

Shedding (a Thousand Points of) Light on Biased Language

Tae Yano

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA
taey@cs.cmu.edu

Philip Resnik

Department of Linguistics and UMIACS
University of Maryland
College Park, MD 20742, USA
resnik@umiacs.umd.edu

Noah A. Smith

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA
nasmith@cs.cmu.edu

Abstract

This paper considers the linguistic indicators of bias in political text. We used Amazon Mechanical Turk judgments about sentences from American political blogs, asking annotators to indicate whether a sentence showed bias, and if so, in which political direction and through which word tokens. We also asked annotators questions about their own political views. We conducted a preliminary analysis of the data, exploring how different groups perceive bias in different blogs, and showing some lexical indicators strongly associated with perceived bias.

1 Introduction

Bias and framing are central topics in the study of communications, media, and political discourse (Scheufele, 1999; Entman, 2007), but they have received relatively little attention in computational linguistics. What are the linguistic indicators of bias? Are there lexical, syntactic, topical, or other clues that can be computationally modeled and automatically detected?

Here we use Amazon Mechanical Turk (MTurk) to engage in a systematic, empirical study of linguistic indicators of bias in the political domain, using text drawn from political blogs. Using the MTurk framework, we collected judgments connected with the two dominant schools of thought in American politics, as exhibited in single sentences. Since no one person can claim to be an unbiased judge of political bias in language, MTurk is an attractive framework that lets us measure perception of bias across a population.

2 Annotation Task

We drew sentences from a corpus of American political blog posts from 2008. (Details in Section 2.1.) Sentences were presented to participants one at a time, without context. Participants were asked to judge the following (see Figure 1 for interface design):

- To what extent a sentence or clause is biased (none, somewhat, very);
- The nature of the bias (very liberal, moderately liberal, moderately conservative, very conservative, biased but not sure which direction); and
- Which words in the sentence give away the author’s bias, similar to “rationale” annotations in Zaidan et al. (2007).

For example, a participant might identify a moderate liberal bias in this sentence,

Without Sestak’s challenge, we would have Specter, comfortably ensconced as a Democrat in name only.

adding checkmarks on the underlined words. A more neutral paraphrase is:

Without Sestak’s challenge, Specter would have no incentive to side more frequently with Democrats.

It is worth noting that “bias,” in the sense we are using it here, is distinct from “subjectivity” as that topic has been studied in computational linguistics. Wiebe et al. (1999) characterize subjective sentences as those that “are used to communicate the speaker’s evaluations, opinions, and speculations,” as distinguished from sentences whose primary intention is “to objectively communicate material that is factual to the reporter.” In contrast, a biased sentence reflects a “tendency or preference towards a particular perspective, ideology or result.”¹ A subjective sentence can be unbiased (*I think that movie was terrible*), and a biased sentence can purport to communicate factually (*Nationalizing our health care system*

¹<http://en.wikipedia.org/wiki/Bias> as of 13 April, 2010.

is a point of no return for government interference in the lives of its citizens²).

In addition to annotating sentences, each participant was asked to complete a brief questionnaire about his or her own political views. The survey asked:

1. Whether the participant is a resident of the United States;
2. Who the participant voted for in the 2008 U.S. presidential election (Barack Obama, John McCain, other, decline to answer);
3. Which side of political spectrum he/she identified with for social issues (liberal, conservative, decline to answer); and
4. Which side of political spectrum he/she identified with for fiscal/economic issues (liberal, conservative, decline to answer).

This information was gathered to allow us to measure variation in bias perception as it relates to the stance of the annotator, e.g., whether people who view themselves as liberal perceive more bias in conservative sources, and vice versa.

2.1 Dataset

We extracted our sentences from the collection of blog posts in Eisenstein and Xing (2010). The corpus consists of 2008 blog posts gathered from six sites focused on American politics:

- American Thinker (conservative),³
- Digby (liberal),⁴
- Hot Air (conservative),⁵
- Michelle Malkin (conservative),⁶
- Think Progress (liberal),⁷ and
- Talking Points Memo (liberal).⁸

13,246 posts were gathered in total, and 261,073 sentences were extracted using WebHarvest⁹ and OpenNLP 1.3.0.¹⁰ Conservative and liberal sites are evenly represented (130,980 sentences from conservative sites, 130,093 from liberal sites). OpenNLP was also used for tokenization.

²Sarah Palin, http://www.facebook.com/note.php?note_id=113851103434, August 7, 2009.

³<http://www.americanthinker.com>

⁴<http://digbysblog.blogspot.com>

⁵<http://hotair.com>

⁶<http://michellemalkin.com>

⁷<http://thinkprogress.org>

⁸<http://www.talkingpointsmemo.com>

⁹<http://web-harvest.sourceforge.net>

¹⁰<http://opennlp.sourceforge.net>

Liberal	Conservative
thinkprogress org	exit question
video thinkprogress	hat tip
et rally	ed lasky
org 2008	hot air
gi bill	tony rezko
wonk room	ed morrissey
dana perino	track record
phil gramm	confirmed dead
senator mccain	american thinker
abu ghraib	illegal alien

Table 1: Top ten “sticky” partisan bigrams for each side.

2.2 Sentence Selection

To support exploratory data analysis, we sought a diverse sample of sentences for annotation, but we were also guided by some factors known or likely to correlate with bias. We extracted sentences from our corpus that matched at least one of the categories below, filtering to keep those of length between 8 and 40 tokens. Then, for each category, we first sampled 100 sentences without replacement. We then randomly extracted sentences up to 1,100 from the remaining pool. We selected the sentences this way so that the collection has variety, while including enough examples for individual categories. Our goal was to gather at least 1,000 annotated sentences; ultimately we collected 1,041. The categories are as follows.

“Sticky” partisan bigrams. One likely indicator of bias is the use of terms that are particular to one side or the other in a debate (Monroe et al., 2008). In order to identify such terms, we independently created two lists of “sticky” (i.e., strongly associated) bigrams in liberal and conservative subcorpora, measuring association using the log-likelihood ratio (Dunning, 1993) and omitting bigrams containing stopwords.¹¹ We identified a bigram as “liberal” if it was among the top 1,000 bigrams from the liberal blogs, as measured by strength of association, and was also not among the top 1,000 bigrams on the conservative side. The reverse definition yielded the “conservative” bigrams. The resulting liberal list contained 495 bigrams, and the conservative list contained 539. We then manually filtered cases that were clearly remnant HTML tags and other markup, arriving at lists of 433 and 535, respectively. Table 1 shows the strongest weighted bigrams.

As an example, consider this sentence (with a preceding sentence of context), which contains *gi bill*. There is no reason to think the bigram itself is inherently biased (in contrast to, for example, *death tax*, which we would

¹¹We made use of Pedersen’s *N*-gram Statistics Package (Banerjee and Pedersen, 2003).

perceive as biased in virtually any unquoted context), but we do perceive bias in the full sentence.

Their hard fiscal line softens in the face of American imperialist adventures. According to CongressDaily the Bush dogs are also whining because one of their members, Stephanie Herseht Sandlin, didn't get HER GI Bill to the floor in favor of Jim Webb's .

Emotional lexical categories. Emotional words might be another indicator of bias. We extracted four categories of words from Pennebaker's LIWC dictionary: Negative Emotion, Positive Emotion, Causation, and Anger.¹² The following is one example of a biased sentence in our dataset that matched these lexicons, in this case the Anger category; the match is in bold.

A bunch of **ugly** facts are nailing the biggest scare story in history.

The five most frequent matches in the corpus for each category are as follows.¹³

Negative Emotion: war attack* problem* numb* argu*

Positive Emotion: like well good party* secur*

Causation: how because lead* make why

Anger: war attack* argu* fight* threat*

Kill verbs. Greene and Resnik (2009) discuss the relevance of syntactic structure to the perception of sentiment. For example, their psycholinguistic experiments would predict that when comparing *Millions of people starved under Stalin* (inchoative) with *Stalin starved millions of people* (transitive), the latter will be perceived as more negative toward Stalin, because the transitive syntactic frame tends to be connected with semantic properties such as intended action by the subject and change of state in the object. "Kill verbs" provide particularly strong examples of such phenomena, because they exhibit a large set of semantic properties canonically associated with the transitive frame (Dowty, 1991). The study by Greene and Resnik used 11 verbs of killing and similar action to study the effect of syntactic "packaging" on perceptions of sentiment.¹⁴ We included membership on this list (in any morphological form) as a selection criterion, both because these verbs may be likely

¹²<http://www.liwc.net>. See Pennebaker et al. (2007) for detailed description of background theory, and how these lexicons were constructed. Our gratitude to Jamie Pennebaker for the use of this dictionary.

¹³Note that some LIWC lexical entries are specified as prefixes/stems, e.g. *ug1**, which matches *ugly uglier*, etc.

¹⁴The verbs are: *kill, slaughter, assassinate, shoot, poison, strangle, smother, choke, drown, suffocate*, and *starve*.

to appear in sentences containing bias (they overlap significantly with Pennebaker's Negative Emotion list), and because annotation of bias will provide further data relevant to Greene and Resnik's hypothesis about the connections among semantic properties, syntactic structures, and positive or negative perceptions (which are strongly connected with bias).

In our final 1,041-sentence sample, "sticky bigrams" occur 235 times (liberal 113, conservative 122), the lexical category features occur 1,619 times (Positive Emotion 577, Negative Emotion 466, Causation 332, and Anger 244), and "kill" verbs appear as a feature in 94 sentences. Note that one sentence often matches multiple selection criteria. Of the 1,041-sentence sample, 232 (22.3%) are from American Thinker, 169 (16.2%) from Digby, 246 (23.6%) from Hot Air, 73 (7.0%) from Michelle Malkin, 166 (15.9%) from Think Progress, and 155 (14.9%) from Talking Points Memo.

3 Mechanical Turk Experiment

We prepared 1,100 Human Intelligence Tasks (HITs), each containing one sentence annotation task. 1,041 sentences were annotated five times each (5,205 judgements total). One annotation task consists of three bias judgement questions plus four survey questions. We priced each HIT between \$0.02 and \$0.04 (moving from less to more to encourage faster completion). The total cost was \$212.¹⁵ We restricted access to our tasks to those who resided in United States and who had above 90% approval history, to ensure quality and awareness of American political issues. We also discarded HITs annotated by workers with particularly low agreement scores. The time allowance for each HIT was set at 5 minutes.

3.1 Annotation Results

3.1.1 Distribution of Judgments

Overall, more than half the judgments are "not biased," and the "very biased" label is used sparingly (Table 2). There is a slight tendency among the annotators to assign the "very conservative" label, although moderate bias is distributed evenly on both side (Table 3). Interestingly, there are many "biased, but not sure" labels, indicating that the annotators are capable of perceiving bias (or manipulative language), without fully decoding the intent of the author, given sentences out of context.

Bias	1	1.5	2	2.5	3
% judged	36.0	26.6	25.5	9.4	2.4

Table 2: Strength of perceived bias per sentence, averaged over the annotators (rounded to nearest half point). Annotators rate bias on a scale of 1 (no bias), 2 (some bias), and 3 (very biased).

¹⁵This includes the cost for the discarded annotations.

Label political leanings

Please read the following sentence.

Jason Horowitz sums up the problem succinctly : Obama wanted nothing to do with the Clintons.

Do you think this sentence shows the author's preference towards a particular position (bias) in politics? If so, to what extent does this sentence display political bias?

I don't see any bias
 Somewhat biased
 Very much biased

If there is a bias, to what extent is it liberal or conservative? Please check 'No obvious bias' if the sentence has no obvious bias.

Very liberal
 Moderately liberal
 I don't see any bias
 Moderately conservative
 Very conservative
 I think it is biased, but am not sure which side

1: Jason
 2: Horowitz
 3: sums
 4: up
 5: the
 6: problem
 7: succinctly
 8: :
 9: Obama
 10: wanted
 11: nothing
 12: to
 13: do
 14: with
 15: the
 16: Clintons.

Figure 1: HIT: Three judgment questions. We first ask for the strength of bias, then the direction. For the word-level annotation question (right), workers are asked to check the box to indicate the region which “give away” the bias.

Bias type	VL	ML	NB	MC	VC	B
% judged	4.0	8.5	54.8	8.2	6.7	17.9

Table 3: Direction of perceived bias, per judgment (very liberal, moderately liberal, no bias, moderately conservative, very conservative, biased but not sure which).

	Economic				
	L	M	C	NA	
Social	L	20.1	10.1	4.9	0.7
	M	0.0	21.9	4.7	0.0
	C	0.1	0.4	11.7	0.0
	NA	0.1	0.0	11.2	14.1

Table 4: Distribution of judgements by annotators' self-identification on social issues (row) and fiscal issue (column); {L, C, M, NA} denote liberal, conservative, moderate, and decline to answer, respectively.

3.1.2 Annotation Quality

In this study, we are interested in where the wisdom of the crowd will take us, or where the majority consensus on bias may emerge. For this reason we did not contrive a gold standard for “correct” annotation. We are, however, mindful of its overall quality—whether annotations have

reasonable agreement, and whether there are fraudulent responses tainting the results.

To validate our data, we measured the pair-wise Kappa statistic (Cohen, 1960) among the 50 most frequent workers¹⁶ and took the average over all the scores.¹⁷ The average of the agreement score for the first question is 0.55, and the second 0.50. Those are within the range of reasonable agreement for moderately difficult task. We also inspected per worker average scores for frequent workers¹⁸ and found one with consistently low agreement scores. We discarded all the HITs by this worker from our results. We also manually inspected the first 200 HITs for apparent frauds. The annotations appeared to be consistent. Often annotators agreed (many “no bias” cases were unanimous), or differed in only the degree of strength (“very biased” vs. “biased”) or specificity (“biased but I am not sure” vs. “moderately liberal”). The direction of bias, if specified, was very rarely inconsistent.

Along with the annotation tasks, we asked workers how we could improve our HITs. Some comments were

¹⁶258 workers participated; only 50 of them completed more than 10 annotations.

¹⁷Unlike traditional subjects for a user-annotation study, our annotators have not judged all the sentences considered in the study. Therefore, to compute the agreement, we considered only the case where two annotators share 20 or more sentences.

¹⁸We consider only those with 10 or more annotations.

insightful for our study (as well as for the interface design). A few pointed out that an impolite statement or a statement of negative fact is not the same as bias, and therefore should be marked separately from bias. Others mentioned that some sentences are difficult to judge out of context. These comments will be taken into account in future research.

4 Analysis and Significance

In the following section we report some of the interesting trends we found in our annotation results. We consider a few questions and report the answers the data provide for each.

4.1 Is a sentence from a liberal blog more likely be seen as liberal?

In our sample sentence pool, conservatives and liberals are equally represented, though each blog site has a different representation.¹⁹ We grouped sentences by source site, then computed the percentage representation of each site within each bias label; see Table 5. In the top row, we show the percentage representation of each group in overall judgements.

In general, a site yields more sentences that match its known political leanings. Note that in our annotation task, we did not disclose the sentence’s source to the workers. The annotators formed their judgements solely based on the content of the sentence. This result can be taken as confirming people’s ability to perceive bias within a sentence, or, conversely, as confirming our *a priori* categorizations of the blogs.

	at	ha	mm	db	tp	tpm
Overall	22.3	23.6	7.0	16.2	15.9	14.9
NB	23.7	22.3	6.1	15.7	17.0	15.3
VC	24.8	32.3	19.3	6.9	7.5	9.2
MC	24.4	33.6	8.0	8.2	13.6	12.2
ML	16.6	15.2	3.4	21.1	22.9	20.9
VL	16.7	9.0	4.3	31.0	22.4	16.7
B	20.1	25.4	7.2	19.5	12.3	13.7

Table 5: Percentage representation of each site within bias label pools from question 2 (direction of perceived bias): very liberal, moderately liberal, no bias, moderately conservative, very conservative, biased but not sure which. Rows sum to 100. Bold-face indicates rates higher than the site’s overall representation in the pool.

4.2 Does a liberal leaning annotator see more conservative bias?

In Table 5, we see that blogs are very different from each other in terms of the bias annotators perceive in their lan-

¹⁹Posts appear on different sites at different rates.

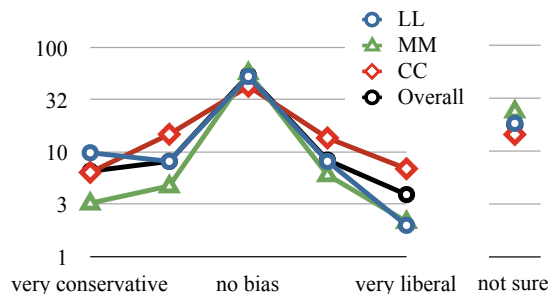


Figure 2: Distribution of bias labels (by judgment) for social and economic liberals (LL), social and economic moderates (MM), and social and economic conservatives (CC), and overall. Note that this plot uses a logarithmic scale, to tease apart the differences among groups.

guage. In general, conservative sites seemingly produced much more identifiable partisan bias than liberal sites.²⁰

This impression, however, might be an artifact of the distribution of the annotators’ own bias. As seen in Table 4, a large portion of our annotators identified themselves as liberal in some way. People might call a statement biased if they disagree with it, while showing leniency toward hyperbole more consistent with their opinions.

To answer this question, we break down the judgement labels by the annotators’ self-identification, and check the percentage of each bias type within key groups (see Figure 2). In general, moderates perceive less bias than partisans (another useful reality check, in the sense that this is to be expected), but conservatives show a much stronger tendency to label sentences as biased, in *both* directions. (We caution that the underrepresentation of self-identifying conservatives in our worker pool means that only 608 judgments from 48 distinct workers were used to estimate these statistics.) Liberals in this sample are less balanced, perceiving conservative bias at double the rate of liberal bias.

4.3 What are the lexical indicators of perceived bias?

For a given word type w , we calculate the frequency that it was marked as indicating bias, normalized by its total number of occurrences. To combine the judgments of different annotators, we increment w ’s count by k/n whenever k judgments out of n marked the word as showing bias. We perform similar calculations with a restriction to liberal and conservative judgments on the sentence as a

²⁰Liberal sites cumulatively produced 64.9% of the *moderately liberal* bias label and 70.1 % of *very liberal*, while conservative sites produced 66.0% of *moderately conservative* and 76.4% of *very conservative*, respectively.

Overall		Liberal		Conservative		Not Sure Which	
bad	0.60	Administration	0.28	illegal	0.40	pass	0.32
personally	0.56	Americans	0.24	Obama’s	0.38	bad	0.32
illegal	0.53	woman	0.24	corruption	0.32	sure	0.28
woman	0.52	single	0.24	rich	0.28	blame	0.28
single	0.52	personally	0.24	stop	0.26	they’re	0.24
rich	0.52	lobbyists	0.23	tax	0.25	happen	0.24
corruption	0.52	Republican	0.22	claimed	0.25	doubt	0.24
Administration	0.52	union	0.20	human	0.24	doing	0.24
Americans	0.51	torture	0.20	doesn’t	0.24	death	0.24
conservative	0.50	rich	0.20	difficult	0.24	actually	0.24
doubt	0.48	interests	0.20	Democrats	0.24	exactly	0.22
torture	0.47	doing	0.20	less	0.23	wrong	0.22

Table 6: Most strongly biased words, ranked by relative frequency of receiving a bias mark, normalized by total frequency. Only words appearing five times or more in our annotation set are ranked.

whole. Top-ranked words for each calculation are shown in Table 6.

Some of the patterns we see are consistent with what we found in our automatic method for proposing biased bigrams. For example, the bigrams tended to include terms that refer to members or groups on the opposing side. Here we find that *Republican* and *Administration* (referring in 2008 to the Bush administration) tends to show liberal bias, while *Obama’s* and *Democrats* show conservative bias.

5 Discussion and Future Work

The study we have conducted here represents an initial pass at empirical, corpus-driven analysis of bias using the methods of computational linguistics. The results thus far suggest that it is possible to automatically extract a sample that is rich in examples that annotators would consider biased; that naïve annotators can achieve reasonable agreement with minimal instructions and no training; and that basic exploratory analysis of results yields interpretable patterns that comport with prior expectations, as well as interesting observations that merit further investigation.

In future work, enabled by annotations of biased and non-biased material, we plan to delve more deeply into the linguistic characteristics associated with biased expression. These will include, for example, an analysis of the extent to which explicit “lexical framing” (use of partisan terms, e.g., Monroe et al., 2008) is used to convey bias, versus use of more subtle cues such as syntactic framing (Greene and Resnik, 2009). We will also explore the extent to which idiomatic usages are connected with bias, with the prediction that partisan “memes” tend to be more idiomatic than compositional in nature.

In our current analysis, the issue of subjectivity was not directly addressed. Previous work has shown that opin-

ions are closely related to subjective language (Pang and Lee, 2008). It is possible that asking annotators about sentiment while asking about bias would provide a deeper understanding of the latter. Interestingly, annotator feedback included remarks that mere negative “facts” do not convey an author’s opinion or bias. The nature of subjectivity as a factor in bias perception is an important issue for future investigation.

6 Conclusion

This paper considered the linguistic indicators of bias in political text. We used Amazon Mechanical Turk judgments about sentences from American political blogs, asking annotators to indicate whether a sentence showed bias, and if so, in which political direction and through which word tokens; these data were augmented by a political questionnaire for each annotator. Our preliminary analysis suggests that bias can be annotated reasonably consistently, that bias perception varies based on personal views, and that there are some consistent lexical cues for bias in political blog data.

Acknowledgments

The authors acknowledge research support from HP Labs, help with data from Jacob Eisenstein, and helpful comments from the reviewers, Olivia Buzek, Michael Heilman, and Brendan O’Connor.

References

- Satanjeev Banerjee and Ted Pedersen. 2003. The design, implementation and use of the ngram statistics package. In *the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*.
- Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.
- David Dowty. 1991. Thematic Proto-Roles and Argument Selection. *Language*, 67:547–619.

- Ted Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74.
- Jacob Eisenstein and Eric Xing. 2010. The CMU 2008 political blog corpus. Technical report CMU-ML-10-101.
- Robert M. Entman. 2007. Framing bias: Media in the distribution of power. *Journal of Communication*, 57(1):163–173.
- Stephan Greene and Philip Resnik. 2009. More than words: Syntactic packaging and implicit sentiment. In *NAACL*, pages 503–511, June.
- Burt L. Monroe, Michael P. Colaresi, and Kevin M. Quinn. 2008. Fightin’ words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4):372–403, October.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.
- J.W. Pennebaker, C.K Chung, M. Ireland, A Gonzales, and R J. Booth, 2007. *The development and psychometric properties of LIWC2007*.
- Dietram A. Scheufele. 1999. Framing as a theory of media effects. *Journal of Communication*, 49(1):103–122.
- Janyce M. Wiebe, Rebecca F. Bruce, and Thomas P. O’Hara. 1999. Development and use of a gold standard data set for subjectivity classifications. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 246–253.
- Omar Zaidan, Jason Eisner, and Christine Piatko. 2007. Using “annotator rationales” to improve machine learning for text categorization. In *NAACL*, pages 260–267, April.