$	0.907	0.915	0.215	0.207	0.78
V_{deno}	0.922	0.944	0.160	0.080	0.37
V_{conno}	0.904	0.914	0.147	0.153	0.64

Table 6: Retrieval metrics. For α-nDCG, higher means more diverse; for Gini, lower means more diverse.

7 Related Work

Embedding Augmentation. At the lexical level, there is substantial literature that supplements pretrained representations with desired information (Faruqui et al., 2015; Bamman et al., 2014) or improves their interpretability (Murphy et al., 2012; Arora et al., 2018; Lauretig, 2019). However, existing works tend to focus on evaluating the dictionary definitions of words, less so on grounding words to specific real world referents and, to our knowledge, no major attempt yet in interpreting and manipulating the denotation and connotation dimensions of meaning as suggested by the semantic theories discussed in §2. While we do not claim to do full justice to conceptual role semantics either, this paper furnishes a first attempt at implementing a school of semantics introduced by philosophers of language and increasingly popular among cognitive scientists.

Style Transfer. At the sentence level, adversarial setups similar to ours have been previously ex-

plored for differentiating style and content. For example, Romanov et al. (2019); Yamshchikov et al. (2019); John et al. (2019) converted informal English to formal English and Yelp reviews from positive to negative sentiment. The motivation for such models is primarily natural language generation and the personalization thereof (Li et al., 2016). Additionally, our framing in terms of Frege's sense and reference adds clarity to the sometimes ill-defined problems explored in style transfer (e.g., treating sentiment as "style"). For example, "she is an undocumented immigrant" and "she is an illegal alien" have the same truth conditions but different connotations, whereas "the cafe is great" and "the cafe is terrible" have different truth conditions.

Modeling Political Language. There is a wealth of work on computational approaches for modeling political language (Glavaš et al., 2019). Within NLP, such efforts tend to focus more on describing how language differs between political subgroups, rather than recognizing similarities in denotation across ideological stances, which is the primary goal of our work. For example, Preoţiuc-Pietro et al. (2017); Han et al. (2019) attempt to predict a person's political ideology from their social media posts, Sim et al. (2013) detect ideological trends present in political speeches, Fulgoni et al. (2016) predict political leaning of news articles, and Padó et al. (2019) focuses on modeling the network structure of policy debates within society. Also highly related is work analyzing linguistic framing in news (Greene and Resnik, 2009; Choi et al., 2012; Baumer et al., 2015).

Echo Chambers and Search. The dangers of ideological "echo chambers" have received significant attention across NLP, information retrieval, and social science research communities. Dori-Hacohen et al. (2015) discuss the challenges of deploying information retrieval systems in controversial domains, and Puschmann (2019) looks specifically at the effects of search personalization on election-related information. Many approaches have been proposed to improve the diversity of search results, typically by identifying search facets *a priori* and then training a model to optimize for diversity (Tintarev et al., 2018; Tabrizi and Shakery, 2019; Lunardi, 2019). In terms of linguistic analyses, Rashkin et al. (2017) and Potthast et al. (2018) analyze stylistic patterns that distinguish fake news from real news. Duseja and Jhamtani (2019) study linguistic patterns that distinguish whether individuals are within social media echo chambers.

8 Summary

In this paper, we describe the problem of pretrained word embeddings conflating denotation and connotation. We address this issue by introducing an adversarial network that explicitly represents the two properties as two different vector spaces. We confirm that our decomposed spaces encode the desired structure of denotation or connotation by both quantitatively measuring their homogeneity and qualitatively evaluating their clusters and their representation of well-known political euphemisms. Lastly, we show that our decomposed spaces are capable of improving the diversity of document rankings in an information retrieval task.

Acknowledgment

We are grateful to Jesse Shapiro, Stephen Bach, Yongming Han, Tucker Berckmann, Daniel Smits, Jessica Forde, Dylan Ebert, Aaron Traylor, Charles Lovering, and Roma Patel for comments and discussions on the (many) early drafts of this paper. This research was supported by the Google Faculty Research Awards Program.

References

Sanjeev Arora, Yuanzhi Li, Yingyu Liang, Tengyu Ma, and Andrej Risteski. 2018. Linear algebraic structure of word senses, with applications to polysemy.

Payal Bajaj, Daniel Campos, Nick Craswell, Li Deng, Jianfeng Gao, Xiaodong Liu, Rangan Majumder, Andrew McNamara, Bhaskar Mitra, Tri Nguyen, et al. 2016. Ms marco: A human generated machine reading comprehension dataset. *arXiv preprint arXiv:1611.09268*.

David Bamman, Chris Dyer, and Noah A. Smith. 2014. Distributed representations of geographically situated language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834, Baltimore, Maryland. Association for Computational Linguistics.

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 *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1472–1482, Denver, Colorado. Association for Computational Linguistics.

Ned Block. 1986. Advertisement for a semantics for psychology. *Midwest studies in philosophy*, 10:615–678.

Susan Carey. 2009. *The Origin of Concepts*. Oxford series in cognitive development. Oxford University Press.

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 *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, pages 70–79, Jeju, Republic of Korea. Association for Computational Linguistics.

Charles LA Clarke, Maheedhar Kolla, Gordon V Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. 2008. Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 659–666.

Shiri Dori-Hacohen, Elad Yom-Tov, and James Allan. 2015. Navigating controversy as a complex search task. In *SCST@ ECIR*. Citeseer.

Nikita Duseja and Harsh Jhamtani. 2019. A sociolinguistic study of online echo chambers on twitter. In *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science*, pages 78–83, Minneapolis, Minnesota. Association for Computational Linguistics.

Manaal Faruqui, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. 2015. Retrofitting word vectors to semantic lexicons. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1606–1615, Denver, Colorado. Association for Computational Linguistics.

Gottlob Frege. 1892. Über sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25–50.

Dean Fulgoni, Jordan Carpenter, Lyle Ungar, and Daniel Preoţiuc-Pietro. 2016. An empirical exploration of moral foundations theory in partisan news sources. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3730–3736, Portorož, Slovenia. European Language Resources Association (ELRA).

Luca Gasparri and Diego Marconi. 2019. Word Meaning. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, fall 2019 edition. Metaphysics Research Lab, Stanford University.

Matthew Gentzkow, Jesse M Shapiro, and Matt Taddy. 2019. Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica*, 87(4):1307–1340.

Goran Glavaš, Federico Nanni, and Simone Paolo Ponzetto. 2019. Computational analysis of political texts: Bridging research efforts across communities. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 18–23, Florence, Italy. Association for Computational Linguistics.

Mark Greenberg and Gilbert Harman. 2005. Conceptual role semantics.

Stephan Greene and Philip Resnik. 2009. More than words: Syntactic packaging and implicit sentiment. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 503–511, Boulder, Colorado. Association for Computational Linguistics.

Jiafeng Guo, Yixing Fan, Qingyao Ai, and W Bruce Croft. 2016. A deep relevance matching model for ad-hoc retrieval. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 55–64.

Jiyoung Han, Youngin Lee, Junbum Lee, and Meeyoung Cha. 2019. The fallacy of echo chambers: Analyzing the political slants of user-generated news comments in Korean media. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pages 370–374, Hong Kong, China. Association for Computational Linguistics.

Vineet John, Lili Mou, Hareesh Bahuleyan, and Olga Vechtomova. 2019. Disentangled representation learning for non-parallel text style transfer. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 424–434, Florence, Italy. Association for Computational Linguistics.

Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, and Martin Potthast. 2019. SemEval-2019 task 4: Hyperpartisan news detection. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 829–839, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Ezra Klein. 2020. *Why We're Polarized*. Profile Books.

Saul A Kripke. 1972. Naming and necessity. In *Semantics of natural language*, pages 253–355. Springer.

Adam Lauretig. 2019. Identification, interpretability, and Bayesian word embeddings. In *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science*, pages 7–17, Minneapolis, Minnesota. Association for Computational Linguistics.

Frances E Lee. 2009. *Beyond ideology: Politics, principles, and partisanship in the US Senate*. University of Chicago Press.

Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan. 2016. A persona-based neural conversation model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 994–1003, Berlin, Germany. Association for Computational Linguistics.

Gabriel Machado Lunardi. 2019. Representing the filter bubble: Towards a model to diversification in news. In *International Conference on Conceptual Modeling*, pages 239–246. Springer.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.

Brian Murphy, Partha Talukdar, and Tom Mitchell. 2012. Learning effective and interpretable semantic models using non-negative sparse embedding. In *Proceedings of COLING 2012*, pages 1933–1950, Mumbai, India. The COLING 2012 Organizing Committee.

Sebastian Padó, Andre Blessing, Nico Blokker, Erenay Dayanik, Sebastian Haunss, and Jonas Kuhn. 2019. Who sides with whom? towards computational construction of discourse networks for political debates. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2841–2847, Florence, Italy. Association for Computational Linguistics.

Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2018. A stylometric inquiry into hyperpartisan and fake news. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 231–240, Melbourne, Australia. Association for Computational Linguistics.

Daniel Preoţiuc-Pietro, Ye Liu, Daniel Hopkins, and Lyle Ungar. 2017. Beyond binary labels: Political ideology prediction of twitter users. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 729–740, Vancouver, Canada. Association for Computational Linguistics.

Cornelius Puschmann. 2019. Beyond the bubble: Assessing the diversity of political search results. *Digital Journalism*, 7(6):824–843.

Hilary Putnam. 1975. The meaning of 'meaning'. *Philosophical papers*, 2.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.

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

Stephen E Robertson, Steve Walker, Susan Jones, Micheline M Hancock-Beaulieu, Mike Gatford, et al. 1995. Okapi at trec-3. *Nist Special Publication Sp*, 109:109.

Alexey Romanov, Anna Rumshisky, Anna Rogers, and David Donahue. 2019. Adversarial decomposition of text representation. 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 815–825, Minneapolis, Minnesota. Association for Computational Linguistics.

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

Shayan A Tabrizi and Azadeh Shakery. 2019. Perspective-based search: a new paradigm for bursting the information bubble. *FACETS*, 4(1):350–388.

Nava Tintarev, Emily Sullivan, Dror Guldin, Sihang Qiu, and Daan Odjik. 2018. Same, same, but different: algorithmic diversification of viewpoints in news. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, pages 7–13.

Ivan P. Yamshchikov, Viacheslav Shibaev, Aleksander Nagaev, Jürgen Jost, and Alexey Tikhonov. 2019. Decomposing textual information for style transfer. In *Proceedings of the 3rd Workshop on Neural Generation and Translation*, pages 128–137, Hong Kong. Association for Computational Linguistics.

A Hyperparameters

- All classifier probes are 4-layer MLPs with hidden size 300, ReLU as nonlinearity, and dropout with $p = 0.33$.
- Decomposers D and C are embedding matrices of shape (vocab_size, 300). Recomposer R concatenates denotation and connotation as a 600-dimensional vector and then feed it into a linear layer of size (600, 300).
- The skip-gram loss follows the parameters recommended by Mikolov et al. (2013). Context window radius = 5. Negative samples per true context word = 10. We also subsample frequent words in exactly the same way as the original paper (equation 5) did with their threshold of 10^{-5}.
- We use Adam as our optimizer throughout. Learning rate = 1×10^{-3} for homogeneity and 1×10^{-5} for Luntz-esque models. Other parameters left as PyTorch default.
- We train 30 epochs for large corpora (CR PROXY and PN PROXY). 150 epochs for smaller corpora (CR TOPIC and CR BILL).
- With batch size = 1024, the smaller corpora take about half an hour to train on an RTX 2080 Ti or comparable GPUs. With batch size = 8192, The larger corpora take about 50 hours to train.
- PyTorch version = 1.6. CUDA version = 10.2.

B Preprocessing Procedures for Congressional Record

We use Stanford Stanza (Qi et al., 2020) for tokenization, part-of-speech tag, dependency parsing, and named entity recognition. We replace multi-word phrases with an atomic token. We source our phrases of interests from the following three pipelines:

1. Named entity recognizer.
2. Frequency-based collocation. (We experimented with PMI-based collocation, which yielded results that were more prone to artifacts and arbitrary threshold setting.)
3. Bigram and trigram constituents of parse trees that are (a) POS-tagged as noun phrase or verb phrase; (b) contain no stop words as in `nltk.corpus.stop_words`; (c) contain no parliamentary procedural words as in {"yield", "motion", "order", "ordered", "quorum", "roll", "unanimous", "mr.", "madam", "speaker", "chairman", "president", "senator", "gentleman", "colleague", "colleagues"}

From these sources, we filter vocabulary with minimum frequency = 15 for small corpora, 30 for large corpora. We then replace each phrase in the corpus by their respective tokens joined by an underscore. When words can be replaced by multiple phrases, longer phrases take priority, and then more frequent phrases take second priority.

Finally, we discard sentences with less than 5 words. We truncate sentences more than 20 words.

Luntz-esque: estate tax, death tax, capitalism, free market, undocumented, illegal aliens, foreign trade, international trade, public option, governmentrun, political speech, campaign spending, cut taxes, trickledown, **Random (CR):** cerro, brownfields, redtape, soon as possible, implicit, sup, habits, granted, personality, luis, internationally, itemize, fidel castro, centralize, restraint, pleadings, amendment before us, child custody protection, cheney, illegal aliens, **Random (PN):** reigniting, hurst, see happen, wandering, wp, conveying, obama obama, global politics, really serious, faggot, permanent normal, syrian observatory, native american, strength among, orbiting, protege, exclaimed, tunis, snopes staff, administration also, **High Partisan (CR):** the usa patriot act, mining, patterns, public safety, gorge, spills, wall street, joliet, bridges, tax code, registrants, freedom of speech, compensatory time, college education, shelter, hunger, oil companies, scourge, somalia, traders, **High Partisan (PN):** mrs. romney, pesticides, zionists, u.s. support, pacific northwest, economics defense, light bulbs, east asian, burton, smog, abdel fattah, banksters, work requirements, greenhouse gases, duggars, nigeria security, bolling, geopolitics, teng, newsom said

Table 7: Sample words from each of our test sets as described in §5.2.

C Preprocessing Procedures for Partisan News

Kiesel et al. (2019) includes 600k articles for train and 150k articles for validation, each labeled with a 5-way partisanship by their publisher. We only train on their validation set because it is comparable in size with Congressional Record and it requires less data cleaning. We discard duplicate sentences, and the rest of the processing pipeline is the same as the Congressional Record.

As mentioned in the main paper, we find the

corpus-given "left" vs. "left-center" and "right" and "right-center" labels are prone to artifacts of particular publishers. For example, many foreign policy related phrases dominate the "right-center" category simply because the publisher *Foreign Affairs* is labeled as "right-center", but this distinction is unsupported in ground truth. Therefore, we collapse "left-center" and "left" as one class, and we collapse "right-center" and "right" as one class.

D Grounding Bill Titles and Topics

We first filter out bills that are mentioned less than 3 times in its corresponding two-year congressional session. The vast majority of bills are only mentioned one time (when they were introduced) or twice (often a bipartisanship poster-child co-sponsor repeats the spiel.)

After manual inspection, we define three speeches after the bill mentioned speech as context speeches and thus assigned the same denotation label (bill or topic) as the bill mentioned speech. Statistics of bill mentioned for each congressional session is summarized in Table 8. Subsequent tables show examples of bill topics, their frequency, and example bill mentioned speeches.

Session	Bills Scraped	Bill Title RegEx Matches	Bills with $\geqslant$ 3 mentions	Speeches with those Bills Mentioned	Num. Sentences
97	1471	539	43	464	20372
98	1633	688	51	665	33242
99	1895	360	45	273	16128
100	2092	440	47	358	18376
101	2633	805	82	684	35903
102	2778	626	58	503	26944
103	2261	443	42	325	16500
104	2120	548	46	440	21664
105	2587	1174	97	931	51878
106	3421	1317	115	1033	64605
107	3225	1007	92	752	44901
108	3039	688	75	436	26783
109	3363	817	62	616	31838
110	3928	1052	102	865	41601
111	3714	868	73	740	36026

Table 8: Corpus with regular expression search for bill titles.

Example Topic	Example Bill Short Titles
Health	National Diabetes Act
	Medical Devices Safety Act
	Emergency Medical Services Systems Act
Education	Women's Educational Equity Act
	Elementary and Secondary Drug Abuse Eradication Act
	Community Education Development Act
Government Operations and Politics	Nonpartisan Commission on Campaign Reform Act
	Government in the Sunshine Act
	Congressional Disclosure of Income Act

Table 9: Example bill topics.

Freq. per sentence	Topic
45815	Health
38339	Education
33993	Government operations and politics
33462	Labor and employment
28392	Taxation
26435	Crime and law enforcement
24204	Finance and financial sector
22273	Commerce
21451	Transportation and public works
20865	International affairs
18560	Public lands and natural resources
17369	Armed forces and national security
16376	Economics and public finance
15660	Law
14702	Environmental protection
14472	Foreign trade and international finance
13353	Families
11752	Energy
11741	Agriculture and food
10512	Science, technology, communications
7050	Civil rights and liberties, minority issues
6599	Housing and community development
6066	Social welfare
5019	Native Americans
3582	Water resources development
3566	Commemorations
3457	Emergency management
2160	Immigration
2116	Congress
1640	Animals
1559	Sports and recreation
1303	Day care
552	Arts, culture, religion
545	Awards, medals, prizes
473	Public works
389	Federal aid to handicapped services
344	Monuments and memorials
241	Administrative procedure
157	Arms control
123	Mines and mineral resources
94	Fires

Table 10: CR TOPIC

Example Speeches with Bill Mentions
"Auto Stock for Every Taxpayer Act" These companies did all of this when the main company decided that the subsidiary was not consistent with the core business. That is what we should do with General Motors–give taxpayers its shares and get General Motors back in the marketplace where it belongs. This idea is fast. it is simple. and it creates a market for the shares... I ask unanimous consent to have printed in the RECORD newspaper articles supporting the Auto Stock for Every Taxpayer Act.
"Radioactive Import Deterrence Act" Mr. Speaker. the Radioactive Import Deterrence Act is a bipartisan bill that would ban the importation of lowlevel radioactive waste unless the President provides a waiver. Lowlevel radioactive waste is generated by medical facilities. university research labs. and utility companies. This waste is generated all over the United States. but finding permanent disposal sites has proven difficult. Currently. 36 States and the District of Columbia have only one approved site to store all the waste generated by those industries. That site is located in Utah...
"Help Find the Missing Act" I yield myself such time as I may consume. Madam Speaker. the Help Find the Missing Act. or Billys Law. will help families of missing persons find their loved ones by strengthening Federal databases about missing persons and unidentified remains. Every year. tens of thousands of Americans go missing and are never found. In the subcommittee we heard moving testimony from Ms. Janice Smolinski. whose son. Billy. went missing in 2004. While she has not found her son. she has dedicated her life to improving the system for others. including highlighting the need to strengthen and expand access to our missing persons databases. I thank her for her dedication to this worthy cause...
"Emergency Aid to American Survivors of the Haiti Earthquake Act" Madam Speaker. I yield myself such time as I may consume. I rise in support of this Senate bill. S. 2949. As Representative MCDERMOTT described. it will provide assistance to thousands of Americans returning from Haiti following the devastating January 12 earthquake there. Let me reiterate that we are helping American citizens with this legislation. The bill. entitled Emergency Aid to American Survivors of the Haiti Earthquake Act. will ensure that State and local governments and charitable agencies on the ground in Florida...
"Enhanced Oversight of State and Local Economic Recovery Act" Mr. Speaker. I rise to thank my colleagues for favorable consideration of H.R. 2182. the Enhanced Oversight of State and Local Economic Recovery Act. I was pleased to cosponsor this legislation. which was introduced by the chairman of the Oversight and Government Reform Committee. At a hearing of that committee. we learned that dedicated oversight funding for State and local governments could improve oversight of money appropriated through the American Recovery and Reinvestment Act...
"Veterans Dog Training Therapy Act" I yield myself such time as I may consume. Madam Speaker. I rise today in support of H.R. 3885. the Veterans Dog Training Therapy Act. I want to thank the ranking member of the Health Subcommittee. Congressman BROWN from South Carolina. for bringing us this legislation. Madam Speaker. we all recognize how damaging the invisible wounds of war can be. The need for effective treatments for posttraumatic stress disorder and for other conditions. such as depression and substance abuse. is apparent. I think. to all Americans. This act recognizes and meets this need by exploring an innovative and promising new form of treatment. using the training of service dogs as a therapeutic medium...
"Prevent Deceptive Census Look Alike Mailings Act" Mr. Speaker. entering its 23rd decade. the U.S. Census is the longest running national census in the world. Our founders wrote it into the Constitution. because taking a fair count is an essential part of fair government. A comprehensive. accurate Census helps ensure that our common resources are distributed where they are most needed. so that our communities can get the roads. schools. and police protection that they need. Theres nothing partisan about that goal. Unfortunately. some groups have set out to deceive Americans by disguising their own private mailings as Census documents...

Table 11: Seven random samples of bill mentions from the 111th Congress. Speeches truncated to fit the table.