A Multidimensional Lexicon for Interpersonal Stancetaking

Umashanthi Pavalanathan
Georgia Institute of Technology
Atlanta, GA
umashanthi@gatech.edu

Jim Fitzpatrick
University of Pittsburgh
Pittsburgh, PA
jim.fitzpatrick@gmail.com

Scott F. Kiesling
University of Pittsburgh
Pittsburgh, PA
kiesling@pitt.edu

Jacob Eisenstein
Georgia Institute of Technology
Atlanta, GA
jacobe@gatech.edu

Abstract

The sociolinguistic construct of stancetaking describes the activities through which discourse participants create and signal relationships to their interlocutors, to the topic of discussion, and to the talk itself. Stancetaking underlies a wide range of interactional phenomena, relating to formality, politeness, affect, and subjectivity. We present a computational approach to stancetaking, in which we build a theoretically-motivated lexicon of stance markers, and then use multidimensional analysis to identify a set of underlying stance dimensions. We validate these dimensions intrinsically and extrinsically, showing that they are internally coherent, match pre-registered hypotheses, and correlate with social phenomena.

1 Introduction

What does it mean to be welcoming or standoffish, light-hearted or cynical? Such interactional styles are performed primarily with language, yet little is known about how linguistic resources are arrayed to create these social impressions. The sociolinguistic concept of *interpersonal stancetaking* attempts to answer this question, by providing a conceptual framework that accounts for a range of interpersonal phenomena, subsuming formality, politeness, and subjectivity (Du Bois, 2007).[1] This framework has been applied almost exclusively through qualitative methods, using close readings of individual texts or dialogs to uncover how language is used to position individuals with respect to their interlocutors and readers.

We attempt the first large-scale operationalization of stancetaking through computational methods. Du Bois (2007) formalizes stancetaking as a multi-dimensional construct, reflecting the relationship of discourse participants to (a) the audience or interlocutor; (b) the topic of discourse; (c) the talk or text itself. However, the multidimensional nature of stancetaking poses problems for traditional computational approaches, in which labeled data is obtained by relying on annotator intuitions about scalar concepts such politeness (Danescu-Niculescu-Mizil et al., 2013) and formality (Pavlick and Tetreault, 2016).

Instead, our approach is based on a theoretically-guided application of unsupervised learning, in the form of factor analysis, applied to lexical features. Stancetaking is characterized in large part by an array of linguistic features ranging from discourse markers such as *actually* to backchannels such as *yep* (Kiesling, 2009). We therefore first compile a lexicon of stance markers, combining prior lexicons from Biber and Finegan (1989) and the Switchboard Dialogue Act Corpus (Jurafsky et al., 1998). We then extend this lexicon to the social media domain using word embeddings. Finally, we apply multi-dimensional analysis of co-occurrence patterns to identify a small set of *stance dimensions*.

To measure the internal coherence (construct validity) of the stance dimensions, we use a word

[1] Stancetaking is distinct from the notion of *stance* which corresponds to a position in a debate (Walker et al., 2012). Similarly, Freeman et al. (2014) correlate phonetic features with the *strength* of such argumentative stances.

intrusion task (Chang et al., 2009) and a set of preregistered hypotheses. To measure the utility of the stance dimensions, we perform a series of extrinsic evaluations. A predictive evaluation shows that the membership of online communities is determined in part by the interactional stances that predominate in those communities. Furthermore, the induced stance dimensions are shown to align with annotations of politeness and formality.

Contributions We operationalize the sociolinguistic concept of stancetaking as a multidimensional framework, making it possible to measure at scale. Specifically,

- we contribute a lexicon of stance markers based on prior work and adapted to the genre of online interpersonal discourse;
- we group stance markers into latent dimensions;
- we show that these stance dimensions are internally coherent;
- we demonstrate that the stance dimensions predict and correlate with social phenomena.[2]

2 Related Work

From a theoretical perspective, we build on prior work on interactional meaning in language. Methodologically, our paper relates to prior work on lexicon-based analysis and contrastive studies of social media communities.

2.1 Linguistic Variation and Social Meaning

In computational sociolinguistics (Nguyen et al., 2016), language variation has been studied primarily in connection with macro-scale social variables, such as age (Argamon et al., 2007; Nguyen et al., 2013), gender (Burger et al., 2011; Bamman et al., 2014), race (Eisenstein et al., 2011; Blodgett et al., 2016), and geography (Eisenstein et al., 2010). This parallels what Eckert (2012) has called the "first wave" of language variation studies in sociolinguistics, which also focused on macro-scale variables.

More recently, sociolinguists have dedicated increased attention to situational and stylistic variation, and the *interactional meaning* that such variation can convey (Eckert and Rickford, 2001). This linguistic research can be aligned with computational efforts to quantify phenomena such as subjectivity (Riloff and Wiebe, 2003), sentiment (Wiebe et al., 2005), politeness (Danescu-Niculescu-Mizil et al., 2013), formality (Pavlick and Tetreault, 2016), and power dynamics (Prabhakaran et al., 2012). While linguistic research on interactional meaning has focused largely on qualitative methodologies such as discourse analysis (e.g., Bucholtz and Hall, 2005), these computational efforts have made use of crowdsourced annotations to build large datasets of, for example, polite and impolite text. These annotation efforts draw on the annotators' intuitions about the meaning of these sociolinguistic constructs.

Interpersonal stancetaking represents an attempt to unify concepts such as sentiment, politeness, formality, and subjectivity under a single theoretical framework (Jaffe, 2009; Kiesling, 2009). The key idea, as articulated by Du Bois (2007), is that stancetaking captures the speaker's relationship to (a) the topic of discussion, (b) the interlocutor or audience, and (c) the talk (or writing) itself. Various configurations of these three legs of the "stance triangle" can account for a range of phenomena. For example, epistemic stance relates to the speaker's certainty about what is being expressed, while affective stance indicates the speaker's emotional position with respect to the content (Ochs, 1993).

The framework of stancetaking has been widely adopted in linguistics, particularly in the discourse analytic tradition, which involves close reading of individual texts or conversations (Kärkkäinen, 2006; Keisanen, 2007; Precht, 2003; White, 2003). But despite its strong theoretical foundation, we are aware of no prior efforts to operationalize stancetaking at scale. Since annotators may not have strong intuitions about stance — in the way that they do about formality and politeness — we cannot rely on the annotation methodologies employed in prior work. We take a different approach, performing a multidimensional analysis of the distribution of likely stance markers.

2.2 Lexicon-based Analysis

Our operationalization of stancetaking is based on the induction of lexicons of stance markers. The lexicon-based methodology is related to earlier work from social psychology, such as the General Inquirer (Stone, 1966) and LIWC (Tausczik and Pennebaker, 2010). In LIWC, the basic categories were identified first, based on psychological

[2]Lexicons and stance dimensions are available at https://github.com/umashanthi-research/multidimensional-stance-lexicon

constructs (e.g., positive emotion, cognitive processes, drive to power) and syntactic groupings of words and phrases (e.g., pronouns, prepositions, quantifiers). The lexicon designers then manually contructed lexicons for each category, augmenting their intuitions by using distributional statistics to suggest words that may have been missed (Pennebaker et al., 2015). In contrast, we follow the approach of Biber (1991), using multidimensional analysis to identify latent groupings of markers based on co-occurrence statistics. We then use crowdsourcing and extrinsic comparisons to validate the coherence of these dimensions.

2.3 Multicommunity Studies

Social media platforms such as Reddit, Stack Exchange, and Wikia can be considered *multicommunity environments*, in that they host multiple subcommunities with distinct social and linguistic properties. Such subcommunities can be contrasted in terms of topics (Adamic et al., 2008; Hessel et al., 2014) and social networks (Backstrom et al., 2006). Our work focuses on Reddit, emphasizing community-wide differences in norms for interpersonal interaction. In the same vein, Tan and Lee (2015) attempt to characterize stylistic differences across subreddits by focusing on very common words and parts-of-speech; Tran and Ostendorf (2016) use language models and topic models to measure similarity across threads within a subreddit. One distinction of our approach is that the use of multidimensional analysis gives us interpretable dimensions of variation. This makes it possible to identify the specific interpersonal features that vary across communities.

3 Data

Reddit, one of the internet's largest social media platforms, is a collection of subreddits organized around various topics of interest. As of January 2017, there were more than one million subreddits and nearly 250 million users, discussing topics ranging from politics (*r/politics*) to horror stories (*r/nosleep*).[3] Although Reddit was originally designed for sharing hyperlinks, it also provides the ability to post original textual content, submit comments, and vote on content quality (Gilbert, 2013). Reddit's conversation-like threads are therefore well suited for the study of interpersonal social and linguistic phenomena.

[3] http://redditmetrics.com/

Subreddits	126,789
Authors	6,401,699
Threads	52,888,024
Comments	531,804,658

Table 1: Dataset size

For example, the following are two comments from the subreddit *r/malefashionadvice*, posted in response to a picture posted by a user asking for fashion advise.

U_1: *"I think the beard* ***looks pretty good****.* ***Definitely*** *not the goatee. Clean shaven is always the safe option."*

U_2: *"****Definitely*** *the beard. But keep it trimmed."*

The phrases in **bold** face are markers of stance, indicating a *evaluative* stance. The following example is a part of a thread in the subreddit *r/photoshopbattles* where users discuss an edited image posted by the original poster *OP*. The phrases in **bold** face are markers of stance, indicating an *involved* and *interactional* stance.

U_3: *"****Ha ha*** *awesome!"*
U_4: "are those..... furries?"
OP: *"****yes****, sir. They are!"*
U_4: *"****Oh cool****. That makes sense!"*

We used an archive of 530 million comments posted on Reddit in 2014, retrieved from the public archive of Reddit comments.[4] This dataset consists of each post's textual content, along with metadata that identifies the subreddit, thread, author, and post creation time. More statistics about the full dataset are shown in Table 1.

4 Stance Lexicon

Interpersonal stancetaking can be characterized in part by an array of linguistic features such as hedges (e.g., *might*, *kind of*), discourse markers (e.g., *actually*, *I mean*), and backchannels (e.g., *yep*, *um*). Our analysis focuses on these markers, which we collect into a lexicon.

4.1 Seed lexicon

We began with a seed lexicon of stance markers from Biber and Finegan (1989), who compiled an

[4] https://archive.org/details/2015_reddit_comments_corpus

extensive list by surveying dictionaries, previous studies on stance, and texts in several genres of English. This list includes certainty adverbs (e.g., *actually*, *of course*, *in fact*), affect markers (e.g., *amazing*, *thankful*, *sadly*), and hedges (e.g., *kind of*, *maybe*, *something like*) among other adverbial, adjectival, verbal, and modal markers of stance. In total, this list consists of 448 stance markers.

The Biber and Finegan (1989) lexicon is primarily based on written genres from the pre-social media era. Our dataset — like much of the recent work in this domain — consists of online discussions, which differ significantly from printed texts (Eisenstein, 2013). One difference is that online discussions contain a number of dialog act markers that are characteristic of spoken language, such as *oh yeah*, *nah*, *wow*. We accounted for this by adding 74 dialog act markers from the Switchboard Dialog Act Corpus (Jurafsky et al., 1998). The final seed lexicon consists of 517 unique markers, from these two sources. Note that the seed lexicon also includes markers that contain multiple tokens (e.g. *kind of*, *I know*).

4.2 Lexicon expansion

Online discussions differ not only from written texts, but also from spoken discussions, due to their use of non-standard vocabulary and spellings. To measure stance accurately, these genre differences must be accounted for. We therefore expanded the seed lexicon using automated techniques based on distributional statistics. This is similar to prior work on the expansion of sentiment lexicons (Hatzivassiloglou and McKeown, 1997; Hamilton et al., 2016).

Our lexicon expansion approach used word embeddings to find words that are distributionally similar to those in the seed set. We trained word embeddings on a corpus of 25 million Reddit comments and a vocabulary of 100K most frequent words on Reddit using the structured skip-gram models of both WORD2VEC (Mikolov et al., 2013) and WANG2VEC (Ling et al., 2015) with default parameters. The WANG2VEC method augments WORD2VEC by accounting for word order information. We found the similarity judgments obtained from WANG2VEC to be qualitatively more meaningful, so we used these embeddings to construct the expanded lexicon.[5]

[5]We used the following default parameters: 100 dimensions, a window size of five, a negative sampling size of ten, five-epoch iterations, and a sub-sampling rate of 10^{-4}.

Seed term	Expanded terms
(*Example seeds from Biber and Finegan (1989)*)	
significantly	considerably, substantially, dramatically
certainly	surely, frankly, definitely
incredibly	extremely, unbelievably, exceptionally
(*Example seeds from Jurafsky et al. (1998)*)	
nope	nah, yup, nevermind
great	fantastic, terrific, excellent

Table 2: Stance lexicon: seed and expanded terms.

To perform lexicon expansion, we constructed a dictionary of candidate terms, consisting of all unigrams that occur with a frequency rate of at least 10^{-7} in the Reddit comment corpus. Then, for each single-token marker in the seed lexicon, we identified all terms from the candidate set whose embedding has cosine similarity of at least 0.75 with respect to the seed marker.[6] Table 2 shows examples of seed markers and related terms we extracted from word embeddings. Through this procedure, we identified 228 additional markers based on similarity to items in the seed list from Biber and Finegan (1989), and 112 additional markers based on the seed list of dialog acts. In total, our stance lexicon contains 812 unique markers.

5 Linguistic Dimensions of Stancetaking

To summarize the main axes of variation across the lexicon of stance markers, we apply a multidimensional analysis (Biber, 1992) to the distributional statistics of stance markers across subreddit communities. Each dimension of variation can then be viewed as a spectrum, characterized by the stance markers and subreddits that are associated with the positive and negative extremes. Multidimensional analysis is based on singular value decomposition, which has been applied successfully to a wide range of problems in natural language processing and information retrieval (e.g., Landauer et al., 1998). While Bayesian topic models are an appealing alternative, singular value decomposition is fast and deterministic, with a minimal number of tuning parameters.

[6]We tried different thresholds on the similarity value and the corpus frequency, and the reported values were chosen based on the quality of the resulting related terms. This was done prior to any of the validations or extrinsic analyses described later in the paper.

5.1 Extracting Stance Dimensions

Our analysis is based on the co-occurrence of stance markers and subreddits. This is motivated by our interest in comparisons of the interactional styles of online communities within Reddit, and by the premise that these distributional differences reflect socially meaningful communicative norms. A pilot study applied the same technique to the co-occurrence of stance markers and individual authors, and the resulting dimensions appeared to be less stylistically coherent.

Singular value decomposition is often used in combination with a transformation of the co-occurrence counts by pointwise mutual information (Bullinaria and Levy, 2007). This transformation ensures that each cell in the matrix indicates how much more likely a stance marker is to co-occur with a given subreddit than would happen by chance under an independence assumption. Because negative PMI values tend to be unreliable, we use positive PMI (PPMI), which involves replacing all negative PMI values with zeros (Niwa and Nitta, 1994). Therefore, we obtain stance dimensions by applying singular value decomposition to the matrix constructed as follows:

$$X_{m,s} = \left(\log \frac{\Pr(\text{marker} = m, \text{subreddit} = s)}{\Pr(\text{marker} = m) \Pr(\text{subreddit} = s)} \right)_+ .$$

Truncated singular value decomposition performs the approximate factorization $X \approx U \Sigma V^\top$, where each row of the matrix U is a k-dimensional description of each stance marker, and each row of V is a k-dimensional description of each subreddit. We included the 7,589 subreddits that received at least 1,000 comments in 2014.

5.2 Results: Stance Dimensions

From the SVD analysis, we extracted the six principal latent dimensions that explain the most variation in our dataset.[7] The decision to include only the first six dimensions was based on the strength of the singular values corresponding to the dimensions. Table 3 shows the top five stance markers for each extreme of the six dimensions. The stance dimensions convey a range of concepts, such as involved versus informational language, narrative versus dialogue-oriented writing, standard versus non-standard variation, and positive versus negative affect. Figure 1 shows the distribution of subreddits along two of these dimensions.

[7] Similar to factor analysis, the top few dimensions of SVD explain the most variation, and tend to be most interpretable. A scree plot (Cattell, 1966) showed that the amount of variation explained dropped after the top six dimensions, and qualitative interpretation showed that the remaining dimension were less interpretable.

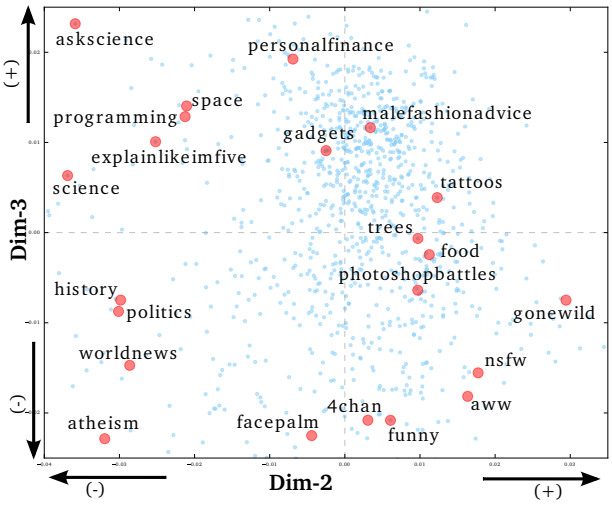

Figure 1: Mapping of subreddits in dimension two and dimension three, highlighting especially popular subreddits. Picture-oriented subreddits *r/gonewild* and *r/aww* map high on dimension two and low on dimension three, indicating involved and informal style of discourse. Subreddits dedicated for knowledge sharing discussions such as *r/askscience* and *r/space* map low on dimension two and high on dimension three indicating informational and formal style.

6 Construct Validity

Evaluating model output against gold-standard annotations is appropriate when there is some notion of a correct answer. As stancetaking is a multidimensional concept, we have taken an unsupervised approach. Therefore, we use evaluation techniques based on the notion of *validity*, which is the extent to which the operationalization of a construct truly captures the intended quantity or concept. Validation techniques for unsupervised content analysis are widely found in the social science literature (Weber, 1990; Quinn et al., 2010) and have also been recently used in the NLP and machine learning communities (e.g., Chang et al., 2009; Murphy et al., 2012; Sim et al., 2013).

We used several methods to validate the stance dimensions extracted from the corpus of Reddit comments. This section describes intrinsic evaluations, which test whether the extracted stance dimensions are linguistically coherent and mean-

		Stance markers	Subreddits
Dim-1	-	beautifully, pleased, thanks, spectacular, delightful	philosophy, history, science
	+	just, even, all, no, so	pcmasterrace, leagueoflegends, gaming
Dim-2	-	suggests that, demonstrates, conclude, demonstrated, demonstrate	philosophy, science, askscience,
	+	lovely, awww, hehe, aww, haha	gonewild, nsfw, aww
Dim-3	-	funnier, hilarious, disturbing, creepy, funny	cringe, creepy, cringepics
	+	thanks, ideally, calculate, estimate, calculation	askscience, personalfinance, space
Dim-4	-	phenomenal, bummed, enjoyed, fantastic, disappointing	movies, television, books
	+	hello, thx, hehe, aww, hi	philosophy, 4chan, atheism
Dim-5	-	lovely, stunning, wonderful, delightful, beautifully	gonewild, aww, tattoos
	+	nvm, cmon, smh, lmao, disappointing	nfl, soccer, cringe
Dim-6	-	stunning, fantastic, incredible, amazing, spectacular	philosophy, gonewild, askscience
	+	anxious, stressed, exhausted, overwhelmed, relieved	relationships, sex, nosleep

Table 3: For each of the six dimensions extracted by our method, we show the five markers and three subreddits (among the 100 most popular subreddits) with the highest loadings.

ingful, thereby testing the *construct* or *content validity* of the proposed stance dimensions (Quinn et al., 2010). Extrinsic evaluations are presented in section 7.

6.1 Word Intrusion Task

A word intrusion task is used to measure the coherence and interpretability of a group of words. Human raters are presented with a list of terms, all but one of which are selected from a target concept; their task is to identify the intruder. If the target concept is internally coherent, human raters should be able to perform this task accurately; if not, their selections should be random. Word intrusion tasks have previously been used to validate the interpretability of topic models (Chang et al., 2009) and vector space models (Murphy et al., 2012).

We deployed a word intrusion task on Amazon Mechanical Turk (AMT), in which we presented the top four stance markers from one end of a dimension, along with an intruder marker selected from the top four markers of the opposite end of that dimension. In this way, we created four word intrusion tasks for each end of each dimension. The main reason for including only the top four words in each dimension is the expense of conducting crowd-sourced evaluations. In the most relevant prior work, Chang et al. (2009) used the top five words from each topic in their evaluation of topic models.

Worker selection We required that the AMT workers ("turkers") have completed a minimum of 1,000 HITs and have at least 95% approval rate Furthermore, because our task is based on analysis of English language texts, we required the turkers to be native speakers of English living in one of the majority English speaking countries. As a further requirement, we required the turkers to obtain a qualification which involves an English comprehension test similar to the questions in standardized English language tests. These requirements are based on best practices identified by Callison-Burch and Dredze (2010).

Task specification Each AMT human intelligence task (HIT) consists of twelve word intrusion tasks, one for each end of the six dimensions. We provided minimal instructions regarding the task, and did not provide any examples, to avoid introducing bias.[8] As a further quality control, each HIT included three questions which ask the turkers to pick the best synonym for a given word from a list of five answers, where one answer was clearly correct; Turkers who gave incorrect answers were to be excluded, but this situation did not arise in practice. Altogether each HIT consists of 15 questions, and was paid US$1.50. Five different turkers performed each HIT.

Results We measured the interrater reliability using Krippendorf's α (Krippendorff, 2007) and the *model precision* metric of Chang et al. (2009). Results on both metrics were encouraging. We obtained a value of $\alpha = 0.73$, on a scale where

[8]The prompt for the word intrusions task was: "Select the intruder word/phrase: you will be given a list of five English words/phrases and asked to pick the word/phrase that is least similar to the other four words/phrases when used in online discussion forums."

$\alpha = 0$ indicates chance agreement and $\alpha = 1$ indicates perfect agreement. The model precision was 0.82; chance precision is 0.20. To offer a sense of typical values for this metric, Chang et al. (2009) report model precisions in the range 0.7–0.83 in their analysis of topic models. Overall, these results indicate that the multi-dimensional analysis has succeeded at identifying dimensions that reflect natural groupings of stance markers.

6.2 Pre-registered Hypotheses

Content validity was also assessed using a set of pre-registered hypotheses. The practice of pre-registering hypotheses before an analysis and testing the correctness is widely used in the social sciences; it was adopted by Sim et al. (2013) to evaluate the induction of political ideological models from text. Before performing the muti-dimensional analysis, we identified two groups of hypotheses that are expected to hold with respect to the latent stancetaking dimensions using our prior linguistic knowledge:

- **Hypothesis I:** Stance markers that are synonyms should not appear on the opposite ends of a stance dimension.
- **Hypothesis II:** If at least one stance marker from a predefined *stance feature group* (defined below) appears on one end of a stance dimension, then other markers from the same feature group will tend not to appear at the opposite end of the same dimension.

6.2.1 Synonym Pairs

For each marker in our stance lexicon, we extracted synonyms from Wordnet, focusing on markers that appear in only one Wordnet synset, and not including pairs in which one term was an inflection of the other.[9] Our final list contains 73 synonym pairs (e.g., *eventually/finally*, *grateful/thankful*, *yea/yeah*). Of these pairs, there were 59 cases in which both terms appeared in either the top or bottom 200 positions of a stance dimension. In 51 of these cases (86%), the two terms appeared on the same side of the dimension. The chance rate would be 50%, so this supports Hypothesis I and further validates the stance dimensions. More details of the results are shown in Table 4. Note that synonym pairs may differ in aspects such as formality (e.g., *said/informed*, *want/desire*), which is one of the main dimensions of stancetaking. Therefore, perfect support for Hypothesis I is not expected.

[9]It is possible that inflections are semantically similar, because by definition they are changes in the form of a word to mark distinctions such as tense, person, or number. However, different inflections of a single word form might be used to mark different stances (e.g., some stances might be associated with the past while others might be associated with the present or future).

Stance Dimension	Number of synonym pairs On same end	On opposite ends
DIMENSION 1	6	3
DIMENSION 2	12	2
DIMENSION 3	2	1
DIMENSION 4	11	0
DIMENSION 5	10	2
DIMENSION 6	10	0
Total	51/59	8/59

Table 4: Results for pre-registered hypothesis that stance dimensions will not split synonym pairs.

6.2.2 Stance Feature Groups

Biber and Finegan (1989) group stance markers into twelve "feature groups", such as certainty adverbs, doubt adverbs, affect expressions, and hedges. Ideally, the stance dimensions should preserve these groupings. To test this, for each of the seven feature groups with at least ten stance markers in the lexicon, we counted the number of terms appearing among the top 200 positions in both ends (high/low) of each dimension. Under the null hypothesis, the stance dimensions are random with respect to the feature groups, so we would expect roughly an equal number of markers on both ends. As shown in Table 5, for five of the seven feature groups, it is possible to reject the null hypothesis at $p < .007$, which is the significance threshold at $\alpha = 0.05$, after correcting for multiple comparisons using the Bonferroni correction. This indicates that the stance dimensions are aligned with predefined stance feature groups.

7 Extrinsic Evaluations

The evaluations in the previous section test internal validity; we now describe evaluations testing whether the stance dimensions are relevant to external social and interactional phenomena.

7.1 Predicting Cross-posting

Online communities can be considered as *communities of practice* (Eckert and McConnell-Ginet, 1992), where members come together to engage in shared linguistic practices. These practices

Feature group	#Stance marker	χ^2	p-value	Reject null?
Certainty adv.	38	16.94	$4.6e^{-03}$	✓
Doubt adv.	23	13.21	$2.2e^{-02}$	×
Certainty verbs	36	48.99	$2.2e^{-09}$	✓
Doubt verbs	55	30.45	$1.2e^{-05}$	✓
Certainty adj.	28	29.73	$1.7e^{-05}$	✓
Doubt adj.	12	14.80	$1.1e^{-02}$	×
Affect exp.	227	97.17	$2.1e^{-19}$	✓

Table 5: Results for preregistered hypothesis that stance dimensions will align with stance feature groups of Biber and Finegan (1989).

evolve simultaneously with membership, coalescing into shared norms. The memberships of multiple subreddits on the same topic (e.g., *r/science* and *r/askscience*) often do not overlap considerably. Therefore we hypothesize that users of Reddit have preferred interactional styles, and that participation in subreddit communities is governed not only by topic interest, but also by these interactional preferences. The proposed stancetaking dimensions provide a simple measure of interactional style, allowing us to test whether it is predictive of community membership decisions.

Classification task We design a classification task, in which the goal is to determine whether a pair of subreddits is *high-crossover* or *low-crossover*. In high-crossover subreddit pairs, individuals are especially likely to participate in both. For the purpose of this evaluation, individuals are considered to participate in a subreddit if they contribute posts or comments. We compute the pointwise mutual information (PMI) with respect to cross-participation among the 100 most popular subreddits. For each subreddit s, we identify the five highest and lowest PMI pairs $\langle s, t \rangle$, and add these to the high-crossover and low-crossover sets, respectively. Example pairs are shown in Table 6. After eliminating redundant pairs, we identify 437 unique high-crossover pairs, and 465 unique low-crossover pairs. All evaluations are based on multiple random training/test splits over this dataset.

Classification approaches A simple classification approach is to predict that subreddits with similar text will have high crossover. We measure similarity using TF-IDF weighted cosine similarity, using two possible lexicons: the 8,000 most frequent words on reddit (BOW), and the stance lexicon (STANCE MARKERS). The similarity threshold between high-crossover and low-crossover pairs was estimated on the training data. We also tested the relevance of multi-dimensional analysis, by applying SVD to both lexicons. For each pair of subreddits, we computed a feature set of the absolute difference across the top six latent dimensions, and applied a logistic regression classifier. Regularization was tuned by internal cross-validation.

Cross-Community Participation	
High-Scoring Pairs	**Low-Scoring Pairs**
r/blog, r/announcements	r/gonewild, r/leagueoflegends
r/pokemon, r/wheredidthesodago	r/soccer, r/nosleep
r/politics, r/technology	r/programming, r/gonewild
r/LifeProTips, r/dataisbeautiful	r/nfl, r/leagueoflegends
r/Unexpected, r/JusticePorn	r/Minecraft, r/personalfinance

Table 6: Examples of subreddit pairs that have large and small amount of overlap of contributing members.

	Cosine	SVD
BOW	66.13%	77.48%
STANCE MARKERS	64.31%	84.93%

Table 7: Accuracy for prediction of subreddit cross-participation.

Results Table 7 shows average accuracies for these models. The stance-based SVD features are considerably more accurate than the BOW-based SVD features, indicating that interactional style does indeed predict cross-posting behavior.[10] Both are considerably more accurate than the bag-of-words models based on cosine similarity.

7.2 Politeness and Formality

The utility of the induced stance dimensions depends on their correlation with social phenomena of interest. Prior work has used crowdsourcing to annotate texts for politeness and formality. We now evaluate the stancetaking properties of these annotated texts.

Data We used the politeness corpus of Wikipedia edit requests from Danescu-Niculescu-Mizil et al. (2013), which includes the textual content of the edit requests, along with scalar annotations of politeness. Following the original

[10] We use BOW+SVD as the most comparable content-based alternative to our stancetaking dimensions. While there may be more accurate discriminative approaches, our goal is a direct comparison of stance and content-based features, not an exhaustive comparison of classification approaches.

authors, we compare the text for the messages ranked in the first and fourth quartiles of politeness scores. For formality, we used the corpus from Pavlick and Tetreault (2016), focusing on the blogs domain, which is most similar to our domain of Reddit. Each sentence in this corpus was annotated for formality levels from -3 to $+3$. We considered only the sentences with mean formality score greater than $+1$ (more formal) and less than -1 (less formal).

Stance dimensions For each document in the above datasets, we compute the stance properties, as follows: for each dimension, we compute the total frequency of the hundred most positive terms and the hundred most negative terms, and then take the difference. Instances containing no terms from either list are excluded. We focus on stance dimensions two and five (summarized in Table 3), because they appeared to be most relevant to politeness and formality. Dimension two contrasts informational and argumentative language against emotional and non-standard language. Dimension five contrasts positive and formal language against non-standard and somewhat negative language.

Results A kernel density plot of the resulting differences is shown in Figure 2. The effect sizes of the resulting differences are quantified using Cohen's d statistic (Cohen, 1988). Effect sizes for all differences are between 0.3 and 0.4, indicating small-to-medium effects — except for the evaluation of formality on dimension five, where the effect size is close to zero. The relatively modest effect sizes are unsurprising, given the short length of the texts. However, these differences lend insight to the relationship between formality and politeness, which may seem to be closely related concepts. On dimension two, it is possible to be polite while using non-standard language such as *hehe* and *awww*, so long as the sentiment expressed is positive; however, these markers are not consistent with formality. On dimension five, we see that positive sentiment terms such as *lovely* and *stunning* are consistent with politeness, but not with formality. Indeed, the distribution of dimension five indicates that both ends of dimension five are consistent only with informal texts.

Overall, these results indicate that interactional phenomena such as politeness and formality are reflected in our stance dimensions, which are induced in an unsupervised manner. Future work may consider the utility of these stance dimensions to predict these social phenomena, particularly in cross-domain settings where lexical classifiers may overfit.

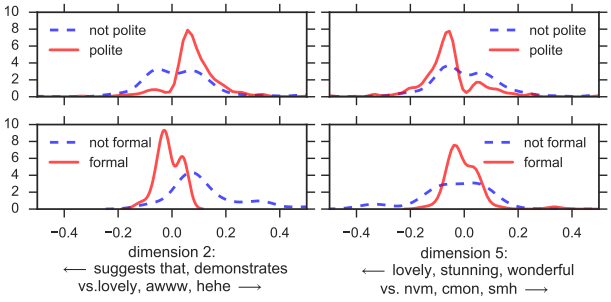

Figure 2: Kernel density distributions for stance dimensions 2 and 5, plotted with respect to annotations of politeness and formality.

8 Conclusion

Stancetaking provides a general perspective on the various linguistic phenomena that structure social interactions. We have identified a set of several hundred stance markers, building on previously-identified lexicons by using word embeddings to perform lexicon expansion. We then used multi-dimensional analysis to group these markers into stance dimensions, which we show to be internally coherent and extrinsically useful. Our hope is that these stance dimensions will be valuable as a convenient building block for future research on interactional meaning.

Acknowledgments Thanks to the anonymous reviewers for their useful and constructive feedback on our submission. This research was supported by Air Force Office of Scientific Research award FA9550-14-1-0379, by National Institutes of Health award R01-GM112697, and by the National Science Foundation awards 1452443 and 1111142. We thank Tyler Schnoebelen for helpful discussions; C.J. Hutto, Tanushree Mitra, and Sandeep Soni for assistance with Mechanical Turk experiments; and Ian Stewart for assistance with creating word embeddings. We also thank the Mechanical Turk workers for performing the word intrusion task, and for feedback on a pilot task.

References

Lada A. Adamic, Jun Zhang, Eytan Bakshy, and Mark S. Ackerman. 2008. Knowledge sharing and yahoo answers: Everyone knows something. In

Proceedings of the Conference on World-Wide Web (WWW). pages 665–674.

Shlomo Argamon, Moshe Koppel, James W. Pennebaker, and Jonathan Schler. 2007. Mining the blogosphere: Age, gender and the varieties of self-expression. *First Monday* 12(9).

Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. 2006. Group formation in large social networks: Membership, growth, and evolution. In *Proceedings of Knowledge Discovery and Data Mining (KDD)*. pages 44–54.

David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. 2014. Gender identity and lexical variation in social media. *Journal of Sociolinguistics* 18(2):135–160.

Douglas Biber. 1991. *Variation across speech and writing*. Cambridge University Press.

Douglas Biber. 1992. The multi-dimensional approach to linguistic analyses of genre variation: An overview of methodology and findings. *Computers and the Humanities* 26(5-6):331–345.

Douglas Biber and Edward Finegan. 1989. Styles of stance in english: Lexical and grammatical marking of evidentiality and affect. *Text* 9(1):93–124.

Su Lin Blodgett, Lisa Green, and Brendan OConnor. 2016. Demographic dialectal variation in social media: A case study of african-american english. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*. pages 1119–1130.

M. Bucholtz and K. Hall. 2005. Identity and interaction: A sociocultural linguistic approach. *Discourse studies* 7(4-5):585–614.

John A Bullinaria and Joseph P Levy. 2007. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior research methods* 39(3):510–526.

John D. Burger, John Henderson, George Kim, and Guido Zarrella. 2011. Discriminating gender on twitter. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*. pages 1301–1309.

Chris Callison-Burch and Mark Dredze. 2010. Creating speech and language data with amazon's mechanical turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*. Association for Computational Linguistics, pages 1–12.

Raymond B Cattell. 1966. The scree test for the number of factors. *Multivariate behavioral research* 1(2):245–276.

Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L Boyd-graber, and David M Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Neural Information Processing Systems (NIPS)*. Vancouver, pages 288–296.

Jacob Cohen. 1988. *Statistical power analysis for the behavioral sciences*. Lawrence Earlbaum Associates, Hillsdale, NJ.

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the Association for Computational Linguistics (ACL)*. Sophia, Bulgaria, pages 250–259.

John W. Du Bois. 2007. The stance triangle. In Robert Engelbretson, editor, *Stancetaking in discourse*, John Benjamins Publishing Company, Amsterdam/Philadelphia, pages 139–182.

Penelope Eckert. 2012. Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41:87–100.

Penelope Eckert and Sally McConnell-Ginet. 1992. Think practically and look locally: Language and gender as community-based practice. *Annual review of anthropology* 21:461–490.

Penelope Eckert and John R Rickford. 2001. *Style and sociolinguistic variation*. Cambridge University Press.

Jacob Eisenstein. 2013. What to do about bad language on the internet. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*. pages 359–369.

Jacob Eisenstein, Amr Ahmed, and Eric P. Xing. 2011. Sparse additive generative models of text. In *Proceedings of the International Conference on Machine Learning (ICML)*. pages 1041–1048.

Jacob Eisenstein, Brendan O'Connor, Noah A. Smith, and Eric P. Xing. 2010. A latent variable model for geographic lexical variation. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*. pages 1277–1287.

Valerie Freeman, Richard Wright, Gina-Anne Levow, Yi Luan, Julian Chan, Trang Tran, Victoria Zayats, Maria Antoniak, and Mari Ostendorf. 2014. Phonetic correlates of stance-taking. *The Journal of the Acoustical Society of America* 136(4):2175–2175.

Eric Gilbert. 2013. Widespread underprovision on reddit. In *Proceedings of Computer-Supported Cooperative Work (CSCW)*. pages 803–808.

William L. Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky. 2016. Inducing domain-specific sentiment lexicons from unlabeled corpora. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*. pages 595–605.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the Association for Computational Linguistics (ACL)*. Madrid, Spain, pages 174–181.

Jack Hessel, Chenhao Tan, and Lillian Lee. 2014. Science, askscience, and badscience: On the coexistence of highly related communities. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*. AAAI Publications, Menlo Park, California, pages 171–180.

Alexandra Jaffe. 2009. *Stance: Sociolinguistic Perspectives*. Oxford University Press.

Daniel Jurafsky, Elizabeth Shriberg, Barbara Fox, and Traci Curl. 1998. Lexical, prosodic, and syntactic cues for dialog acts. In *Proceedings of ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers*. pages 114–120.

Elise Kärkkäinen. 2006. Stance taking in conversation: From subjectivity to intersubjectivity. *Text & Talk-An Interdisciplinary Journal of Language, Discourse Communication Studies* 26(6):699–731.

Tiina Keisanen. 2007. Stancetaking as an interactional activity: Challenging the prior speaker. *Stancetaking in discourse: Subjectivity, evaluation, interaction* pages 253–81.

Scott Fabius Kiesling. 2009. Style as stance. *Stance: sociolinguistic perspectives* pages 171–194.

Klaus Krippendorff. 2007. Computing krippendorff's alpha reliability. *Departmental papers (ASC)* page 43.

Thomas Landauer, Peter W. Foltz, and Darrel Laham. 1998. Introduction to latent semantic analysis. *Discource Processes* 25:259–284.

Wang Ling, Chris Dyer, Alan W Black, and Isabel Trancoso. 2015. Two/too simple adaptations of word2vec for syntax problems. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*. Denver, CO, pages 1299–1304.

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*. pages 3111–3119.

Brian Murphy, Partha Pratim Talukdar, and Tom Mitchell. 2012. Learning effective and interpretable semantic models using non-negative sparse embedding. In *Proceedings of the International Conference on Computational Linguistics (COLING)*. Mumbai, India, pages 1933–1949.

Dong Nguyen, A Seza Doğruöz, Carolyn P Rosé, and Franciska de Jong. 2016. Computational sociolinguistics: A survey. *Computational Linguistics* 42(3):537–593.

Dong Nguyen, Rilana Gravel, Dolf Trieschnigg, and Theo Meder. 2013. "How Old Do You Think I Am?" A Study of Language and Age in Twitter. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*. pages 439–448.

Yoshiki Niwa and Yoshihiko Nitta. 1994. Co-occurrence vectors from corpora vs. distance vectors from dictionaries. In *Proceedings of the International Conference on Computational Linguistics (COLING)*. Kyoto, Japan, pages 304–309.

Elinor Ochs. 1993. Constructing social identity: A language socialization perspective. *Research on language and social interaction* 26(3):287–306.

Ellie Pavlick and Joel Tetreault. 2016. An empirical analysis of formality in online communication. *Transactions of the Association for Computational Linguistics (TACL)* 4:61–74.

James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015. Technical report.

Vinodkumar Prabhakaran, Owen Rambow, and Mona Diab. 2012. Predicting overt display of power in written dialogs. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*. pages 518–522.

Kristen Precht. 2003. Stance moods in spoken english: Evidentiality and affect in british and american conversation. *Text - Interdisciplinary Journal for the Study of Discourse* 23(2):239–258.

Kevin M Quinn, Burt L Monroe, Michael Colaresi, Michael H Crespin, and Dragomir R Radev. 2010. How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science* 54(1):209–228.

Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*. pages 105–112.

Yanchuan Sim, Brice Acree, Justin H Gross, and Noah A Smith. 2013. Measuring ideological proportions in political speeches. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.

Philip J. Stone. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.

Chenhao Tan and Lillian Lee. 2015. All who wander: On the prevalence and characteristics of multi-community engagement. In *Proceedings of the Conference on World-Wide Web (WWW)*. pages 1056–1066.

Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29(1):24–54.

Trang Tran and Mari Ostendorf. 2016. Characterizing the language of online communities and its relation to community reception. In *Proceedings of*

Empirical Methods for Natural Language Processing (EMNLP).

Marilyn A Walker, Pranav Anand, Robert Abbott, and Ricky Grant. 2012. Stance classification using dialogic properties of persuasion. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*. pages 592–596.

Robert Philip Weber. 1990. *Basic content analysis*. 49. Sage.

Peter RR White. 2003. Beyond modality and hedging: A dialogic view of the language of intersubjective stance. *Text - Interdisciplinary Journal for the Study of Discourse* 23(2):259–284.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language resources and evaluation* 39(2):165–210.