Connotation Frames: A Data-Driven Investigation

Hannah Rashkin **Sameer Singh** **Yejin Choi**
Computer Science & Engineering
University of Washington
{hrashkin, sameer, yejin}@cs.washington.edu

ACL 2016

担当：乾健太郎（東北大学）

Connotation flames

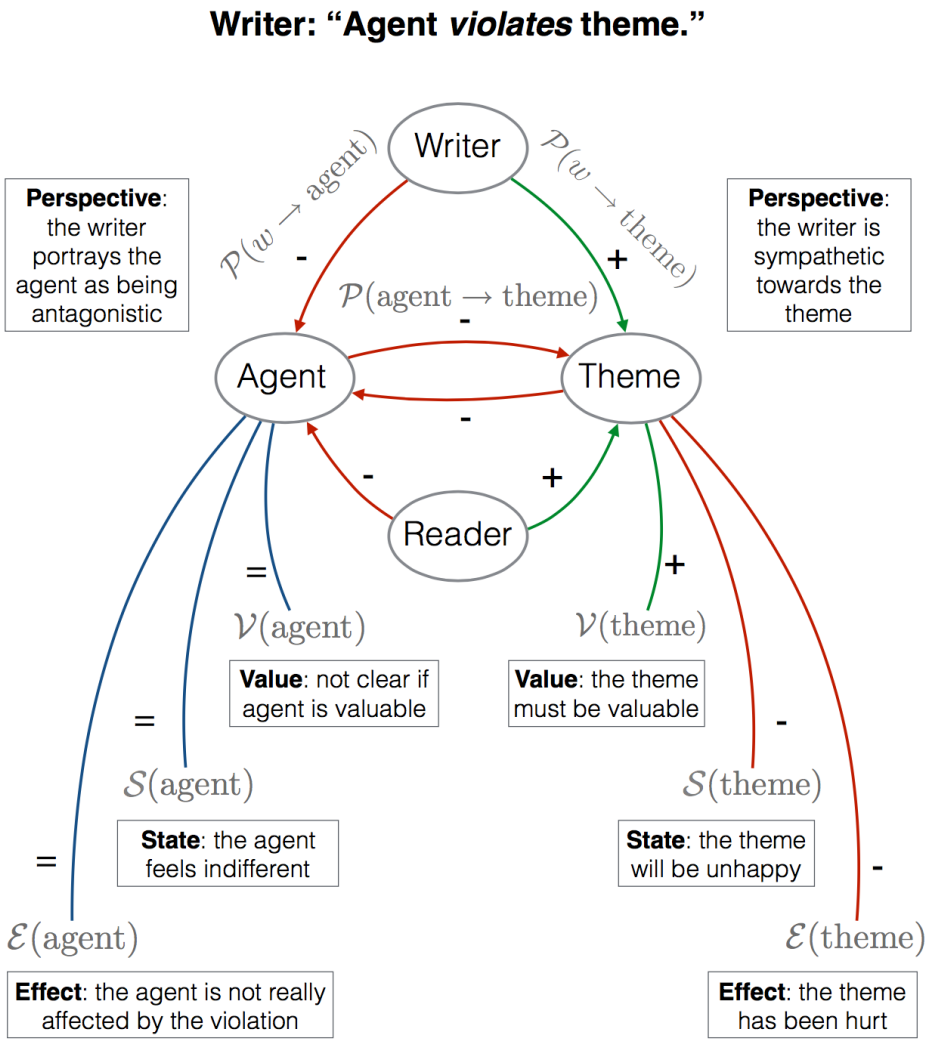


Figure 1: An example connotation frame of "violate" as a set of typed relations: perspective $\mathcal{P}(x \rightarrow y)$, effect $\mathcal{E}(x)$, value $\mathcal{V}(x)$, and mental state $\mathcal{S}(x)$.

x violated y のコノテーション

(1) writer's perspective: the writer is projecting x as an "antagonist" and y as a "victim", eliciting negative perspective from readers toward x (i.e., blaming x) and positive perspective toward y (i.e., sympathetic or supportive toward y).

(2) entities' perspective: y most likely feels negatively toward x as a result of being violated.

(3) effect: something bad happened to y.

(4) value: y is something valuable, since it does not make sense to violate something worthless. In other words, the writer is presupposing a positive value of y as a fact.

(5) mental state: y is most likely unhappy about the outcome.[1]

Contributions

1. New formalism, model, and annotated dataset for studying connotation frames from large-scale natural language data and statistics
2. New data-driven insights into the dynamics among different typed relations within each frame
3. Analytic study showing the potential use of connotation frames for analyzing subtle biases in journalism.

Connotation flames: Formalism

Given a predicate v, we define a connotation frame $\mathcal{F}(v)$ as a collection of typed relations and their polarity assignments: (i) **perspective** $\mathcal{P}^v(a_i \rightarrow a_j)$: a directed sentiment from the entity a_i to the entity a_j, (ii) **value** $\mathcal{V}^v(a_i)$: whether a_i is presupposed to be valuable, (iii) **effect** $\mathcal{E}^v(a_i)$: whether the event denoted by the predicate v is good or bad for the entity a_i, and (iv) **mental state** $\mathcal{S}^v(a_i)$: the likely mental state of the entity a_i as a result of the event.

We assume that each typed relation can have one of the three connotative polarities $\in \{+, -, =\}$, i.e., positive, negative, or neutral. Our goal in this paper is to focus on the general connotation of the predicate considered out of context. We leave contextual interpretation of connotation as future work.

Verb	Subset of Typed Relations		Example Sentences
suffer	$\mathcal{P}(w \rightarrow \text{agent}) = +$ $\mathcal{P}(w \rightarrow \text{theme}) = -$ $\mathcal{P}(\text{agent} \rightarrow \text{theme}) = -$	$\mathcal{E}(\text{agent}) = -$ $\mathcal{V}(\text{agent}) = +$ $\mathcal{S}(\text{agent}) = -$	The story begins in Illinois in 1987, when *a 17-year-old girl* **suffered** *a botched abortion.*
guard	$\mathcal{P}(w \rightarrow \text{agent}) = +$ $\mathcal{P}(w \rightarrow \text{theme}) = +$ $\mathcal{P}(\text{agent} \rightarrow \text{theme}) = +$	$\mathcal{E}(\text{theme}) = +$ $\mathcal{V}(\text{theme}) = +$ $\mathcal{S}(\text{theme}) = +$	In August, *marshals* **guarded** *25 clinics* in 18 cities.
uphold	$\mathcal{P}(w \rightarrow \text{theme}) = +$ $\mathcal{P}(\text{agent} \rightarrow \text{theme}) = +$	$\mathcal{E}(\text{theme}) = +$ $\mathcal{V}(\text{theme}) = +$	A hearing is scheduled to make a decision on whether to **uphold** *the clinic's suspension.*

Table 1: Example typed relations (perspective $\mathcal{P}(x \rightarrow y)$, effect $\mathcal{E}(x)$, value $\mathcal{V}(x)$, and mental state $\mathcal{S}(x)$).

Data collection by crowdsourcing

- Amazon Mechanical Turk
- For each of 1000 most frequent verbs (NYT)
  - 5 generic sentences (subj-verb-obj from Google n-gram)
  - 3 annotators (for each sent)
- *"How do you think the Subject feels about the event described in this sentence?"*
  - 5 choices: pos, pos-neu, neu, neg-neu, neg
- Average Krippendorff alpha is 0.25, indicating stronger than random agreement
- NC agreement is pretty high
- Some aspects are highly skewed

Aspect	% Agreement		Distribution	
	Strict	NC	% +	% -
$\mathcal{P}(w \rightarrow o)$	75.6	95.6	36.6	4.6
$\mathcal{P}(w \rightarrow s)$	76.1	95.5	47.1	7.9
$\mathcal{P}(s \rightarrow o)$	70.4	91.9	45.8	5.0
$\mathcal{E}(o)$	52.3	94.6	50.3	20.24
$\mathcal{E}(s)$	53.5	96.5	45.1	4.7
$\mathcal{V}(o)$	65.2	-	78.64	2.7
$\mathcal{V}(s)$	71.9	-	90.32	1.4
$\mathcal{S}(o)$	79.9	98.0	12.8	14.5
$\mathcal{S}(s)$	70.4	92.5	50.72	8.6

Table 4: Label Statistics: % Agreement refers to pairwise inter-annotator agreement. The strict agreement counts agreement over 3 classes ("positive or neutral" was counted as agreeing with either + or neutral), while non-conflicting (NC) agreement also allows agreements between neutral and -/+ (no direct conflicts). Distribution shows the final class distribution of -/+ labels created by averaging annotations.

[7] We take the average to obtain scalar value between $[-1., 1.]$ for each aspect of a verb's connotation frame. For simplicity, we cutoff the ranges of negative, neutral and positive polarities as $[-1, -0.25)$, $[-0.25, 0.25]$ and $(0.25, 1]$, respectively.

Contributions

1. New formalism, model, and annotated dataset for studying connotation frames from large-scale natural language data and statistics
2. New data-driven insights into the dynamics among different typed relations within each frame
3. Analytic study showing the potential use of connotation frames for analyzing subtle biases in journalism.

Dynamics over typed relations

Polarity assignments of typed relations are interdependent

Perspective Triad: If A is positive towards B, and B is positive towards C, then we expect A is also positive towards C. Similar dynamics hold for the negative case. $$\mathcal{P}_{w \to a_1} = \neg \left(\mathcal{P}_{w \to a_2} \oplus \mathcal{P}_{a_1 \to a_2} \right)$$
Perspective – Effect: If a predicate has a positive effect on the Subject, then we expect that the interaction between the Subject and Object was positive. Similar dynamics hold for the negative case and for other perspective relations. $$\mathcal{E}_{a_1} = \mathcal{P}_{a_2 \to a_1}$$
Perspective – Value: If A is presupposed as valuable, then we expect that the writer also views A positively. Similar dynamics hold for the negative case. $$\mathcal{V}_{a_1} = \mathcal{P}_{w \to a_1}$$
Effect – Mental State: If the predicate has a positive effect on A, then we expect that A will gain a positive mental state. Similar dynamics hold for the negative case. $$\mathcal{S}_{a_1} = \mathcal{E}_{a_1}$$

Table 3: Potential Dynamics among Typed Relations: we propose models that parameterize these dynamics using log-linear models (frame-level model in §3).

Modeling connotation frames

frame level classifier

We define $\mathbf{Y}_i := \{\mathcal{P}_{wo}, \mathcal{P}_{ws}, \mathcal{P}_{so}, \mathcal{E}_o, \mathcal{E}_s, \mathcal{V}_o, \mathcal{V}_s, \mathcal{S}_o, \mathcal{S}_s\}$ as the set of relational aspects for the i^{th} verb. The factor graph for $\mathbf{Y}_i$, is illustrated in Figure 2, and we will describe the factor potentials in more detail in the rest of this section. The probability of an assignment of polarities to the nodes in $\mathbf{Y}_i$ is:

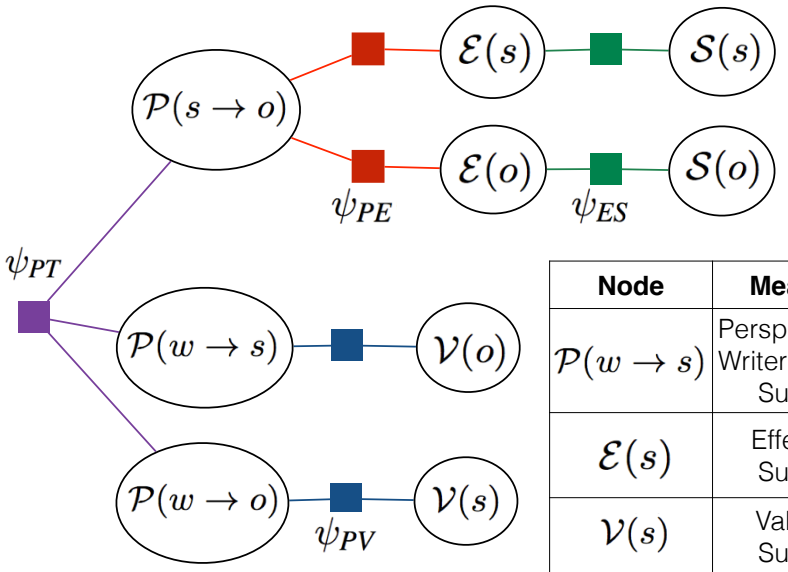
$$P(\mathbf{Y}_i) \propto \psi_{\text{PV}}(\mathcal{P}_{ws}, \mathcal{V}_s)\, \psi_{\text{PV}}(\mathcal{P}_{wo}, \mathcal{V}_o)\, \psi_{\text{PE}}(\mathcal{P}_{so}, \mathcal{E}_s)\, \psi_{\text{PE}}(\mathcal{P}_{so}, \mathcal{E}_o)\, \psi_{\text{ES}}(\mathcal{E}_s, \mathcal{S}_s)\, \psi_{\text{ES}}(\mathcal{E}_o, \mathcal{S}_o)\, \psi_{\text{PT}}(\mathcal{P}_{wo}, \mathcal{P}_{ws}, \mathcal{P}_{so}) \prod_{y \in \mathbf{Y}_i} \psi_{emb}(y)$$

$$\psi_{emb}(\mathcal{V}_o) = e^{w_{\mathcal{V}_o} \cdot f(\mathcal{V}_o)}$$

one-hot feature vector (+,,or =) representing the results of aspect-level classifier

MaxEnt classifier to predict the aspect label for a given 300 dimensional word-embedding

$$\psi_{PT}(\mathcal{P}_{wo}, \mathcal{P}_{ws}, \mathcal{P}_{so}) = e^{w_{PT} \cdot f(\mathcal{P}_{wo}, \mathcal{P}_{ws}, \mathcal{P}_{so})}$$



Node	Meaning
$\mathcal{P}(w \to s)$	Perspective of Writer towards Subject
$\mathcal{E}(s)$	Effect on Subject
$\mathcal{V}(s)$	Value of Subject
$\mathcal{S}(s)$	Mental State of Subject

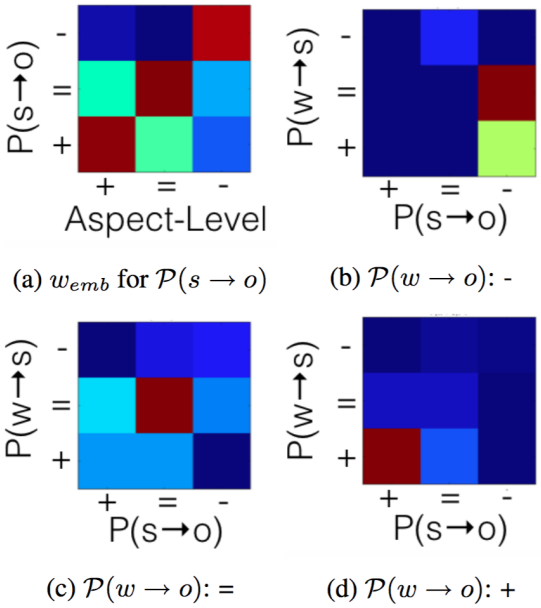
Figure 2: A factor graph for predicting the polarities of the typed relations that define a connotation frame for a given verb predicate. The factor graph also includes unary factors (ψ_{emb}), which we left out for brevity.

Experiments

- Annotated verbs divided into training, dev, and held-out test sets of equal size (300 verbs each)
- Aspect-level and frame-level models consistently outperform baselines (3-NN, GRAPH PROP)
- Frame-level model makes a small improvement

Algorithm	Acc.	Avg F_1
Graph Prop	58.81	41.46
3-nn	63.71	47.30
Aspect-Level	67.93	53.17
Frame-Level	68.26	53.50

Table 6: Performance on the test set. Results are averaged over the different aspects.



(a) w_{emb} for $\mathcal{P}(s \rightarrow o)$ (b) $\mathcal{P}(w \rightarrow o)$: - (c) $\mathcal{P}(w \rightarrow o)$: = (d) $\mathcal{P}(w \rightarrow o)$: +

Figure 3: Learned weights of embedding factor for the perspective of subject to object and the weights the perspective triad (PT) factor. Red is for weights that are more positive, whereas blue are more neg-

Contributions

1. New formalism, model, and annotated dataset for studying connotation frames from large-scale natural language data and statistics
2. New data-driven insights into the dynamics among different typed relations within each frame
3. Analytic study showing the potential use of connotation frames for analyzing subtle biases in journalism.

Data analytics of media bias

KBA Stream Corpus 2014

- 30 news sources indicated as exhibiting liberal or conservative leanings
- http://trec-kba.org/kba-stream-corpus-2014.shtml

Estimating entity polarities

- Selected 70 million news articles
- Extracted 1.2 billion unique tuples of the form (url,subj,verb,obj,count)
- Measure entity-to-entity sentiment using Connotation Frames

Observations

- Democrats positive: “nancy pelosi”, “unions”, “gun control”, etc.
- Republicans positive: “the pipeline”, “gop leaders”, “budget cuts”, etc.

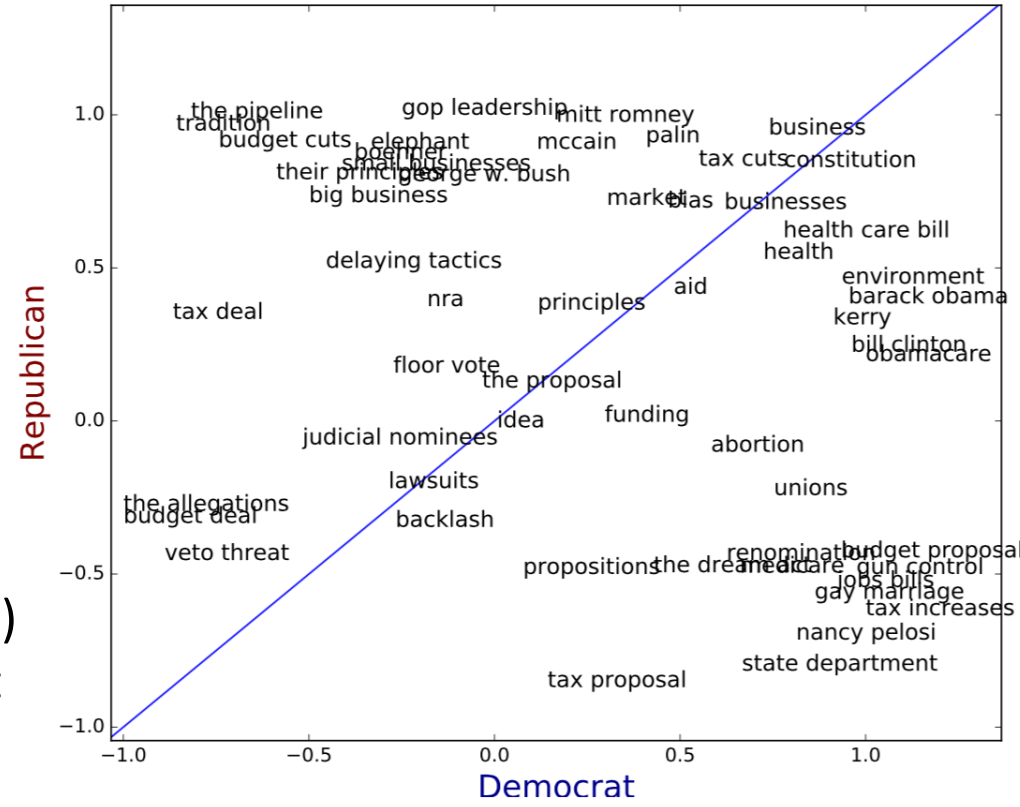


Figure 4: Average sentiment of Democrats and Republicans (as subjects) to selected nouns (as their objects), aggregated over a large corpus using the

Related work

- Frame semantics (Baker+98; Palmer+05)
  - Only denotational meanings
- Sentiment analysis
  - implied sentiment analysis (Feng+13; Greene+09)
  - opinion implicature (Deng&Wiebe14)
  - opinion role induction (Wiegand&Ruppenhofer15)
  - effect analysis (Choi&Wiebe14)

  → This work organizes various aspects of the connotative information into coherent frames
- Media bias, etc.
  - modeling framing (Greene&Resnik09; Hasan&Ng13)
  - biased language (Recasens+13)
  - ideology detection (Yano+10)

  → Connotation frame lexicon will be useful for them