

Written Dialog and Social Power: Manifestations of Different Types of Power in Dialog Behavior

Vinodkumar Prabhakaran

Department of Computer Science
Columbia University
New York, NY
vinod@cs.columbia.edu

Owen Rambow

Center for Computational Learning Systems
Columbia University
New York, NY
rambow@ccls.columbia.edu

Abstract

Dialog behavior is affected by power relations among the discourse participants. We show that four different types of power relations (hierarchical power, situational power, influence, and power over communication) affect written dialog behavior in different ways. We also present a system that can identify power relations given a written dialog.

1 Introduction

The recent increase in online social interactions has triggered great interest in computationally analyzing such interactions to gain insights about the discourse participants (DPs). Within the field of analyzing online interactions, there is a growing interest in finding how social power relations between participants are reflected in the various facets of interactions, and whether the power relations can be detected using computational means (Rowe et al., 2007; Bramsen et al., 2011). More recent work has shown that an analysis of the dialog structure (and not just the message content) helps detecting power relations (Biran et al., 2012; Danescu-Niculescu-Mizil et al., 2012).

Understanding the relation between dialog and power may help in various applications. For example, if a dialog system is engineered to behave appropriately given the user’s expectation of relative power (for different types of power), then the user may experience the interaction with the system as more natural. Turning to dialog analysis rather than generation, we can build a computational system to analyze power relations between participants in an interaction. Such a system could have various applications. Power analysis in online forums and communities could be useful in

determining relevance to a user searching the forum. For example, a user may want to limit his search to posts authored by the DPs with higher power. Power analysis may also aid law enforcement agencies to detect leaders and influencers in suspicious online communities. This is especially useful since the real identities of the members of such communities are often not revealed and their hierarchies may not be available to the law enforcement agencies.

The power differential between the DPs may be based on a multitude of factors such as status, authority, role, knowledge and so on. Early computational approaches to analyzing power in interactions relied solely on static power structures such as corporate hierarchies as the source of the power differential (Rowe et al., 2007; Bramsen et al., 2011). More recent studies have looked into dynamic notions of power as well, such as influence (Biran et al., 2012). However, not much work has been done to understand how different types of power differ in the ways they affect how people interact in dialog.

In this paper, we study four different types of power — hierarchical power, situational power, influence and power over communication. We investigate whether all four social power relations are manifested in dialog behavior; we restrict our attention to written dialog, specifically email exchanged in an American corporation. By “dialog behavior”, we mean the choices a DP makes while engaging in dialog. Dialog behavior includes choices that affect dialog structure, such as the choice of when to participate (e.g., does the DP initiate the dialog?), how much to contribute (e.g., is the DP terse or loquacious?), what sort of contribution to make (e.g., which dialog acts does the DP perform? how does the contribution link

to previous dialog contributions?), and what form the contribution should take (e.g., whether to make an overt display of power). The main contribution of this paper is to show that the four types of power we consider are in fact different from one another and that they affect the DPs' behavior in written dialog in different but predictable ways. We analyze these manifestations in the language as well as the dialog structure of interactions. We also present a system to detect the DPs with one of these types of power from threaded email interactions.

In Section 2, we discuss related work in the field. Section 3-4 presents the data, annotations, and inter-rater agreement studies on the annotations. Section 5 summarizes the dimensions of interactions we analyze. We then present the main contributions of this paper: Section 6 analyzes the variations in the manifestations of power among the four types, and Section 7 describes a system to predict persons with any of the four types of power. We then conclude and discuss future work.

2 Related Work

Within the dialog community, researchers have studied notions of control and initiative in dialogs (e.g. (Walker and Whittaker, 1990; Jordan and Di Eugenio, 1997)). Walker and Whittaker (1990) define "control of communication" in terms of whether the discourse participants are providing new, unsolicited information. They use utterance level rules to determine which discourse participant (whether the speaker or the hearer) is in control, and extend it to segments of discourse. Their notion of control differs from our notion of power over communication. They model control locally over discourse segments. What we are interested in (and what our annotations capture) is the possession of controlling power by one (or more) participant(s) across the entire dialog, i.e. how a participant controls the communication in a dialog thread in order to achieve its intended goals. Despite this difference in definition, we show in Section 6 that our notion of power over communication correlates with Walker and Whittaker (1990)'s notion of control over discourse segments. Jordan and Di Eugenio (1997) suggest that "initiative" applies to the level of problem solving, just as "control" applies to the dialog level. We leave the investigation into the relation between initiative and situational power for future work.

In social sciences, different typologies of power have been proposed. Wartenberg (1990) makes the distinction between power-over and power-to in the context of interactions. Power-over refers to relationships between interactants set by external power structures, while power-to refers to the ability an interactant possesses within the interaction, even if it is temporary. Our notions of hierarchical power and influence are special cases of power-over. Hierarchical power is determined by organizational hierarchy, while influence is determined by knowledge, expertise etc. Similarly, our notions of situational power and power over communication are special cases of power-to. Situational power applies to the situation or task at hand, while power over communication applies to the interaction itself. French and Raven (1959) proposed five bases of power: Coercive, Reward, Positional, Referent, and Expert. They are widely used to study power in sociology. We consider hierarchical power, situational power and power over communication to be positional in nature; although the former two can also have bases in coercion and rewards. The bases of influence are mainly referent and expert power.

Studies in sociolinguistics have also explored the relation between dialog behavior and social power. O'Barr (1982) shows that power relations are manifested in language use in courtroom dialogs. Locher (2004) studies politeness in dialogs in relation to the exercise of power. The correlation between discourse structure and perceived influence of participants has also been studied (Ng et al., 1993; Ng et al., 1995). Specifically, factors such as frequency of contribution, proportion of turns, and number of successful interruptions have been identified as important indicators of influence (Reid and Ng, 2000). This work was done entirely on spoken dialog. In our work, we show that the core insight — conversation is a resource for influence — carries over to written dialog; we also show that it carries over to other forms of power. However, some of the characteristics of spoken dialog do not carry over directly to written dialog, most prominently among them the issue of interruptions: there is no interruption in written dialog.

We now look at various computational approaches to extract power relations from online dialogs. Several studies have used Social Network Analysis (e.g., (Rowe et al., 2007)) to extract social relations from online communication.

Researchers have also applied NLP techniques on message content to detect power relations. Earlier approaches used simple lexical features (e.g. (Bramsen et al., 2011; Gilbert, 2012)) while later studies have performed deeper discourse analysis and used features such as linguistic coordination (Danescu-Niculescu-Mizil et al., 2012), language uses such as attempts to persuade and various other dialog patterns (Biran et al., 2012). We present a more detailed discussion of the above mentioned studies and how they differ from our line of research in (Prabhakaran et al., 2012c).

Our research also falls into the category of studies that go beyond pure lexical features and use dialog structure based features to extract social power relations. In (Prabhakaran et al., 2012c), we studied the notion of situational power in depth and presented a system to detect persons with situational power using dialog features. In this paper as well, we use the system described in (Prabhakaran et al., 2012c). However, this work differs from (Prabhakaran et al., 2012c) and other studies described above in that our focus is on how different types of power are manifested differently in the dialog behavior of the participants. We show that the types of power we consider are in fact different and vary in the ways they manifest in dialogs (Section 6). We also present a system that predicts different types of power (Section 7), not just hierarchical or situational power.

3 Data and Annotations

We use the subset of the Enron email corpus with power annotations presented in Prabhakaran et al. (2012a) for our experiments. The corpus also contains manual dialog act annotations by Hu et al. (2009), which enable us to perform the analysis of how power affects dialog behavior. The corpus contains 122 email threads with a total of 360 messages and 20,740 word tokens. There are about 8.5 participants per thread. There are 221 active participants (participants of a thread who has sent at least one email message in the thread) in the corpus. Table 1 presents the counts and percentages of active participants with each type of power in the corpus. We now define the four types of power we investigate in this paper.

Hierarchical Power (HP): We use the gold organizational hierarchy for Enron released by Agarwal et al. (2012) to model hierarchical power. It contains relations between 1,518 employees, and

Type of power	Count	Percentage
Hierarchical Power (HP)	18	8.1
Situational Power (SP)	81	36.7
Power over Communication (PC)	127	57.5
Influence (INFL)	11	5.0

Table 1: Annotation statistics

13,724 dominance pairs (pairs of employees such that the first dominates the second in the hierarchy, not necessarily immediately). We labeled a participant to have hierarchical power within a thread if there exist a dominance pair in the gold hierarchy such that he/she dominates any other participant in the same thread.

For the other three types of power — situational power, power over communication, and influence, we utilize the manual annotations present in the corpus of (Prabhakaran et al., 2012a).¹ We labeled a participant to have one of these types of power within a thread if he or she was judged to have that type of power over any other participant in the same thread. We explain the annotations in detail below with an example thread and corresponding annotations shown in Table 2; the email body contains dialog act and link annotations in [square brackets] which will be explained in Section 3.1.

Situational Power (SP): Person₁ is said to have situational power over person₂ if person₁ has power or authority to direct and/or approve person₂’s actions in the current situation or while a particular task is being performed, based on the communication in the current thread. Situational power is independent of organizational hierarchy: person₁ with situational power may or may not be above person₂ in the organizational hierarchy (or there may be no organizational hierarchy at all). In our example thread, our annotator judged Kathryn to possess situational power over Leslie, Sara and Brent because Kathryn is following up on and assigning a task to others, and because Kathryn uses language that shows that she is in charge of the situation.

Power over Communication (PC): A person is said to have power over communication if he actively attempts to achieve the intended goals of the communication.² These are people who ask questions, request others to take action, etc., and not

¹The manual annotations also capture the perception of hierarchical power. In this work, we use only the actual gold hierarchy (Agarwal et al., 2012) as described above.

²In (Prabhakaran et al., 2012a), power over communication was called “control of communication”.

From: Kathryn Cordes To: Leslie Hansen, Sara Shackleton, Brent Hendry CC: Mark Greenberg, Erik Eller, Thomas D Gros	
M1.1. Leslie Sara, and Brent: [Conventional]	
M1.2. Could I get an update on were we are with the top 20 customer amendments. [Req-Info]; [Flink1.2]	
M1.3. Last week we got 5 amendments for power physical but we still haven't received any amendments for financial. [Inform]	
M1.4. Entergy-Koch is very interested in ConfirmLogic and have asked for the amendments. [Inform]	
M1.5. When can we get the amendments for Entergy-Koch completed? [Req-Info]; [Flink1.5]	
M1.6. Thanks, [Conventional]	
M1.7. KC [Conventional]	
From: Brent Hendry To: Mark Taylor	
M2.1. I have just finished the draft for our internal legal review and sent it around. [Inform]	
M2.2. There are still a lot of work to be done but I do not know when everyone will have time to look at this considering how much other work there is. [Inform]	
M2.3. How should we respond considering she has copied Tom Gros? [Req-Info]; [Blink1.5]; [Flink2.3]	
Person with SP	Kathryn Cordes
Person with PC	N/A
Person with INFL	Mark Taylor
Overt Display of Power	M1.2

Table 2: Example thread with power annotations

people who simply respond to questions or perform actions when directed to do so. There could be multiple such participants in a given thread. In our example thread, no one was judged to have power over communication since the communication is broken into two separate interactions of just one message each — one from Kathryn to everyone and the other between Brent and Mark.

Influence (INFL): A person is defined to have influence if she 1) has credibility in the group, 2) persists in attempting to convince others, even if some disagreement occurs, 3) introduces topics/ideas that others pick up on or support, and 4) is a group participant but not necessarily active in the discussion(s) where others support/credit her. In addition, the influencer's ideas or language may be adopted by others and others may explicitly recognize influencer's authority. In our example,

our annotator judged Mark to have influence over Brent since the latter seeks advice from the former on how to deal with the situation.

3.1 Dialog Act Annotations

The corpus we used contains manual dialog act annotations as described in Hu et al. (2009). We use these annotations to model the dialog structure of the communication thread. For each message, Hu et al. (2009) assign a Dialog Act (DA) label to each segment of text with a coherent communicative function. The label could be one of the following: ReqAction, ReqInfo, Inform, InformOffline,³ Conventional, and Commit. In addition, the segments are linked by three types of links to reflect the dialog structure. These links capture the patterns of local alternation between an initiating dialog act and a responding one. A **forward link** (Flink) is the analog of a “first pair-part” of an adjacency pair, is restricted to ReqInfo and ReqAction segments. The responses to such requests are assigned a **backward link** (Blink). If an utterance can be interpreted as a response to a preceding segment, it gets a Blink even where the preceding segment has no Flink. The preceding segment taken to be the “first pair-part” of the link is assigned a **secondary forward link** (SFlink).

3.2 Overt Display of Power

Our corpus also contains the **overt display of power** (ODP) (Prabhakaran et al., 2012b) annotations. An utterance is defined to have an ODP if it is interpreted as creating additional constraints on the response beyond those imposed by the general dialog act. Syntactically, an ODP can be an imperative, a question, or a declarative sentence. In our example thread, utterance M1.2 is an instance of ODP. The inter-annotator agreement value (κ) of ODP annotations was 0.67.

4 Reliability of Annotations

The power annotations in the corpus are performed by a single annotator and capture her perception of the overall power structure among the participants of the interaction. To verify the reliability of these annotations, we performed an independent inter-annotator agreement (IAA) study on a subset of 47 threads from the corpus. We trained

³Sometimes, the Inform act refers to a previous act of communication which did not happen in the email thread itself. Such cases are marked as Offline.

two annotators — AnnA and AnnB — using the same annotation manual described in (Prabhakaran et al., 2012a) and compared the annotations they produced on the selected threads. Annotators were asked to read the entire thread before performing the annotations. They are also asked to provide, in free-form English, a short “power narrative” which describes their perception of the overall power structure among the discourse participants of that thread. Annotators build a fairly consistent mental image of a power narrative — an outline of the power structure between the participants — based on various indicators from across the thread. Their individual power annotations are based on this power narrative. Hence, the cognitive process behind labeling a participant to have a particular type of power is not a binary decision the annotator makes for each participant. However, evaluating agreement on such a formulation is not straightforward. Thus, for the purpose of this IAA study, we port this task into a binary decision task of identifying whether participant X has power of type P or not.

There were 289 participants in the selected 47 threads. The κ values obtained for each type of power is shown in Table 3 under Round 1. Since the κ values obtained in round 1 were only fair to moderate, we performed another round of training and inter annotator study. For this round, AnnB was not available, and we hired another annotator AnnC. The κ values obtained between AnnA and AnnC on another set of 10 threads is presented in Table 3 under Round 2.

Type of power	Round 1	Round 2
Situational Power (SP)	0.47	0.47
Power over Communication (PC)	0.27	0.76
Influence (INFL)	0.50	0.79

Table 3: Inter Rater Agreement (κ)

The κ values obtained in both round 1 and round 2 are in the range of those previously reported for similar tasks (e.g., 0.18 for managerial influence and 0.52 for establishing solidarity (Bracewell et al., 2012); 0.72 for influence (Biran et al., 2012)). The agreement in round 2 improved considerably for both PC and INFL after the second round of training. The issue of moderate agreement for SP and its possible reasons are discussed in detail in (Prabhakaran et al., 2012c). For the rest of this paper, we use the original annotations that were present in the corpus.

5 Dialog Behavior

We use five sets of features to capture the dialog behavior of participants: dialog act percentages (**DAP**), dialog link counts (**DLC**), positional (**PST**), verbosity (**VRB**), and overt displays of power (**ODP**). The specific features within each set are listed in Table 4. PST and VRB are readily derivable from the data, without any annotations.

Set	Features
DAP	ReqAction, ReqInform, Inform, InformOffline, Conventional, Commit
DLC	Flink, Slink, Blink, Clink, Dlink, DlinkRatio
PST	Initiator, FirstMsg, LastMsg
VRB	MsgCount, MsgRatio, TokenCount, TokenRatio, TokensPerMsg
ODP	ODPCount

Table 4: Feature Sets

DAP captures the percentages of each dialog act labels in each participant’s utterances. DLC captures the metrics on various kinds of links in each participant’s messages. Flink, Slink and Blink corresponds to counts of respective link annotations in participants’ messages. We refer to Flinks with one or more backward links as connected links (Clink) and those with no matching Blink as dangling links (Dlink). A dangling link denotes a request that was ignored. The DlinkRatio is the ratio of Dlinks to Flinks for a participant. This captures what percentage of a participant’s requests went unanswered. PST captures the positions within the thread where the participant joined and left the conversation. Initiator is a binary feature capturing whether the participant initiated the thread or not. FirstMsg and LastMsg are real valued features between 0 and 1, capturing the relative position of the first and last messages by the participant. VRB features are self explanatory. ODP captures the number of instances of ODP in the messages by each participant.

6 Variations in Manifestations of Power

In this section, we present the results of a statistical analysis of the dialog features with respect to people with the four types of power. For each type of power (HP, SP, INFL and PC), we consider two populations of people who participated in the dialog: \mathcal{P} , those judged to have that type of power, and \mathcal{N} , those not judged to have that power. Then, for each feature, we perform a two-sample, two-tailed t-test comparing means of feature values of

Set	Features	HP	SP	PC	INFL
DAP	ReqAction	0.10 0.02 _{0.23}	0.07 0.01 _{0.01}	0.03 0.04 _{0.48}	0.0 0.04 _{6.9E-5}
	ReqInform	0.10 0.11 _{0.87}	0.10 0.12 _{0.70}	0.11 0.11 _{0.91}	0.09 0.11 _{0.73}
	Inform	0.56 0.60 _{0.63}	0.56 0.63 _{0.10}	0.60 0.61 _{0.79}	0.78 0.59 _{0.01}
	InformOffline	0.00 0.005 _{0.04}	0.003 0.005 _{0.62}	0.008 0.0 _{0.04}	0.0 0.005 _{0.04}
	Conventional	0.23 0.24 _{0.96}	0.25 0.23 _{0.35}	0.24 0.23 _{0.81}	0.13 0.24 _{0.04}
	Commit	0.0 0.002 _{0.21}	0.001 0.003 _{0.51}	0.001 0.004 _{0.44}	0.0 0.002 _{0.21}
DLC	Flink	0.56 0.74 _{0.27}	0.98 0.59 _{0.03}	0.91 0.49 _{6.2E-3}	0.45 0.74 _{0.35}
	SFlink	0.16 0.34 _{0.09}	0.49 0.24 _{0.02}	0.43 0.21 _{0.01}	0.64 0.32 _{0.07}
	Blink	0.94 0.61 _{0.23}	0.72 0.59 _{0.40}	0.41 0.94 _{1.7E-4}	1.00 0.61 _{0.39}
	Clink	0.27 0.61 _{0.04}	0.83 0.44 _{7.1E-3}	0.75 0.35 _{6.9E-4}	0.73 0.57 _{0.46}
	Dlink	0.44 0.49 _{0.79}	0.64 0.39 _{0.08}	0.58 0.35 _{0.06}	0.36 0.49 _{0.67}
	DlinkRatio	0.39 0.24 _{0.24}	0.33 0.21 _{0.05}	0.27 0.24 _{0.57}	0.18 0.26 _{0.55}
PST	Initiator	0.27 0.57 _{0.02}	0.68 0.48 _{3.3E-3}	0.88 0.11 _{3.4E-44}	0.64 0.55 _{0.58}
	FirstMsg	0.34 0.19 _{0.02}	0.13 0.24 _{1.1E-3}	0.05 0.40 _{1.4E-28}	0.16 0.21 _{0.55}
	LastMsg	0.47 0.37 _{0.08}	0.41 0.36 _{0.21}	0.31 0.47 _{1.9E-5}	0.32 0.38 _{0.51}
VRB	MsgCount	1.33 1.46 _{0.47}	1.68 1.32 _{0.03}	1.62 1.22 _{1.3E-3}	1.45 1.45 _{0.99}
	MsgRatio	0.48 0.52 _{0.47}	0.54 0.50 _{0.18}	0.61 0.39 _{2.8E-15}	0.45 0.52 _{0.19}
	TokenCount	53.22 91.53 _{0.06}	113.04 74.19 _{0.02}	121.38 43.90 _{1.1E-8}	143.55 85.54 _{0.10}
	TokenRatio	0.35 0.54 _{0.04}	0.62 0.47 _{2.1E-3}	0.72 0.26 _{1.0E-28}	0.63 0.52 _{0.26}
	TokensPerMsg	39.73 63.45 _{0.13}	73.22 54.76 _{0.07}	78.27 38.91 _{1.3E-5}	118.94 58.52 _{0.09}
ODP	ODPCount	0.50 0.36 _{0.30}	0.78 0.14 _{6.0E-8}	0.49 0.21 _{2.6E-3}	0.09 0.39 _{0.01}

Table 5: Variations in manifestations of power on feature values: $\text{mean}(\mathcal{P}) | \text{mean}(\mathcal{N})_{p\text{-value}}$
 \mathcal{P} : people judged to have power; \mathcal{N} : people judged not to have power; Values with $p \leq 0.05$ are boldfaced
Types of power - SP: Situational power, HP: Hierarchical power, PC: Power over Communication, INFL: Influence;
Features - DAP: Dialog acts, DLC: Dialog links, PST: Positional, VRB: Verbosity, ODP: Overt display of power

\mathcal{P} and \mathcal{N} . Table 5 presents means of each feature value for both populations \mathcal{P} and \mathcal{N} (as “ $\text{mean}(\mathcal{P}) | \text{mean}(\mathcal{N})$ ”) along with the p-value associated with the t-test as the subscript. For $p < 0.05$, we reject the null hypothesis and consider the feature to be statistically significant (boldfaced in Table 5).

We find many features which are statistically significant, which suggests that power types are reflected in the dialog structure. The t-test results also show that significance of features differ considerably from one type of power to another, which suggests that different power types are reflected differently in the dialog structure, and that they are thus indeed different types of power.

For HP, we find that people with HP are less active in threads than those without. For example, persons with hierarchical power tend to talk less within a thread (TokenRatio). They tend to start participating much later in the threads (FirstMsg) and do not initiate threads often (Initiator). SP and PC manifest in stark contrast from HP. Persons with SP and persons with PC both tend to talk more within a thread (TokenRatio). They also tend to be the initiators of the thread (Initiator) or start participating in the thread closer to the beginning (FirstMsg). SP and PC have many other features which are also statistically significant. For

example, they send significantly more messages (MsgCount). They also have significantly more instances of overt displays of power (ODPCount) than others. It is interesting to note that ODPCount was not a significant feature for HP. It suggests that bosses don’t always display their power overtly when they interact. SP and PC also differ from one another. For example, those with SP tend to request actions (ReqAction) significantly more than those without. However, this was not significant in case of PC. Similarly, the number of back links (Blink) was not a significant feature for SP. But, people with PC tend to have significantly fewer back links (Blink) than those without.

This finding — people with PC have fewer back links — is interesting, since it aligns PC with the characterization of control by Walker and Whittaker (1990). According to them, control over a discourse segment is determined by whether the participant provide unsolicited information in the dialog or not. In the dialog act annotation scheme we use, solicited information (in other words, responses to requests and commands) places an obligatory Blink on the corresponding text segment. Hence, the fact that people with PC have significantly larger contributions to the dialog (VRB features), but with fewer back

links, suggest that most of their contribution is unsolicited information. This is in line with Walker and Whittaker (1990)’s definition of control over discourse segments.

Although INFL has fewer data points, we found a few significant features for INFL. People with INFL never request actions (ReqAction) as opposed to those with SP who request actions more frequently than others. Also, people with INFL tend to have significantly more inform utterances (Inform). They also have significantly fewer overt displays of power (ODPCount) than others, a stark contrast to those with SP and PC.

The statistical measures presented in previous section are exploratory in nature, presenting tests on all combinations of features and power types. We do not draw theoretical conclusions from the specific combination of interactions that are found statistically significant. Hence, we did not apply any corrections for multiple tests in statistical significance for individual features. When we apply, the Bonferroni correction for multiple tests to adjust the p-value for number of test performed (threshold = $0.05/84 = 6.0E-4$), 10 features would still remain statistically significant. Hence the global null hypothesis that the features we considered do not interact with the power types would still be rejected.

7 Predicting Persons with Power

In this section, we present a system to predict whether a person has a given type of power in the context of an email thread. We show that different sets of features are helpful to detect different types of power. We build a separate binary classifier for each power type predicting whether or not a given participant in a communication thread has that type of power or not. Since our dataset is skewed especially for HP & INFL (with very few persons with power), we balanced our dataset by up-sampling minority class instances in the training step. This has proven useful in cases of unbalanced datasets (Japkowicz, 2000). All results presented below have been obtained after balancing the training folds in cross validation; the test folds remain unchanged. We used the tokenizer, POS tagger, lemmatizer and SVMLight (Joachims, 1999) wrapper in the ClearTK (Ogren et al., 2008) package. The ClearTK wrapper for SVMLight internally shifts the prediction threshold based on a posterior probabilistic score calcu-

Type	Feature set	P	R	F
HP	Random	16.6	38.9	11.3
	AlwaysTrue	8.1	100.0	15.0
	LEX	0.0	0.0	0.0
	VRB	16.7	44.4	24.2
	PST	13.8	72.2	23.2
	DAP	16.0	22.2	18.6
	DLC	15.3	61.1	24.4
SP	ODP	15.3	50.0	23.4
	VRB+PST+ODP	20.9	50.0	29.5
	Random	36.7	49.4	42.1
	AlwaysTrue	36.7	100.0	53.6
	LEX	54.9	55.6	55.2
	VRB	43.9	70.4	54.0
	PST	45.1	67.9	54.2
PC	DAP	40.9	75.3	53.0
	DLC	49.6	75.3	59.8
	ODP	71.2	51.9	60.0
	DLC+ODP	59.4	70.4	64.4
	Random	57.5	51.2	54.2
	AlwaysTrue	57.5	100.0	73.0
	LEX	70.2	78.0	73.9
INFL	VRB	78.7	84.3	81.4
	PST	91.8	88.2	90.0
	DAP	60.5	92.9	73.3
	DLC	74.3	81.9	77.9
	ODP	74.6	34.7	47.3
	PST	91.8	88.2	90.0
	Random	5.2	54.6	9.5
INFL	AlwaysTrue	5.0	100.0	9.5
	LEX	0.0	0.0	0.0
	VRB	8.1	81.8	14.8
	PST	4.6	45.5	8.4
	DAP	6.9	63.6	12.4
	DLC	13.7	63.6	22.6
	ODP	6.2	90.9	11.6
DLC	13.7	63.6	22.6	

Table 6: Cross validation results

SP: Situational power, HP: Hierarchical power
PC: Power over communication, INFL: Influence
VRB: Verbosity, PST: Positional, DAP: Dialog acts, DLC: Dialog links, LEX: Lexical, ODP: Overt display of power

lated using Lin et al. (2007)’s algorithm.

We first find the best performing subset of features for each feature set by exhaustive search within the set. Once we have the best subset of each feature set, we do another round of exhaustive search combining best performers of each set to find the overall best performing feature subset. We report micro-averaged (P)recision, (R)ecall and (F)-measure on 5-fold cross validation for each power type. We experimented with a linear kernel and a quadratic kernel; the latter performed better. All results presented in this paper are obtained using a quadratic kernel.

Table 6 shows cross validation results for all

four types of power for each set of features.⁴ The corpus was split into folds at the thread level. We present two simple baseline measures - **Random** and **AlwaysTrue** and a language-based baseline, **LEX**. In the Random baseline, we predict an active participant to have the particular type of power at random. In AlwaysTrue baseline, we always predict an active participant to have power. For LEX, we use only lexical features (unigrams and bigrams) from messages sent by each participant to train the SVM model described above. For each power type, the table also lists (in the last row) the best performing feature subset combination and corresponding results.

HP is hard to predict, which could partly be due to the very small number of positive training examples in the corpus. For the LEX baseline using purely word ngrams, the system did not get any correct predictions. All feature subsets outperformed the other baselines of 11.3% and 15.0% (for Random and AlwaysTrue respectively), and a combination of VRB, PST and ODP gave the best model obtaining an F measure of 29.5%.

For SP, the best performing individual feature sets are ODP and DLC, both at or near 60.0%. While ODP gave a high precision (71.2%) model, DLC gave a high recall (75.3%) model, the combination of both gave the best performing system with an F measure of 64.4%.

For PC, the best single feature was FirstMsg (relative position of first message). This is because the person with the power over communication is almost always the initiator of the thread. Note that the notion of PC is not defined in terms of positional features: annotators were asked to find the participants who “actively attempt to achieve the intended goals of the communication”. It is our finding that those who are in PC were also the ones who did initiate the thread. It is also worth noting that ODP is the worst performer for PC which is in contrast with the case of SP, supporting the claim that these two types of power are in fact different.

INFL is another very hard class to predict, again, possibly partly due to the very small number of positive training examples. The simple baseline F measures were both 9.5, while the LEX did not produce any correct predictions at all. All feature sets except PST outperformed these baseline measures. The best performance was obtained

by DLC with counts of Blinks, Flinks, Dlinks and SFlinks as features.

For assessing statistical significance of F measure improvements over baseline, we used the Approximate Randomness Test (Yeh, 2000). We found the improvements to be statistically significant for SP ($p = 0.001$), HP ($p=0.001$) and PC ($p = 0.01$) with a threshold for significance at $p = 0.05$. However, for INFL, the improvement was not statistically significant ($p = 0.3$). The statistical significance of SP, HP and PC would hold even after applying Bonferroni correction for multiple tests.

8 Conclusion and Future Work

We studied four types of power between participants of written dialog. We have shown that these types of power are manifested very differently with respect to the features we are using, which validates our claim that these are indeed different types of power. We also presented a supervised learning system to predict persons with one of the types of power in written dialog yielding encouraging results. We have shown that dialog features are very significant in predicting power relations in online written communication.

In future work, we intend to try predicting power relations between pairs of participants. It would be interesting to see how dialog features correlate with the other direction of power; that is from a submitter to an exerciser of power. We will investigate the use of additional features related to the dialog participants, such as gender. We will also investigate using a dialog act tagger, link predictor and an ODP tagger to build a fully automatic power predicting system. We would also like to extend this work to other genres of written communication like discussion forums and blogs.

9 Acknowledgments

This paper is based upon work supported by the Johns Hopkins Human Language Technology Center of Excellence and by the DARPA DEFT Program. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government. We also thank several anonymous reviewers, including three SIGDIAL reviewers, for their constructive feedback.

⁴Results for SP were presented in (Prabhakaran et al., 2012c). We present them here for comparison.

References

- A. Agarwal, A. Omuya, A. Harnly, and O. Rambow. 2012. A Comprehensive Gold Standard for the Enron Organizational Hierarchy. In *Proceedings of the 50th Annual Meeting of the ACL (Vol. 2: Short Papers)*, pages 161–165, Jeju Island, Korea, July. ACL.
- O. Biran, S. Rosenthal, J. Andreas, K. McKeown, and O. Rambow. 2012. Detecting influencers in written online conversations. In *Proceedings of the Second Workshop on Language in Social Media*, pages 37–45, Montréal, Canada, June. ACL.
- D. B. Bracewell, M. Tomlinson, and H. Wang. 2012. A motif approach for identifying pursuits of power in social discourse. In *ICSC*, pages 1–8. IEEE Computer Society.
- P. Bramsen, M. Escobar-Molano, A. Patel, and R. Alonso. 2011. Extracting social power relationships from natural language. In *The 49th Annual Meeting of the ACL*, pages 773–782. ACL.
- C. Danescu-Niculescu-Mizil, L. Lee, B. Pang, and J. Kleinberg. 2012. Echoes of power: language effects and power differences in social interaction. In *Proceedings of the 21st international conference on WWW*, New York, NY, USA. ACM.
- J. R. French and B. Raven. 1959. The Bases of Social Power. In Dorwin Cartwright, editor, *Studies in Social Power*, pages 150–167. Univ. of Mich. Press.
- E. Gilbert. 2012. Phrases that signal workplace hierarchy. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, CSCW '12*, pages 1037–1046, New York, NY, USA. ACM.
- J. Hu, R. Passonneau, and O. Rambow. 2009. Contrasting the Interaction Structure of an Email and a Telephone Corpus: A Machine Learning Approach to Annotation of Dialogue Function Units. In *Proceedings of the SIGDIAL 2009 Conference*, London, UK, September. ACL.
- N. Japkowicz. 2000. Learning from imbalanced data sets: Comparison of various strategies. In *AAAI Workshop on Learning from Imbalanced Data Sets*.
- T. Joachims. 1999. Making Large-Scale SVM Learning Practical. In B. Schölkopf, C. J.C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, Cambridge, MA, USA. MIT Press.
- P. W. Jordan and B. Di Eugenio. 1997. Control and initiative in collaborative problem solving dialogues. In *Working Notes of the AAAI Spring Symposium on Comp. Models for Mixed Initiative*, pages 81–84.
- H. Lin, C. Lin, and R. C. Weng. 2007. A Note on Platt’s Probabilistic Outputs for Support Vector Machines. *Mach. Learn.*, 68:267–276, October.
- M. A. Locher. 2004. *Power and politeness in action: disagreements in oral communication*. Language, power, and social process. M. de Gruyter.
- S. H. Ng, D. Bell, and M. Brooke. 1993. Gaining turns and achieving high in influence ranking in small conversational groups. *British Journal of Social Psychology*, pages 32, 265–275.
- S. H. Ng, M Brooke, , and M. Dunne. 1995. Interruption and in influence in discussion groups. *Journal of Language and Social Psychology*, pages 14(4),369–381.
- W. M. O’Barr. 1982. *Linguistic evidence: language, power, and strategy in the courtroom*. Studies on law and social control. Academic Press.
- P. V. Ogren, P. G. Wetzler, and S. Bethard. 2008. ClearTK: A UIMA toolkit for statistical natural language processing. In *Towards Enhanced Interoperability for Large HLT Systems: UIMA for NLP workshop at Language Resources and Evaluation Conference (LREC)*.
- V. Prabhakaran, O. Rambow, and M. Diab. 2012a. Annotations for power relations on email threads. In *Proceedings of the Eighth conference on LREC*, Istanbul, Turkey, May. ELRA.
- V. Prabhakaran, O. Rambow, and M. Diab. 2012b. Predicting Overt Display of Power in Written Dialogs. In *Proceedings of the HLT-NAACL*, Montreal, Canada, June. ACL.
- V. Prabhakaran, O. Rambow, and M. Diab. 2012c. Who’s (Really) the Boss? Perception of Situational Power in Written Interactions. In *Proceedings of the 24th International Conference on COLING*, Mumbai, India. ACL.
- S. A. Reid and S. H. Ng. 2000. Conversation as a resource for in influence: evidence for prototypical arguments and social identification processes. *European Journal of Social Psych.*, pages 30, 83–100.
- R. Rowe, G. Creamer, S. Hershkop, and S.J. Stolfo. 2007. Automated social hierarchy detection through email network analysis. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web Mining and Social Network Anal.* ACM.
- M. Walker and S. Whittaker. 1990. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proceedings of the 28th annual meeting on ACL*, pages 70–78. ACL.
- T. E. Wartenberg. 1990. *The forms of power: from domination to transformation*. Temple Univ. Press.
- A. Yeh. 2000. More accurate tests for the statistical significance of result differences. In *Proceedings of the 18th conference on COLING - Volume 2*, pages 947–953, Stroudsburg, PA, USA. ACL.