Author Commitment and Social Power: Automatic Belief Tagging to Infer the Social Context of Interactions

Vinodkumar Prabhakaran
Stanford University
vinod@cs.stanford.edu

Premkumar Ganeshkumar
Agolo, Inc.
prem@agolo.com

Owen Rambow
Elemental Cognition, Inc.
owenr@elementalcognition.com

Abstract

Understanding how social power structures affect the way we interact with one another is of great interest to social scientists who want to answer fundamental questions about human behavior, as well as to computer scientists who want to build automatic methods to infer the social contexts of interactions. In this paper, we employ advancements in extra-propositional semantics extraction within NLP to study how author commitment reflects the social context of an interactions. Specifically, we investigate whether the level of commitment expressed by individuals in an organizational interaction reflects the hierarchical power structures they are part of. We find that subordinates use significantly more instances of non-commitment than superiors. More importantly, we also find that subordinates attribute propositions to other agents more often than superiors do — an aspect that has not been studied before. Finally, we show that enriching lexical features with commitment labels captures important distinctions in social meanings.

1 Introduction

Social power is a difficult concept to define, but is often manifested in how we interact with one another. Understanding these manifestations is important not only to answer fundamental questions in social sciences about power and social interactions, but also to build computational models that can automatically infer social power structures from interactions. The availability and access to large digital repositories of naturally occurring social interactions and the advancements in natural language processing techniques in recent years have enabled researchers to perform large scale studies on linguistic correlates of power, such as words and phrases (Bramsen et al., 2011; Gilbert, 2012), linguistic coordination (Danescu-Niculescu-Mizil et al., 2012), agenda control (Taylor et al., 2012), and dialog structure (Prabhakaran and Rambow, 2014).

Another area of research that has recently garnered interest within the NLP community is the modeling of author commitment in text. Initial studies in this area were done in processing hedges, uncertainty and lack of commitment, specifically focused on scientific text (Mercer et al., 2004; Di Marco et al., 2006; Farkas et al., 2010). More recently, researchers have also looked into capturing author commitment in non-scientific text, e.g., levels of factuality in newswire (Saurí and Pustejovsky, 2009), types of commitment of beliefs in a variety of genres including conversational text (Diab et al., 2009; Prabhakaran et al., 2015). These approaches are motivated from an information extraction perspective, for instance in aiding tasks such as knowledge base population.[1] However, it has not been studied whether such sophisticated author commitment analysis can go beyond what is expressed in language and reveal the underlying social contexts in which language is exchanged.

In this paper, we bring together these two lines of research; we study how power relations correlate with the levels of commitment authors express in interactions. We use the power analysis framework built by Prabhakaran and Rambow (2014) to perform this study, and measure author commitment using the committed belief tagging framework introduced by (Diab et al., 2009) that distinguishes different types of beliefs expressed in text. Our contributions are two-fold — statistical analysis of author commitment in relation with power, and enrichment of lexical features with commitment labels to aid in computational prediction of power relations. In the first part, we find that au-

[1] The BeSt track of the 2017 TAC-KBP evaluation aimed at detecting the "belief and sentiment of an entity toward another entity, relation, or event" (http://www.cs.columbia.edu/~rambow/best-eval-2017/).

thor commitment is significantly correlated with the social power relations between their participants — subordinates use more instances of non-commitment, a finding that is in line with sociolinguistics studies in this area. We also find that subordinates use significantly more reported beliefs (i.e., attributing beliefs to other agents) than superiors. This is a new finding; to our knowledge, there has not been any sociolinguistics studies investigating this aspect of interaction in relation with power. In the second part, we present novel ways of incorporating the author commitment information into lexical features that can capture important distinctions in word meanings conveyed through the belief contexts in which they occur; distinctions that are lost in a model that conflates all occurrences of a word into one unit.

We first describe the related work in computational power analysis and computational modeling of cognitive states in Section 2. In Section 3, we describe the power analysis framework we use. Section 4 formally defines the research questions we are investigating, and describes how we obtain the belief information. In Section 5, we present the statistical analysis of author commitment and power. Section 6 presents the utility of enriching lexical features with belief labels in the context of automatic power prediction. Section 7 concludes the paper and summarizes the results.

2 Related Work

The notion of belief that we use in this paper (Diab et al., 2009; Prabhakaran et al., 2015) is closely related to the notion of factuality that is captured in FactBank (Saurí and Pustejovsky, 2009). They capture three levels of factuality, certain (CT), probable (PB), and possible (PS), as well as the underspecified factuality (Uu). They also record the corresponding polarity values, and the source of the factuality assertions to distinguish between factuality assertions by the author and those by the agents/sources introduced by the author. While FactBank offers a finer granularity, they are annotated on newswire text. Hence, we use the corpus of belief annotations (Prabhakaran et al., 2015) that is obtained on online discussion forums, which is closer to our genre.

Automatic hedge/uncertainty detection is a very closely related task to belief detection. The belief tagging framework we use aims to capture the cognitive states of authors, whereas hedges are linguistic expressions that convey one of those cognitive states — non-committed beliefs. Automatic hedge/uncertainty detection has generated active research in recent years within the NLP community. Early work in this area focused on detecting speculative language in scientific text (Mercer et al., 2004; Di Marco et al., 2006; Kilicoglu and Bergler, 2008). The open evaluation as part of the CoNLL shared task in 2010 to detect uncertainty and hedging in biomedical and Wikipedia text (Farkas et al., 2010) triggered further research on this problem in the general domain (Agarwal and Yu, 2010; Morante et al., 2010; Velldal et al., 2012; Choi et al., 2012). Most of this work was aimed at formal scientific text in English. More recent work has tried to extend this work to other genres (Wei et al., 2013; Sanchez and Vogel, 2015) and languages (Velupillai, 2012; Vincze, 2014), as well as building general purpose hedge lexicons (Prokofieva and Hirschberg, 2014). In our work, we use the lexicons from (Prokofieva and Hirschberg, 2014) to capture hedges in text.

Sociolinguists have long studied the association between level of commitment and social contexts (Lakoff, 1973; O'Barr and Atkins, 1980; Hyland, 1998). A majority of this work studies gender differences in the use of hedges, triggered by the influential work by Robin Lakoff (Lakoff, 1973). She argued that women use linguistic strategies such as hedging and hesitations in order to adopt an unassertive communication style, which she terms "women's language". While many studies have found evidence to support Lakoff's theory (e.g., (Crosby and Nyquist, 1977; Preisler, 1986; Carli, 1990)), there have also been contradictory findings (e.g., (O'Barr and Atkins, 1980)) that link the difference in the use of hedges to other social factors (e.g., power). O'Barr and Atkins (1980) argue that the use of hedges is linked more to the social positions rather than gender, suggesting to rename "women's language" to "powerless language". In later work, O'Barr (1982) formalized the notion of powerless language, which formed the basis of many sociolinguistics studies on social power and communication. O'Barr (1982) analyzed courtroom interactions and identified hedges and hesitations as some of the linguistic markers of "powerless" speech. However, there has not been any computational work which has looked into how power relations relate to the level of commitment expressed in text. In this paper, we use com-

putational power analysis to perform a large scale data-oriented study on how author commitment in text reveals the underlying power relations.

There is a large body of literature in the social sciences that studies power as a social construct (e.g., (French and Raven, 1959; Dahl, 1957; Emerson, 1962; Pfeffer, 1981; Wartenberg, 1990)) and how it relates to the ways people use language in social situations (e.g., (Bales et al., 1951; Bales, 1970; O'Barr, 1982; Van Dijk, 1989; Bourdieu and Thompson, 1991; Ng and Bradac, 1993; Fairclough, 2001; Locher, 2004)). Recent years have seen growing interest in computationally analyzing and detecting power and influence from interactions. Early work in computational power analysis used social network analysis based approaches (Diesner and Carley, 2005; Shetty and Adibi, 2005; Creamer et al., 2009) or email traffic patterns (Namata et al., 2007). Using NLP to deduce social relations from online communication is a relatively new area of active research.

Bramsen et al. (2011) and Gilbert (2012) first applied NLP based techniques to predict power relations in Enron emails, approaching this task as a text classification problem using bag of words or ngram features. More recently, our work has used dialog structure features derived from deeper dialog act analysis for the task of power prediction in Enron emails (Prabhakaran and Rambow, 2014; Prabhakaran et al., 2012; Prabhakaran and Rambow, 2013). In this paper, We use the framework of (Prabhakaran and Rambow, 2014), but we analyze a novel aspect of interaction that has not been studied before — what level of commitment do the authors express in language.

There has also been work on analyzing power in other genres of interactions. Strzalkowski et al. (2010) and Taylor et al. (2012) concentrate on lower-level constructs called *Language Uses* such as agenda control to predict power in Wikipedia talk pages. Danescu-Niculescu-Mizil et al. (2012) study how social power and linguistic coordination are correlated in Wikipedia interactions as well as Supreme Court hearings. Bracewell et al. (2012) and Swayamdipta and Rambow (2012) try to identify pursuit of power in discussion forums. Biran et al. (2012) and Rosenthal (2014) study the problem of predicting influence in Wikipedia talk pages, blogs, and other online forums. Prabhakaran et al. (2013) study manifestations of power of confidence in presidential debates.

3 Power in Workplace Email: Data and Analysis Framework

The focus of our study is to investigate whether the level of commitment participants express in their contributions in an interaction is related to the power relations they have with other participants, and how it can help in the problem of predicting social power. In this section, we introduce the power analysis framework as well as the data we use in this study.

3.1 Problem

In order to model manifestations of power relations in interactions, we use our interaction analysis framework from (Prabhakaran and Rambow, 2014), where we introduced the problem of predicting organizational power relations between pairs of participants based on single email threads. The problem is formally defined as follows: given an email thread t, and a related interacting participant pair (p_1, p_2) in the thread, predict whether p_1 is the *superior* or *subordinate* of p_2. In this formulation, a *related interacting participant pair (RIPP)* is a pair of participants of the thread such that there is at least one message exchanged within the thread between them (in either direction) and that they are hierarchically related with a superior/subordinate relation.

3.2 Data

We use the same dataset we used in (Prabhakaran and Rambow, 2014), which is a version of the Enron email corpus in which the thread structure of email messages is reconstructed (Yeh and Harnly, 2006), and enriched by Agarwal et al. (2012) with gold organizational power relations, manually determined using information from Enron organizational charts. The corpus captures dominance relations between 13,724 pairs of Enron employees. As in (Prabhakaran and Rambow, 2014), we use these dominance relation tuples to obtain gold labels for the *superior* or *subordinate* relationships between pairs of participants. We use the same train-test-dev split as in (Prabhakaran and Rambow, 2014). We summarize the number of threads and related interacting participant pairs in each subset of the data in Table 1.

4 Research Hypotheses

Our first objective in this paper is to perform a large scale computational analysis of author com-

Description	Train	Dev	Test
Email threads	18079	8973	9144
# of RIPPs	7510	3578	3920

Table 1: Data Statistics. Row 1: number of threads in subsets of the corpus. Row 2: number of related interacting participant pairs in those subsets. RIPP: Related interacting participant pairs

mitment and power relations. Specifically, we want to investigate whether the commitment authors express towards their contributions in organizational interactions is correlated with the power relations they have with other participants. Sociolinguistics studies have found some evidence to suggest that lack of commitment expressed through hedges and hesitations is associated with lower power status (O'Barr, 1982). However, in our study, we go beyond hedge word lists, and analyze different cognitive belief states expressed by authors using a belief tagging framework that takes into account the syntactic contexts within which propositions are expressed.

4.1 Obtaining Belief Labels

We use the committed belief analysis framework introduced by (Diab et al., 2009; Prabhakaran et al., 2015) to model different levels of beliefs expressed in text. Specifically, in this paper, we use the 4-way belief distinction — COMMITTEDBELIEF, NONCOMMITTEDBELIEF, REPORTEDBELIEF, and NONAPPLICABLE— introduced in (Prabhakaran et al., 2015).[2] (Prabhakaran et al., 2015) presented a corpus of online discussion forums with over 850K words, annotating each propositional head in text with one of the four belief labels. The paper also presented an automatic belief tagger trained on this data, which we use to obtain belief labels in our data. We describe each belief label and our associated hypotheses below.

Committed belief (CB): the writer strongly believes that the proposition is true, and wants the reader/hearer to believe that. E.g.:

(1) a. John will **submit** the report.
b. I know that John is **capable**.

[2]We also performed analysis and experiments using an earlier 3-way belief distinction proposed by (Diab et al., 2009), which also yielded similar findings. We do not report the details of those analyses in this paper.

As discussed earlier, lack of commitment in one's writing/speech is identified as markers of powerless language. We thus hypothesize:

H. 1. *Superiors use more instances of committed belief in their messages than subordinates.*

Non-committed belief (NCB): the writer explicitly identifies the proposition as something which he or she could believe, but he or she happens not to have a strong belief in, for example by using an epistemic modal auxiliary. E.g.:

(2) a. John may **submit** the report.
b. I guess John is **capable**.

This class captures a more semantic notion of non-commitment than hedges, since the belief annotation attempts to model the underlying meaning rather than language uses, and hence captures other linguistic means of expressing non-committedness. Following (O'Barr, 1982), we formulate the below hypothesis:

H. 2. *Subordinates use more instances of non committed belief in their messages than superiors.*

Reported belief (ROB): the writer attributes belief (either committed or non-committed) to another person or group. E.g.:

(3) a. Sara says John will **submit** the report.
b. Sara thinks John may be **capable**.

Note that this label is only applied when the writer's own belief in the proposition is unclear. For instance, if the first example above was *Sara knows John will submit the report on-time*, the writer is expressing commitment toward the proposition that John will submit the report and it will be labeled as committed belief rather than reported belief. Reported belief captures instances where the writer is in effect limiting his/her commitment towards what is stated by attributing the belief to someone else. So, in line with our hypotheses for non-committed beliefs, we formulate the following hypothesis:

H. 3. *Subordinates use more instances of reported beliefs in their messages than superiors.*

Non-belief propositions (NA): – the writer expresses some other cognitive attitude toward the proposition, such as desire or intention (4a), or expressly states that he/she has no belief about the proposition (e.g., asking a question (4b)). E.g.:

(4) a. I need John to **submit** the report.
b. Will John be **capable**?

As per the above definition, requests for information (i.e., questions) and requests for actions are cases where the author is not expressing a belief about the proposition, but rather expressing the desire that some action be done. In the study correlating power with dialog act tags (Prabhakaran and Rambow, 2014), we found that superiors issue significantly more requests than subordinates. Hence, we expect the superiors to have significantly more non belief expressions in their messages, and formulate the following hypothesis:

H. 4. *Superiors use more instances of non beliefs in their messages than subordinates.*

4.2 Testing Belief Tagger Bias

NLP tools are imperfect and may produce errors, which poses a problem when using any NLP tool for sociolinguistic analysis. More than the magnitude of error, we believe that whether the error is correlated with the social variable of interest (i.e., power) is more important; e.g., is the belief-tagger more likely to find ROB false-positives in subordinates text? To test whether this is the case, we performed manual belief annotation on around 500 propositional heads in our corpus. Logistic regression test revealed that the belief-tagger is equally likely to make errors (both false-positives and false-negatives, for all four belief-labels) in sentences written by subordinates as superiors (the null hypothesis accepted at $p > 0.05$ for all eight tests).

5 Statistical Analysis

Now that we have set up the analysis framework and research hypotheses, we present the statistical analysis of how superiors and subordinates differ in their relative use of expressions of commitment.

5.1 Features

For each participant of each pair of related interacting participants in our corpus, we aggregate each of the four belief tags:

- *CBCount*: number of propositional heads tagged as Committed Belief (CB)
- *NCBCount*: number of propositional heads tagged as Non Committed Belief (NCB)
- *ROBCount*: number of propositional heads tagged as Reported Belief (ROB)
- *NACount*: number of propositional heads tagged as Non Belief (NA)

5.2 Hypotheses Testing

Our general hypothesis is that power relations do correlate with the level of commitment people express in their messages; i.e., at least one of H.1 - H.4 is true. In this analysis, each participant of the pair (p_1, p_2) is a data instance. We exclude the instances for which a feature value is undefined.[3]

In order to test whether superiors and subordinates use different types of beliefs, we used a linear regression based analysis. For each feature, we built a linear regression model predicting the feature value using power (i.e., superior vs. subordinate) as the independent variable. Since verbosity of a participant can be highly correlated with each of these feature values (we found it to be highly correlated with subordinates (Prabhakaran and Rambow, 2014)), we added token count as a control variable to the linear regression.

Our linear regression test revealed significant differences in NCB (b=-.095, t(-8.09), p<.001), ROB (b=-.083, t(-7.162), p<.001) and NA (b=.125, t(4.351), p<.001), and no significant difference in CB (b=.007, t(0.227), p=0.821). Figure 1 pictorially demonstrates these results by plotting the difference between the mean values of each commitment feature (here normalized by token count) of superiors vs. subordinates, as a percentage of mean feature value of the corresponding commitment feature for superiors. Dark bars denote statistically significant differences.

5.3 Interpretation of Findings

The results from our statistical analysis validate our original hypothesis that power relations do correlate with the level of commitment people express in their messages. This finding remains statistically significant ($p < 0.001$) even after applying the Bonferroni correction for multiple testing.

The results on NCB confirm our hypothesis that subordinates use more non-committedness in their language. Subordinates' messages contain 48% more instances of non-committed belief than superiors' messages, even after normalizing for the length of messages. This is in line with prior sociolinguistics literature suggesting that people with

[3]These are instances corresponding to participants who did not send any messages in the thread (some of the pairs in the set of related interacting participant pairs only had one-way communication) or whose messages were empty (e.g., forwarding messages).

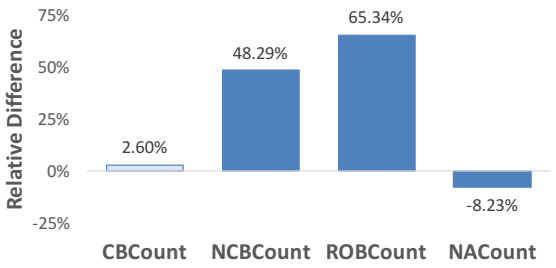

Figure 1: Relative difference (RD) between subordinates and superiors in their use of different types of commitment (counts normalized by word count of contributions). Dark bars: statistical significance at $p < 0.05$. (RD = $\frac{(Mean(Subordinates) - Mean(Superiors)) * 100}{Mean(Superiors)}$).

less power tend to use less commitment, previously measured in terms of hedges. However, in our work, we go beyond hedge dictionaries and use expressions of non-committedness that takes into account the syntactic configurations in which the words appear.

Another important finding is in terms of reported belief (ROB). Our results strongly verify the hypothesis H.3 that subordinates use significantly more reported beliefs than superiors. In fact, it obtained the largest magnitude of relative difference (65.3% more) of all features we analyzed. To our knowledge, ours is the first study that analyzed the manifestation of power in authors attributing beliefs to others. Our results are in line with the finding in (Agarwal et al., 2014) that "if many more people get mentioned to a person then that person is the boss", because as subordinates report other people's beliefs to superiors, they are also likely to mention them.

The finding that superiors use more NAs confirms our hypothesis H.4. As discussed earlier, this is expected since superiors issue more requests (as found by (Prabhakaran and Rambow, 2014)), the propositional heads of which would be tagged as NA by the belief tagger. However, our hypothesis H.1 is proven false. Being a superior or subordinate does not affect how often their messages contain CB, which suggests that power differences are manifested only in terms of lack of commitment.

6 Commitment in Power Prediction

Our next step is to explore whether we can utilize the hedge and belief labels to improve the performance of an automatic power prediction system. For this purpose, we use our POWERPREDICTOR system (Prabhakaran and Rambow, 2014) that predicts the direction of power between a pair of related interacting participants in an email thread. It uses a variety of linguistic and dialog structural features consisting of verbosity features (message count, message ratio, token count, token ratio, and tokens per message), positional features (initiator, first message position, last message position), thread structure features (number of all recipients and those in the *To* and *CC* fields of the email, reply rate, binary features denoting the adding and removing of other participants), dialog act features (request for action, request for information, providing information, and conventional), and overt displays of power, and lexical features (lemma ngrams, part-of-speech ngrams, and mixed ngrams, a version of lemma ngrams with open class words replaced with their part-of-speech tags). The feature sets are summarized in Table 2 ((Prabhakaran and Rambow, 2014) has a detailed description of these features).

Set	Description
VRB	Verbosity (e.g., message count)
PST	Positional (e.g., thread initiator?)
THR	Thread structure (e.g., reply rate)
DIA	Dialog act tagging (e.g., request count)
ODP	Overt displays of power
LEX	Lexical ngrams (lemma, POS, mixed ngrams)

Table 2: POWERPREDICTOR system: Features used

None of the features used in POWERPREDICTOR use information from the parse trees of sentences in the text However, in order to accurately obtain the belief labels, deep dependency parse based features are critical (Prabhakaran et al., 2010). We use the ClearTk wrapper for the Stanford CoreNLP pipeline to obtain the dependency parses of sentences in the email text. To ensure an unified analysis framework, we also use the Stanford CoreNLP for tokenization, part-of-speech tagging, and lemmatization steps, instead of OpenNLP. This change affects our analysis in two ways. First, the source of part-of-speech tags and word lemmas is different from what was presented in the original system, which might affect the performance of the dialog act tagger and overt display of power tagger (DIA and ODP features). Second, we had to exclude 117 threads (0.3%) from the corpus for which the Stanford CoreNLP failed to parse some sentences, resulting in the removal of 11 data points (0.2%), only one of which

was in the test set. On randomly checking, we found that they contained non-parsable text such as dumps of large tables, system logs, or unedited dumps of large legal documents.

In order to better interpret how the commitment features help in power prediction, we use a linear kernel SVM in our experiments. Linear kernel SVMs are significantly faster than higher order SVMs, and our preliminary experiments revealed the performance gain by using a higher order SVM to be only marginal. We use the best performing feature set from (Prabhakaran and Rambow, 2014) as a strong baseline for our experiments. This baseline feature set is the combination of thread structure features (THR) and lexical features (LEX). This baseline system obtained an accuracy of 68.8% in the development set.

6.1 Belief Label Enriched Lexical Features

Adding the belief label counts into the SVM directly as features will not yield much performance improvements, as signal in the aggregate counts would be minimal given the effect sizes of differences we find in Section 5. In this section, we investigate a more sophisticated way of incorporating the belief tags into the power prediction framework. Lexical features are very useful for the task of power prediction. However, it is often hard to capture deeper syntactic/semantic contexts of words and phrases using ngram features. We hypothesize that incorporating belief tags into the ngrams will enrich the representation and will help disambiguate different usages of same words/phrases. For example, let us consider two sentences: *I need the report by tomorrow* vs. *If I need the report, I will let you know*. The former is likely coming from a person who has power, whereas the latter does not give any such indication. Applying the belief tagger to these two sentences will result in *I need(CB) the report ...* and *If I need(NA) the report ....* Capturing the difference between *need(CB)* vs. *need(NA)* will help the machine learning system to make the distinction between these two usages and in turn improve the power prediction performance.

In building the ngram features, whenever we encounter a token that is assigned a belief tag, we append the belief tag to the corresponding lemma or part-of-speech tag in the ngram. We call it the *Append* version of corresponding ngram feature. We summarize the different versions of each type of ngram features below:

- *LN*: the original word lemma ngram; e.g., *i_need_the*.
- LN^{CBApnd}: word lemma ngram with appended belief tags; e.g., *i_need(CB)_the*.
- *PN*: the original part-of-speech ngram; e.g., *PRP_VB_DT*.
- PN^{CBApnd}: part-of-speech ngram with appended belief tags; e.g., *PRP_VB(CB)_DT*.
- *MN*: the original mixed ngram; e.g., *i_VB_the*.
- MN^{CBApnd}: mixed ngram with appended belief tags; e.g., *i_VB(CB)_the*.

Feature Configuration in LEXICAL	Accuracy
LN +PN +MN (*BaseLine*)	68.8
LN^{CBApnd}*+PN +MN*	**69.3**
LN +PN^{CBApnd}*+MN*	68.6
LN +PN +MN^{CBApnd}	69.0
LN^{CBApnd}*+ PN +* MN^{CBApnd}	69.2

Table 3: Power prediction results using different configurations of LEX features. (The full feature set also includes THR.)

In Table 3, we show the results obtained by incorporating the belief tags in this manner to the LEXICAL features of the original baseline feature set. The first row indicates the baseline results and the following rows show the impact of incorporating belief tags using the *Append* method. While the *Append* version of both lemma ngrams and mixed ngrams improved the results, the *Append* version of part of speech ngrams reduced the results. The combination of best performing version of each type of ngram obtained slightly lower result than using the *Append* version of word ngram alone, which posted the overall best performance of 69.3%, a significant improvement ($p<0.05$) over not using any belief information. We use the approximate randomization test (Yeh, 2000) for testing statistical significance of the improvement.

Finally, we verified that our best performing feature sets obtain similar improvements in the unseen test set. The baseline system obtained 70.2% accuracy in the test set. The best performing configuration from Table 3 significantly improved this accuracy to 70.8%. The second best performing configuration of using the *Append* version of both word and mixed ngrams obtained only a small improvement upon the baseline in the test set.

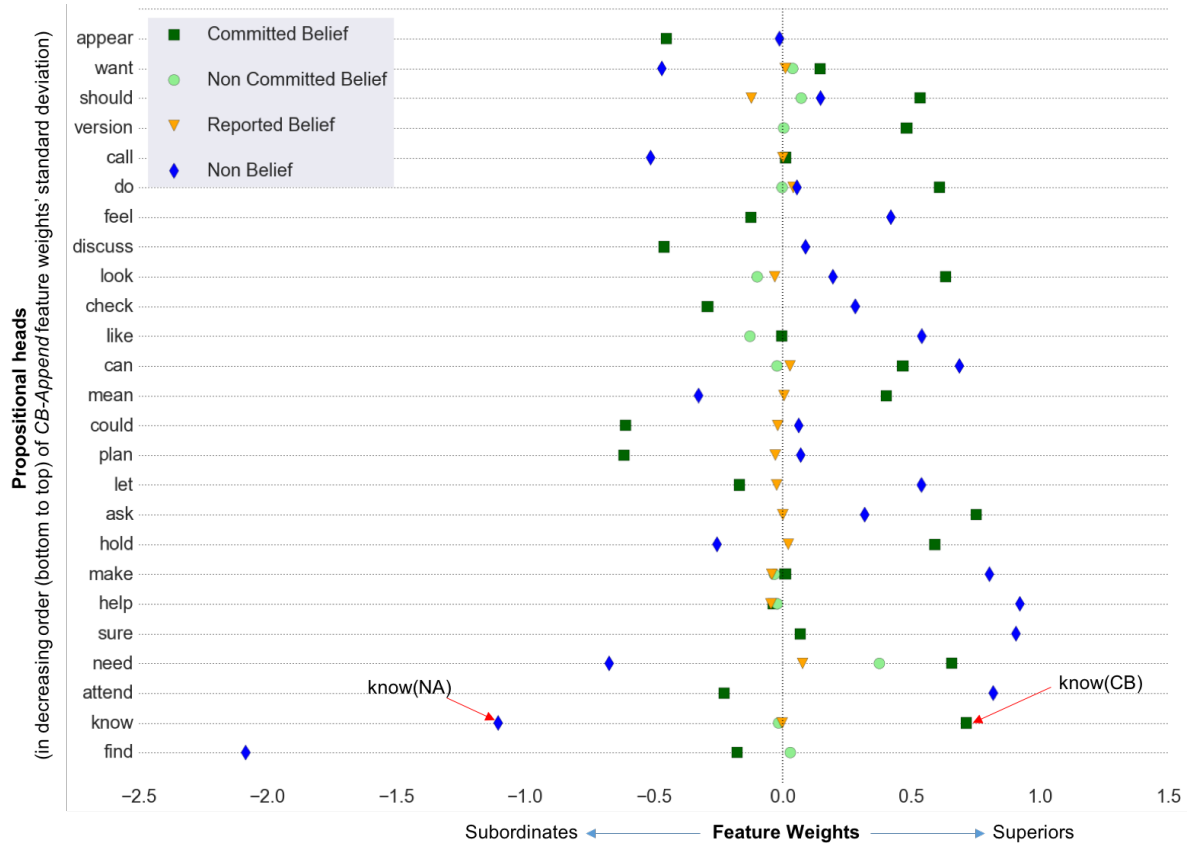

Figure 2: Feature weights of different belief appended versions of 25 propositional heads whose lemma unigrams had the highest standard deviation. Y-axis denotes the propositional heads in decreasing order of standard deviation from bottom to top. X-axis denotes the feature weights.

6.2 Word NGram Feature Analysis

We inspect the feature weights assigned to the LN^{CBApnd} version of lemma ngrams in our best performing model. Each lemma ngram that contains a propositional head (e.g., *need*) has four possible LN^{CBApnd} ngram versions: *need(CB)*, *need(NCB)*, *need(ROB)*, and *need(NA)*. For each lemma ngram, we calculate the standard deviation of weights assigned to different LN^{CBApnd} versions in the learned model as a measure of variation captured by incorporating belief tags into that ngram.[4]

Figure 2 shows the feature weights of different LN^{CBApnd} versions of twenty five propositional heads whose lemma unigrams had the highest standard deviation. The y-axis lists propositional heads arranged in the decreasing order of standard deviation from bottom to top, while the x-axis denotes the feature weights. The markers distinguish the different LN^{CBApnd} versions of each propositional head — square denotes COMMITTEDBELIEF, circle denotes NONCOMMITTEDBELIEF, triangle denotes REPORTEDBELIEF, and diamond denotes NONAPPLICABLE. The feature versions with negative weights are associated more with subordinates' messages, whereas those with positive weights are associated more with superiors' messages. Since NCB and ROB versions are rare, they rarely get high weights in the model.

We find that by incorporating belief labels into lexical features, we capture important distinctions in social meanings expressed through words that are lost in the regular lemma ngram formulation. For example, propositional heads such as *know*, *need*, *hold*, *mean* and *want* are indicators of power when they occur in CB contexts (e.g., *i need ...*), whereas their usages in NA contexts (e.g., *do you need?*, *if i need...*, etc.) are indicators of lack of power. In contrast, the CB version of *attend*, *let*, *plan*, *could*, *check*, *discuss*, and *feel* (e.g., *i will attend/check/plan ...*) are strongly associated with lack of power, while their NA versions (e.g., *can you attend/check/plan?*) are indicators of power.

[4]Not all lemma ngrams have all four versions; we calculated standard deviation using the versions present.

7 Conclusion

In this paper, we made two major contributions. First, we presented a large-scale data oriented analysis of how social power relations between participants of an interaction correlate with different types of author commitment in terms of their relative usage of hedges and different levels of beliefs — committed belief, non-committed belief, reported belief, and non-belief. We found evidence that subordinates use significantly more propositional hedges than superiors, and that superiors and subordinates use significantly different proportions of different types of beliefs in their messages. In particular, subordinates use significantly more non-committed beliefs than superiors. They also report others' beliefs more often than superiors. Second, we investigated different ways of incorporating the belief tag information into the machine learning system that automatically detects the direction of power between pairs of participants in an interaction. We devised a sophisticated way of incorporating this information into the machine learning framework by appending the heads of propositions in lexical features with corresponding belief tags, demonstrating its utility in distinguishing social meanings expressed through the different belief contexts.

This study is based on emails from a single corporation, at the beginning of the 21st century. Our findings on the correlation between author commitment and power may be reflective of the work culture that prevailed in that organization at the time when the emails were exchanged. It is important to replicate this study on emails from multiple organizations in order to assess whether these results generalize across board. It is likely that behavior patterns are affected by factors such as ethnic culture (Cox et al., 1991) of the organization, and the kinds of conversations interactants engage in (for instance, co-operative vs. competitive behavior (Hill et al., 1992)). We intend to explore this line of inquiry in future work.

Acknowledgments

This paper is partially based upon work supported by the DARPA DEFT program under a grant to Columbia University; all three co-authors were at Columbia University when portions of this work were performed. The views expressed here are those of the author(s) and do not reflect the official policy or position of the Department of Defense or the U.S. Government. We thank Dan Jurafsky and the anonymous reviewers for their helpful feedback.

References

Apoorv Agarwal, Adinoyi Omuya, Aaron Harnly, and Owen Rambow. 2012. A comprehensive gold standard for the Enron organizational hierarchy. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Jeju Island, Korea, pages 161–165. http://www.aclweb.org/anthology/P12-2032.

Apoorv Agarwal, Adinoyi Omuya, Jingwei Zhang, and Owen Rambow. 2014. Enron corporation: You're the boss if people get mentioned to you. In *Proceedings of the 2014 International Conference on Social Computing*. ACM, New York, NY, USA, SocialCom '14, pages 2:1–2:4. https://doi.org/10.1145/2639968.2640065.

Shashank Agarwal and Hong Yu. 2010. Detecting hedge cues and their scope in biomedical text with conditional random fields. *Journal of biomedical informatics* 43(6):953–961.

Robert F. Bales. 1970. *Personality and interpersonal behavior*. Holt, Rinehart, and Winston (New York).

Robert F. Bales, Fred L. Strodtbeck, Theodore M. Mills, and Mary E. Roseborough. 1951. Channels of communication in small groups. *American Sociological Review* pages 16(4), 461–468.

Or Biran, Sara Rosenthal, Jacob Andreas, Kathleen McKeown, and Owen Rambow. 2012. Detecting influencers in written online conversations. In *Proceedings of the Second Workshop on Language in Social Media*. Association for Computational Linguistics, Montréal, Canada, pages 37–45. http://www.aclweb.org/anthology/W12-2105.

Pierre Bourdieu and John B. Thompson. 1991. *Language and symbolic power*. Harvard University Press.

David B. Bracewell, Marc Tomlinson, and Hui Wang. 2012. A motif approach for identifying pursuits of power in social discourse. In *ICSC*. IEEE Computer Society, pages 1–8.

Philip Bramsen, Martha Escobar-Molano, Ami Patel, and Rafael Alonso. 2011. Extracting Social Power Relationships from Natural Language. In *ACL*. The Association for Computational Linguistics, pages 773–782.

Linda L. Carli. 1990. Gender, language, and influence. *Journal of Personality and Social Psychology* 59(5):941.

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*. Association for Computational Linguistics, pages 70–79.

Taylor H. Cox, Sharon A. Lobel, and Poppy Lauretta McLeod. 1991. Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. *Academy of management journal* 34(4):827–847.

Germán Creamer, Ryan Rowe, Shlomo Hershkop, and Salvatore J. Stolfo. 2009. Segmentation and automated social hierarchy detection through email network analysis. In Haizheng Zhang, Myra Spiliopoulou, Bamshad Mobasher, C. Lee Giles, Andrew Mccallum, Olfa Nasraoui, Jaideep Srivastava, and John Yen, editors, *Advances in Web Mining and Web Usage Analysis*, Springer-Verlag, Berlin, Heidelberg, pages 40–58.

Faye Crosby and Linda Nyquist. 1977. The female register: An empirical study of lakoff's hypotheses. *Language in Society* 6(03):313–322.

Robert A. Dahl. 1957. The concept of power. *Syst. Res.* 2(3):201–215. https://doi.org/10.1002/bs.3830020303.

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of Power: Language Effects and Power Differences in Social Interaction. In *Proceedings of the 21st international conference on World Wide Web*. ACM, New York, NY, USA, WWW '12. https://doi.org/10.1145/2187836.2187931.

Chrysanne Di Marco, Frederick W. Kroon, and Robert E. Mercer. 2006. Using hedges to classify citations in scientific articles. In *Computing attitude and affect in text: theory and applications*, Springer, pages 247–263.

Mona Diab, Lori Levin, Teruko Mitamura, Owen Rambow, Vinodkumar Prabhakaran, and Weiwei Guo. 2009. Committed Belief Annotation and Tagging. In *Proceedings of the Third Linguistic Annotation Workshop*. Association for Computational Linguistics, Suntec, Singapore, pages 68–73. http://www.aclweb.org/anthology/W09-3012.

Jana Diesner and Kathleen M. Carley. 2005. Exploration of communication networks from the Enron email corpus. In *In Proc. of Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining 2005*. pages 21–23.

Richard M. Emerson. 1962. Power-Dependence Relations. *American Sociological Review* 27(1):31–41. https://doi.org/10.2307/2089716.

Norman Fairclough. 2001. *Language and power*. Pearson Education.

Richárd Farkas, Veronika Vincze, György Móra, János Csirik, and György Szarvas. 2010. The conll-2010 shared task: Learning to detect hedges and their scope in natural language text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, Uppsala, Sweden, pages 1–12. http://www.aclweb.org/anthology/W10-3001.

John R. French and Bertram Raven. 1959. The Bases of Social Power. In Dorwin Cartwright, editor, *Studies in Social Power*, University of Michigan Press, pages 150–167+.

Eric Gilbert. 2012. Phrases that Signal Workplace Hierarchy. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, CSCW '12, pages 1037–1046.

Charles W.L. Hill, Michael A. Hitt, and Robert E. Hoskisson. 1992. Cooperative versus competitive structures in related and unrelated diversified firms. *Organization Science* 3(4):501–521.

Ken Hyland. 1998. *Hedging in scientific research articles*, volume 54. John Benjamins Publishing.

Halil Kilicoglu and Sabine Bergler. 2008. Recognizing speculative language in biomedical research articles: a linguistically motivated perspective. *BMC bioinformatics* 9(Suppl 11):S10.

Robin Lakoff. 1973. Language and Woman's Place. *Language in society* 2(01):45–79.

Miriam A. Locher. 2004. *Power and politeness in action: disagreements in oral communication*. Language, power, and social process. M. de Gruyter. http://books.google.com/books?id=Aa32A4gWb8sC.

Robert E. Mercer, Chrysanne Di Marco, and Frederick W. Kroon. 2004. The frequency of hedging cues in citation contexts in scientific writing. In *Advances in artificial intelligence*, Springer, pages 75–88.

Roser Morante, Vincent Van Asch, and Walter Daelemans. 2010. Memory-based resolution of in-sentence scopes of hedge cues. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task*. Association for Computational Linguistics, pages 40–47.

Jr. Galileo Mark S. Namata, Lise Getoor, and Christopher P. Diehl. 2007. Inferring organizational titles in online communication. In *Proceedings of the 2006 conference on Statistical network analysis*. Springer-Verlag, Berlin, Heidelberg, ICML'06, pages 179–181. http://dl.acm.org/citation.cfm?id=1768341.1768359.

Sik Hung Ng and James J. Bradac. 1993. *Power in language: Verbal communication and social influence.*. Sage Publications, Inc.

William M. O'Barr. 1982. *Linguistic evidence: language, power, and strategy in the courtroom*. Studies on law and social control. Academic Press. http://books.google.com/books?id=bqO0PwAACAAJ.

William M. O'Barr and Bowman K. Atkins. 1980. "women's language" or "powerless language"? *Women and Language in Literature and Society* .

Jeffrey Pfeffer. 1981. *Power in organizations*. Pitman, Marshfield, MA.

Vinodkumar Prabhakaran, Tomas By, Julia Hirschberg, Owen Rambow, Samira Shaikh, Tomek Strzalkowski, Jennifer Tracey, Michael Arrigo, Rupayan Basu, Micah Clark, Adam Dalton, Mona Diab, Louise Guthrie, Anna Prokofieva, Stephanie Strassel, Gregory Werner, Janyce Wiebe, and Yorick Wilks. 2015. A New Dataset and Evaluation for Belief/Factuality. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics (*SEM 2015)*. Association for Computational Linguistics, Denver, USA.

Vinodkumar Prabhakaran, Ajita John, and Dorée D. Seligmann. 2013. Who Had the Upper Hand? Ranking Participants of Interactions Based on Their Relative Power. In *Proceedings of the IJCNLP*. Asian Federation of Natural Language Processing, Nagoya, Japan, pages 365–373. http://www.aclweb.org/anthology/I13-1042.

Vinodkumar Prabhakaran and Owen Rambow. 2013. Written Dialog and Social Power: Manifestations of Different Types of Power in Dialog Behavior. In *Proceedings of the IJCNLP*. Asian Federation of Natural Language Processing, Nagoya, Japan, pages 216–224. http://www.aclweb.org/anthology/I13-1025.

Vinodkumar Prabhakaran and Owen Rambow. 2014. Predicting power relations between participants in written dialog from a single thread. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Baltimore, Maryland, pages 339–344. http://www.aclweb.org/anthology/P14-2056.

Vinodkumar Prabhakaran, Owen Rambow, and Mona Diab. 2010. Automatic Committed Belief Tagging. In *Coling 2010: Posters*. Coling 2010 Organizing Committee, Beijing, China, pages 1014–1022. http://www.aclweb.org/anthology/C10-2117.

Vinodkumar Prabhakaran, Owen Rambow, and Mona Diab. 2012. Who's (Really) the Boss? Perception of Situational Power in Written Interactions. In *24th International Conference on Computational Linguistics (COLING)*. Association for Computational Linguistics, Mumbai, India.

Bent Preisler. 1986. Linguistic sex roles in conversation. *Berlin: Mouton de Gruyter* .

Anna Prokofieva and Julia Hirschberg. 2014. Hedging and speaker commitment. In *5th International Workshop on Emotion, Social Signals, Sentiment and Linked Open Data. LREC*.

Sara Rosenthal. 2014. Detecting Influencers in Social Media Discussions. *XRDS: Crossroads, The ACM Magazine for Students* 21(1):40–45.

Liliana Mamani Sanchez and Carl Vogel. 2015. A hedging annotation scheme focused on epistemic phrases for informal language. In *Proceedings of the IWCS Workshop on Models for Modality Annotation*. Association for Computational Linguistics, London, UK.

Roser Saurí and James Pustejovsky. 2009. FactBank: a corpus annotated with event factuality. *Language Resources and Evaluation* 43:227–268. 10.1007/s10579-009-9089-9. http://dx.doi.org/10.1007/s10579-009-9089-9.

Jitesh Shetty and Jafar Adibi. 2005. Discovering important nodes through graph entropy the case of Enron email database. In *Proceedings of the 3rd international workshop on Link discovery*. ACM, New York, NY, USA, LinkKDD '05, pages 74–81. https://doi.org/http://doi.acm.org/10.1145/1134271.1134282.

Tomek Strzalkowski, George Aaron Broadwell, Jennifer Stromer-Galley, Samira Shaikh, Sarah Taylor, and Nick Webb. 2010. Modeling sociocultural phenomena in discourse. In *Proceedings of the 23rd International Conference on COLING 2010*. Coling 2010 Organizing Committee, Beijing, China. http://www.aclweb.org/anthology/C10-1117.

Swabha Swayamdipta and Owen Rambow. 2012. The Pursuit of Power and Its Manifestation in Written Dialog. *2012 IEEE Sixth International Conference on Semantic Computing* 0:22–29. https://doi.org/http://doi.ieeecomputersociety.org/10.1109/ICSC.2012.49.

Sarah M. Taylor, Ting Liu, Samira Shaikh, Tomek Strzalkowski, George Aaron Broadwell, Jennifer Stromer-Galley, Umit Boz, Xiaoai Ren, Jingsi Wu, and Feifei Zhang. 2012. Chinese and American Leadership Characteristics: Discovery and Comparison in Multi-party On-Line Dialogues. In *ICSC*. IEEE Computer Society, pages 17–21.

Teun A Van Dijk. 1989. Structures of discourse and structures of power. *Annals of the International Communication Association* 12(1):18–59.

Erik Velldal, Lilja Øvrelid, Jonathon Read, and Stephan Oepen. 2012. Speculation and negation: Rules, rankers, and the role of syntax. *Computational Linguistics* 38(2):369–410.

Sumithra Velupillai. 2012. *Shades of certainty: annotation and classification of swedish medical records*. Ph.D. thesis, Department of Computer and Systems Sciences, Stockholm University.

Veronika Vincze. 2014. Uncertainty Detection in Hungarian Texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin City University and Association for Computational Linguistics, Dublin, Ireland, pages 1844–1853. http://www.aclweb.org/anthology/C14-1174.

Thomas E. Wartenberg. 1990. *The forms of power: from domination to transformation*. Temple University Press. http://books.google.sh/books?id=yK52QgAACAAJ.

Zhongyu Wei, Junwen Chen, Wei Gao, Binyang Li, Lanjun Zhou, Yulan He, and Kam-Fai Wong. 2013. An empirical study on uncertainty identification in social media context. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Sofia, Bulgaria, pages 58–62. http://www.aclweb.org/anthology/P13-2011.

Alexander Yeh. 2000. More accurate tests for the statistical significance of result differences. In *Proceedings of the 18th conference on Computational linguistics - Volume 2*. Association for Computational Linguistics, Stroudsburg, PA, USA, COLING '00, pages 947–953. https://doi.org/http://dx.doi.org/10.3115/992730.992783.

Jen-Yuan Yeh and Aaron Harnly. 2006. Email Thread Reassembly Using Similarity Matching. In *CEAS 2006 - The Third Conference on Email and Anti-Spam, July 27-28, 2006, Mountain View, California, USA*. Mountain View, California, USA.