

Who Had the Upper Hand? Ranking Participants of Interactions Based on Their Relative Power

Vinodkumar Prabhakaran
Columbia University
New York, NY, USA
vinod@cs.columbia.edu

Ajita John
Avaya Labs
Basking Ridge, NJ, USA
ajita@avaya.com

Dorée D. Seligmann
Avaya Labs
Basking Ridge, NJ, USA
doree@avaya.com

Abstract

In this paper, we present an automatic system to rank participants of an interaction in terms of their relative power. We find several linguistic and structural features to be effective in predicting these rankings. We conduct our study in the domain of political debates, specifically the 2012 Republican presidential primary debates. Our dataset includes textual transcripts of 20 debates with 4-9 candidates as participants per debate. We model the power index of each candidate in terms of their relative poll standings in the state and national polls. We find that the candidates' power indices affect the way they interact with others and the way others interact with them. We obtained encouraging results in our experiments and we expect these findings to carry across to other genres of multi-party conversations.

1 Introduction

Recently, there has been a rapid increase in social interactions being stored on the World Wide Web. In addition to those interactions that are inherently online such as discussion forums and social networks, offline interactions such as broadcast events, debates and speeches are also captured in real time and stored online in repositories such as YouTube and news media outlets. This growing mass of public data representing various modes of interactions enables researchers to computationally analyze social interactions at a scale which was not feasible previously. Within the field of analyzing online social interactions, there is a growing interest to study how the power or status difference between participants is reflected in

the various facets of interactions and if it can be detected using computational means (Diesner and Carley, 2005; Rowe et al., 2007; Bakshy et al., 2011; Bramsen et al., 2011; Biran et al., 2012; Danescu-Niculescu-Mizil et al., 2012).

When people interact with one another, there is often a power differential that affects the way they interact. This differential may be drawn from a multitude of factors such as social status, authority, experience, age etc. Identifying the dominant participants of an interaction through a power ranking system could have various applications. It could help improve effectiveness of advertisements within online communities. For example, targeting an advertisement to powerful and influential members within an online community might increase its effectiveness and reach to the community members. Power analysis can also help in information retrieval systems. Revealing power dynamics within stored interactions could be useful in determining relevance for a user with information needs. For example, a user may want to limit his search to posts authored by interactants with higher power. Power analysis may also aid intelligence 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 the hierarchies of such communities may not be available to the intelligence agencies.

Most computational efforts to analyze or predict power differentials between participants of interactions have relied on static power structures or hierarchies as sources for the power differential (Rowe et al., 2007; Bramsen et al., 2011; Gilbert, 2012). However, many interactions happen outside the context of a pre-defined static power structure or hierarchy. Examples for such interactions

include political debates, online discussions, and email interactions outside organizational boundaries. Although the participants of these interactions may not be part of an established power structure, there is often a power differential between them drawn from various other factors such as popularity, experience, knowledge etc. In such situations, the interaction itself plays an important role as a medium for the interactants to pursue, gain and maintain power over others. Consequently, the manifestations of power in such interactions will also inherently differ from the cases where a hierarchy is present. However, most computational studies on power within interactions have not explored such a dynamic notion of power.

In this paper, we analyze political debates where the power differential is dynamic. Specifically, we analyze the 2012 Republican presidential primary debates. We present an automatic ranking system to rank debate participants in terms of their relative power. We model the power of each candidate in terms of their relative standings in the polls released prior to the debate. We find that the candidates' power indices affect the way they interact with others and the way others interact with them. To our knowledge, our work is the first to do an in-depth computational analysis of the structure of interactions, modeling patterns of interruptions and mentions of participants, in relation to power. Moreover, the domain we study is particularly interesting since the primary objective of the debate participants is to pursue and maintain power over each other, as opposed to operating within a static power structure. Lastly, the findings of this study are note-worthy as they relate to the domain of political debates, an area which has not been well-studied in this fashion before. We will release the dataset with annotations to the research community to drive more research in this direction.

Next, we review the background and related computational work in the area of power analysis. Section 3 presents the domain of presidential debates and details how we model power in this domain. Section 4 presents the data and Section 5 presents the power ranker and describes features, experiments and results.

2 Background and Related Work

Social power and how it affects the ways people behave in interactions have been studied extensively in social sciences and psychology. Bales

and Slater (1955) studied interactions in small group conversations and suggested language as a reflection and resource of power and influence. Later, Bales (1970) identified the importance of the structure of conversations (e.g. frequency of turns) and argued that "to take up time speaking in a small group is to exercise power over the other members for at least the duration of the time taken, regardless of the content". Ng et al. (1993) found that conversational turns gained by interruptions are stronger indicators of power than turns gained otherwise. In further work Ng and Bradac (1993), they argued that the content also plays a role in influence; a view contrary to (Bales, 1970). More indicators of power in the content of interactions were studied later on. Sexton and Helmreich (1999) found linguistic indicators that could help identify relative status between individuals in social interactions. Locher (2004) studies politeness in interactions in relation to the exercise of power. Our work draws inspiration from many of these studies and looks for correlates of power in both the content and structure of interactions.

Due to the easy availability of data, most of these studies have been performed on written interactions, whereas our study is done on spoken interactions. Early computational approaches to analyze power in interactions relied on network-based approaches. There have been several studies using Social Network Analysis (Diesner and Carley, 2005; Rowe et al., 2007) for extracting social relations from emails. These approaches rely on collections of interactions between a set of individuals to build interaction networks and use various centrality metrics on those networks in order to deduce power relations between interactants. These studies mainly use the meta-data about messages: who sent how many messages to whom and when. Researchers have also analyzed the content of messages using NLP techniques to detect power differentials. For example, Bramsen et al. (2011) and Gilbert (2012) utilize a text classification approach and classify messages in the Enron email corpus as messages sent from a superior to a subordinate, and *vice versa*. Both studies model static hierarchical relationships; our work models a dynamic notion of power in interactions happening outside organizational boundaries. Also, the studies described above consider messages or collections of messages in isolation, but not in the context of the entire interaction.

More recently, a deeper analysis of interactions is shown to be useful in detecting power or influence in interactions. Danescu-Niculescu-Mizil et al. (2012) focus on the notion of language coordination — a metric that measures the extent to which a discourse participant adopts another’s language — in relation to various social attributes such as power, gender, etc. They perform their study on Wikipedia discussion forums and Supreme Court hearings — both of which have enforced power structures. Prabhakaran et al. (2012a) analyze the notion of overt displays of power (ODP) in dialog. Prabhakaran et al. (2012b) and Prabhakaran and Rambow (2013) study how the ODP and other dialog act analysis based features of organizational email interactions correlate with different types of power possessed by the participants. Biran et al. (2012) and Bracewell et al. (2012) use lower-level dialog constructs to model power relations. Biran et al. (2012) use dialog constructs such as attempts to persuade, agreement, disagreement and various dialog patterns in order to find influencers in Wikipedia discussion forums and LiveJournal blogs. Bracewell et al. (2012) try to identify participants pursuing power in discussion forums. They devise a set of eight *social acts* which largely overlaps with the dialog constructs used by (Biran et al., 2012).

Our work also falls into the above category of studies in the sense that we also go beyond pure lexical features and use dialog structure based features in our analysis. However, our work differs in few major ways. Firstly, our domain — political debates — contains spoken interactions while most studies discussed above are performed on written interactions (except Danescu-Niculescu-Mizil et al. (2012) which studies Supreme Court hearings). Secondly, in our domain, the primary purpose of the interactions is to pursue and maintain power, while most studies mentioned earlier deal with domains which are task oriented (Enron, Wikipedia and Supreme Court). Thirdly, in our domain, candidates may gain or lose power in the course of interactions, whereas power is more stable in the studies discussed above. Lastly, our interactions are time-bound, in contrast to online discussions such as Wikipedia forums.

We now turn our attention to related computational work on analyzing conversations in our domain of political debates. Rosenberg and Hirschberg (2009) analyze speeches made in the

context of 2004 Democratic presidential primary election and identify lexical and prosodic cues that signal charisma. More recently, Nguyen et al. (2012) analyze 2008 presidential and vice presidential debates to study how speaker identification helps topic segmentation and how candidates exercise control over conversations by shifting topics. While our domain is also presidential debates, our focus is on how the candidate’s power or confidence affects interactions within the debates.

3 Domain: Political Debates

Before the United States presidential election, a series of presidential primary elections are held in each U.S. state by both major political parties (Republican and Democratic) to select their respective presidential nominees. In recent times, it has become customary that candidates of both parties engage in a series of debates prior to and during their respective parties’ primary elections. In this study, we explore how the power differential between the candidates manifests in these debates. Specifically, we use the 20 debates held between May 2011 and February 2012 as part of the 2012 Republican presidential primaries.¹ There were a total of 10 candidates who took part in these primary debates; some of whom participated only in one or two debates. Interactions in these debates are fairly well structured and follow a pattern of the moderator asking questions and the candidates responding, with some disruptions due to interruptions from other candidates.

Presidential debates serve an important role during the election process. It serves as a platform for candidates to discuss their stances on policy issues and contrast them with other candidates’ stances. In addition, it also serves as a medium for the candidates to pursue and maintain power over other candidates. This makes it an interesting domain to investigate how power dynamics between participants are manifested in an interaction. In addition, the 2012 Republican presidential election campaign was one of the most volatile ones in recent times. Most candidates held the front runner position at some point during the campaign. This prevents the analysis of power dynamics in these debates from being biased on the personal characteristics of a single candidate or a small set of can-

¹There were no Democratic presidential primary debates in 2012, since the incumbent President Barack Obama was the de-facto nominee.

didates. Figure 2 shows the trend of how power indices of candidates (to be defined formally in Section 3.1) varied across debates.

3.1 Modeling Power in Debates

We use the term *Power Index* to denote the power or confidence with which a candidate comes into the debate. The Power Index of a candidate can be influenced by various factors. For example, during the presidential primary election campaigns, candidates get endorsed by various political personalities, newspapers and businesses. We think that such endorsements as well as the funds raised through campaigns positively affect the Power Index of the candidate. However, a more important source of a candidate’s power is their relative standing in recent poll scores. It gives the candidate a sense of how successful he/she is in convincing the electorate of his/her candidature. In this study, we model the Power Index of each candidate based solely on their recent state or national poll standings because we think that this is the most dominant factor. Other components such as the funds raised can be included in a similar fashion in the calculation of Power Index. We leave this to future work.

For each debate D , we denote the set of candidates participating in that debate by C_D . Let $date(D)$ denote the date on which debate D was held and $state(D)$ denote the state in which it was held. Debates from December 2011 onwards were held in states where the primaries were to be held in the near future. In these debates, we assume that their standings in the respective state polls, rather than national polls, would be the dominating factor affecting the power or confidence of candidates. Hence, for those debates, we chose the respective state’s poll scores as the reference. For others, we chose the national polls as the reference. Let $refType$ denote the type of the reference poll we consider for debate D .

$$refType = \begin{cases} state(D), & \text{if } date(D) > 12/01/11 \\ NAT, & \text{otherwise} \end{cases}$$

We show the $refType$ for each debate in Figure 1. For each debate, we find the poll results (national or state) released most recently and use the percentage of electorate supporting each candidate as the power index. If there are multiple polls released on the day the most recent poll was released, then we take the mean of poll scores

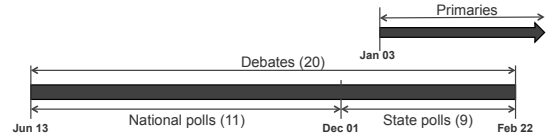


Figure 1: Timeline of Debates and Primaries

Number of debates	20
Interaction time	30-40 hrs
Average number of Candidates per debate	6.6
Average number of Turns per debate	245.2
Average number of Words per debate	20466.6

Table 1: Debate statistics

from all those polls to find the power index. Let $RefPolls(D)$ be the set of polls of type $refType$ released on the most recent date on which one or more such polls were released before $Date(D)$. We define the *Power Index*, $\mathcal{P}(X)$, of candidate $X \in C_D$ as below

$$\mathcal{P}(X) = \frac{1}{|RefPolls(D)|} \sum_{i=1}^{|RefPolls(D)|} p_i$$

where p_i denote the poll percentage X got in the i^{th} poll in $RefPolls(D)$.

4 Data

We obtained the manual transcripts of presidential debates from The American Presidency Project.² The transcripts of all debates follow similar formats, except for a few exceptions. Each debate’s transcript lists the presidential candidates who participated and the moderator(s) of the debate. Transcripts demarcate speaker turns and also contain markups to denote applause, laughter, booing and crosstalk during the debates. Table 1 shows various statistics on the debates. We obtained the state and national poll results from the corresponding Wikipedia pages which kept track of polls from various sources including Gallup, various national and regional news agencies etc.^{3,4} Figure 2 shows the trend of how the power indices of candidates varied across debates. Of the ten candidates, seven of them (everyone except Johnson, Huntsman and Pawlenty) were among the top 3 candidates for at least three debates.

²<http://www.presidency.ucsb.edu/debates.php>

³http://en.wikipedia.org/wiki/Statewide_opinion_polling_for_the_Republican_Party_presidential_primaries,_2012

⁴http://en.wikipedia.org/wiki/Nationwide_opinion_polling_for_the_Republican_Party_2012_presidential_primaries

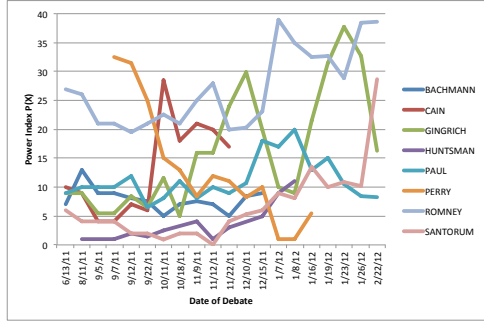


Figure 2: Power Index $P(X)$ variations across debates
 Note: Plots for Pawlenty and Johnson are not shown since they participated only in one or two debates.

5 Automatic Power Ranker

In this section, we present a supervised learning system to rank the participants of the debates based on their power indices. Formally, given a debate D with a set of participants $C_D = \{X_1, X_2, \dots, X_n\}$ and corresponding power indices denoted by $P(X_i)$ for $1 < i < n$, we want to find a ranking function $r : C_D \rightarrow \{1 \dots n\}$ such that for all $1 < i, j < n$,

$$r(X_i) > r(X_j) \iff P(X_i) > P(X_j)$$

We use an SVM based supervised learning system to estimate the ranking function r' that gives an ordering of participants $\{X'_1, X'_2, \dots, X'_n\}$, optimizing on the number of inversions between the orderings produced by r' and r .

5.1 Features

One of the primary ways power is manifested in an interaction is the manner in which people participate. By this, we are referring to the conscious and subconscious choices a participant makes while engaging in interactions. These include the lexical choices of each participant as well as other choices that affect the structure of the interaction - such as how much a participant speaks and on what topics. We used features to capture the language used in the debates as well as the structure of debates. Specifically, we analyze each debate participant in 4 dimensions — what they said (lexical features), how much they spoke (verbosity features), how they argued (argument features), and how they were talked about (mention features). Some structural features such as turns information are readily available from the transcripts, while for some others like arguments and candidate mentions, we use simple heuristics or perform deeper NLP analysis. The features we used are described in detail below

and are summarized in Table 2.

Code	Feature Description
Lexical: What they spoke	
WN	WordNgrams: word sequence of length 1 to 5
PN	PosNgrams: POS sequence of length 1 to 5
Verbosity: How much they spoke	
WD	WordDev: % of words spoken by $X - 1/ C_D $
TD	TurnDev: % of turns by $X - 1/ C_D $
QD	QuestionDev: % of questions to $X - 1/ C_D $
LT	LongestTurn: # of words in the longest turn
WT	WordsPerTurn: average # of words per turn
WS	WordsPerSent: average # of words per sentence
Argument: How they argued	
IOT	InterruptOthersPerTurn: % of candidate X 's turns that were interrupting others
OIT	OthersInterruptPerTurn: % of candidate X 's turns that others interrupted
Mentions: How they were talked about	
MP	MentionPercent: % of candidate X mentions
FN	FirstNamePercent: % of candidate X mentions that were first name mentions
LN	LastNamePercent: % of candidate X mentions that were last name mentions
FLN	FirstAndLastNamePercent: % of candidate X mentions that were first and last name mentions
TN	TitleAndNamePercent: % of candidate X mentions that were mentions using titles

Table 2: Features with respect to candidate X

Lexical - What they said: Ngram based features have been used in previous studies to analyze power in written interactions (Bramsen et al., 2011; Gilbert, 2012). It is expected to capture lexical patterns that denote power relations. We aggregated all turns of a participant and extracted counts for word lemma ngrams (WN) and POS tag ngrams (PN).

Verbosity - How much they spoke: We used features to capture each candidate's proportion of turns, time duration they talked, and number of questions posed to them. We approximated the time duration each speaker spoke by the total number of words spoken by him/her in the entire debate. To find the number of questions asked, we used the heuristic — instances where the candidate spoke right after the moderator are questions the moderator posed to the candidate. The percentage values of these features are dependent on the number of participants in each debate, which varied from 9 to 4. To handle this, for each feature, we measured the deviation of each candidate's percentage for that feature from its expected fair share percentage in the debate. We define the fair share percentage of a feature in a given debate to be $1/|C_D|$ — the percentage each candidate would

```

...
SANTORUM: ... I would ask Governor Romney, do
you believe people who have -- who were felons,
who served their time, who have extended --
exhausted their parole and probation, should
they be given the right to vote?

WILLIAMS: Governor Romney?

ROMNEY: First of all, as you know, the PACs
that run ads on various candidates, as we
unfortunately know in this --

SANTORUM: I'm looking for a question -- an
answer to the question first. [applause]

ROMNEY: We have plenty of time. I'll get there.
I'll do it in the order I want to do. [...]
the super PACs run ads. [...] they said that
you voted to make felons vote? Is that it?

SANTORUM: That's correct. That's what the ad
says.

ROMNEY: And you're saying that you didn't?

SANTORUM: Well, first, I'm asking you to answer
the question, because that's how you got the
time. It's actually my time. [...] should
they be given the right to have a vote?

```

Figure 3: Excerpt from the debate held at Myrtle Beach, SC on January 16 2012

have gotten for that feature if it was equally distributed. We calculate the deviation of each feature — TurnDev (TD), WordDev (WD) and QuestionDev (QD) — as the difference between observed percentage for that feature and $1/|C_D|$. We also investigated three additional structural features - longest turn length (LT), words per turn (WT) — whether they had longer turns on average, and words per sentence (WS) — whether they used shorter sentences.

Argument - How they argued: Modeling arguments and interruptions in interactions is not a straight-forward task. There has been work in the NLP community to detect arguments and interruptions in spoken as well as written interactions (Somasundaran et al., 2007). However, the well-structured nature of interactions that is expected in the debates allows us to use some simple heuristics to detect arguments and interruptions for the purposes of this study. We leave deeper NLP processing of candidate turns to detect interruptions and arguments for future work.

Debates follow a pattern where the candidate is expected to speak only after a moderator prompts him or her to either answer a question or to respond to another candidate. Hence, if a candidate talks immediately after another candidate, he is disrupting the expected pattern of the debate. This holds true even if such an out-of-turn talk may

not have interrupted the previous speaker mid-sentence. We considered such instances where the candidate spoke out-of-turn after another candidate as interruptions to the previous candidate. In most cases, such interruptions lead to back-and-forth exchanges between the candidates until a moderator steps in. We define such exchanges between candidates where they talk with one another without the moderator intervening as an argument. Arguments can extend to many number of turns. In counting interruptions, we counted only the first interruption by each candidate in the series of turns that constitute an argument. An example argument is given in Figure 3 where we counted only one instance of interruption for both Santorum and Romney. We used features to capture interruptions by candidate X as well as interruptions by others while candidate X was speaking. Since the raw counts of these measures are dependent on the number of turns, we used the normalized counts to find the *per-turn* value of these measures as features — InterruptOthersPerTurn (IOT) and OthersInterruptPerTurn (OIT).

Mentions - How they were talked about: Intuitively, how often a candidate was mentioned or referred to by others in the debate is a good indicator of his or her power. The more a candidate is mentioned, the more central he or she is in the the context of that debate. We use the mention count normalized across the total number of mentions of all candidates in a given debate (MP) as a feature.

In addition, we look at the form of addressing used while referring to each candidate. Previous studies in social sciences and linguistics have looked at the form of addressing in relation to the social relations (Brown and Ford, 1961; Dickey, 1997). Building on insights from these studies, we investigated if the modes of addressing candidates change with respect to their power. Specifically, we looked at four modes of addressing — FN (First Name), LN (Last Name) FLN (First and Last Name) and TN (Title followed by Name, first, last or full). As titles, we included common titles such as Mr., Ms. etc. as well as a set of domain-specific titles: Governor, Speaker, Senator, Congresswoman and Congressman. About 68.6% of total candidate mentions across debates were TN mentions, while the other types of mentions accounted for close to 10% each. FN, LN, TN and FLN capture the distribution of each candidate's mentions across these four types of mentions as

percentage of their total mentions.

5.2 Correlation Analysis and Significance

Figure 4 shows the Pearson’s product correlation between each structural feature and candidate’s power index $P(X)$. The darker bars denote statistically significant ($p < 0.05$) correlations. Applying Bonferroni correction for multiple tests, the threshold for p-value for significance would be reduced to 0.0025. Even then, the statistically significant features would retain their significance. We consider three correlation windows — weak (0.2 - 0.39), moderate (0.4 - 0.69) and high (0.7 and above).

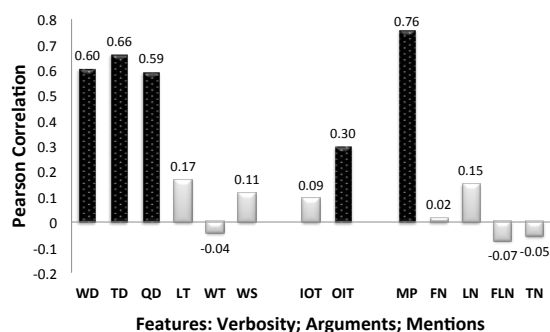


Figure 4: Pearson Correlations for Structural Features
Correlation windows: Weak (0.2 - 0.39); Moderate (0.4 - 0.69); High (≥ 0.7)

We obtained statistically significant moderate positive correlation between the word and turn features and candidates’ power indices. Candidates with higher power indices spoke for significantly more time than others (WD) and they also got significantly more number of turns (TD). This finding is in line with the empirical findings in sociology literature (Ng et al., 1993; Reid and Ng, 2000). We also obtained moderate positive correlation between questions posed to the candidate and his or her power index, which suggests that the candidates with higher power indices were asked significantly more questions by the moderators.

Another interesting observation was on the interruption patterns. We obtained no significant correlation between how powerful a candidate was and how often he/she interrupted others (IOT). Instead, we found statistically significant positive correlation (although weak) for OIT, which means that the candidates with more power were interrupted significantly more by others. This is counter-intuitive and in contrast with previous findings by (Ng et al., 1995) that those who interrupt are more influential or powerful. We believe

that this is a manifestation of the participants pursuing power over each other rather than operating within a static power structure.

We found statistically significant high positive correlation between the power indices of candidates and how often they were referenced/mentioned by others (MP). In other words, as candidates gain more power, they are referenced significantly more by others. However, the distribution of mentions of a candidate across different forms of addressing (FN, LN, TN, FLN) did not have any correlation with the power indices of the candidate. This suggests that while forms of addressing is found to be correlated with power relations by previous studies (Brown and Ford, 1961; Dickey, 1997), they are not affected by the short term variations of power as in our domain.

5.3 Implementation

To build the ranker, we used the ClearTk’s *SVM^{rank}* (Joachims, 2006) wrapper package. We also used the ClearTk wrapper for the Stanford CoreNLP package to perform basic NLP analysis on the speaker turn texts. The basic steps we performed include - tokenization, sentence segmentation, parts-of-speech tagging, lemmatization and named entity tagging.

5.4 Evaluation

We report results on 5-fold cross validation. We report three commonly used evaluation metrics for ranking tasks — Kendall’s Tau, $nDCG$ and $nDCG_3$. Kendall’s Tau measures the similarity between two rankings based on the number of rank inversions (discordant pairings) between original and predicted ranking. $nDCG$ employs a normalized discounted cumulative gain method which penalizes the inversions happening in the top of the ranked list more than those happening in the bottom. $nDCG_3$ focuses only on the top 3 candidates from each debate. $nDCG$ based metrics are more suitable for our purposes since it provides a way to factor in the magnitude of ranking metric (in our case, power index) in the performance assessment. E.g., under $nDCG$, the penalty for swapping a pair of candidates with $P(X)$ values 35.0 and 5.0 will be higher than that for a pair with $P(X)$ values 12.0 and 15.0. Tau treats these mistakes equally if the swaps generate the same number of inversions.

5.5 Results and Discussion

We first find the best performing set of lexical features (Word and POS ngrams) by varying the ngram length from 1 to 5. We then find the best performing feature subset of structural features among all subsets. The small cardinality of the set of structural features makes this feasible. We then use the combination of the best feature subsets from both settings. The results obtained are presented in Table 3. We present a baseline system using word unigrams as features.

	Tau	nDCG	nDCG-3
<i>Baseline (Unigrams)</i>	0.25	0.860	0.733
<i>WN+PN</i>	0.36	0.880	0.779
<i>WD+QD+MP</i>	0.47	0.961	0.921
<i>WD+QD+OIT</i>	0.45	0.960	0.921
<i>WN+PN+WD+QD+MP</i>	0.37	0.902	0.818
<i>WN+PN+WD+QD+OIT</i>	0.37	0.902	0.826

Table 3: Ranker results

We obtain the best configuration of lexical features to be WN+PN, with values of n as 1 and 2 respectively. The PN features improve the performance of the baseline system (unigrams) from 0.25 to 0.36 Tau . Similar improvements are observed in $nDCG$ and $nDCG_3$ as well. The structural features outperform the lexical features and obtain the best overall result of 0.961 for $nDCG$ and 0.921 for $nDCG_3$ for a combination of WordDeviation, QuestionDeviation and MentionPercent. Another feature subset — WordDeviation, QuestionDeviation and OthersInterruptPerTurn — obtained the same performance in $nDCG_3$, but slightly lower numbers for Tau and $nDCG$. The overall best performing features were WD, QD, MP and OIT, which is in line with the findings in the correlation study in Section 5.2. WD suggests that people with more power tend to and/or are allowed to talk more. QD, MP and OIT are reflections of how others’ perception of a candidates power affected the way they interacted with him/her. Surprisingly, combining lexical and structural features did not yield good results. We suspect that this might be due to the high dimensional ngram feature space.

We analyzed the correlation of each structural features with $P(X)$ in Section 5.2. However, it is not feasible to perform such significance studies on ngram features because of the huge feature space. In order to find the ngram features that are most representative for this task, we inspected

the feature weights of the linear kernel model created for the best performing ngram feature set (WN+PN). Table 4 lists few of the interesting features that came in the top 25 positive and negative weighted features, along with corresponding weights. POS tags are capitalized and `_BOS_` stands for `beginning_of_sentence`. It is hard to infer strong conclusions based purely on the SVM feature weights. However, SVM does pick up some interesting signals. E.g., those with power used `you` more, while those with less power used `we` more. Also, those with power used `agree` more, suggesting that they might be less contentious than others. `UH_`, which captures interjections/pauses was assigned a positive weight, which aligns with the finding that those with power get interrupted more. `_BOS__JJ` (-0.11) suggests that the participant with lower power tend to start sentences using an adjective.

Positive weighted	Negative weighted
<i>VBN_NN (0.30)</i>	<i>tell (-0.24)</i>
<i>agree (0.27)</i>	<i>do (-0.23)</i>
<i>UH_ (.18)</i>	<i>WDT (-0.15)</i>
<i>you (0.09)</i>	<i>we (-0.09)</i>
<i>VBP_TO (0.18)</i>	<i>_BOS__JJ (-0.11)</i>

Table 4: Top weighted features from the ngram based model created for WN + PN

6 Conclusion and Future Work

We presented a system to automatically rank participants of an interaction in terms of their relative power. We identified several linguistic and structural features that were effective in predicting these rankings. We conducted this study in the domain of political debates, specifically the 2012 Republican presidential primary debates. We find that candidates’ power indices affected the way they interacted with others in the debates — how much they spoke and how they spoke. We also found that power affected the way others interacted with them — the number of questions directed at them, how often they were interrupted, and how often they were mentioned. Our experiments in this domain yield very encouraging results and we plan to investigate if these findings carry across to other genres of multi-party conversations as a part of our future work. We also plan to perform deeper analysis on the interactions such as looking for dialog patterns which may signal topic control in relation to power.

References

- E. Bakshy, J.M. Hofman, W.A. Mason, and D.J. Watts. 2011. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 65–74. ACM.
- R.F. Bales and P.E. Slater. 1955. Role differentiation in small decision-making groups. *Family, socialization and interaction process*, pages 259–306.
- R. F. Bales. 1970. *Personality and interpersonal behaviour*. New York: Holt, Reinhart, and Winston.
- 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.
- R. Brown and M. Ford. 1961. Address in american english. *The Journal of Abnormal and Social Psychology*, 62(2):375.
- 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.
- E. Dickey. 1997. Forms of address and terms of reference. *Journal of Linguistics*, pages 255–274.
- J. Diesner and K. 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.
- 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.
- T. Joachims. 2006. Training Linear SVMs in Linear Time. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 217–226. ACM.
- M. A. Locher. 2004. *Power and politeness in action: disagreements in oral communication*. Language, power, and social process. M. de Gruyter.
- S. H. Ng and J. J. Bradac. 1993. *Power in language: Verbal communication and social influence*. Sage Publications, Inc.
- 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.
- V. Nguyen, J. Boyd-Graber, and P. Resnik. 2012. SITS: A hierarchical nonparametric model using speaker identity for topic segmentation in multi-party. In *Association for Computational Linguistics*.
- V. Prabhakaran and O. Rambow. 2013. Written Dialog and Social Power: Manifestations of Different Types of Power in Dialog Behavior. In *Proceedings of the IJCNLP*, Nagoya, Japan, October. ACL.
- V. Prabhakaran, O. Rambow, and M. Diab. 2012a. 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. 2012b. 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.
- A. Rosenberg and J. Hirschberg. 2009. Charisma perception from text and speech. *Speech Communication*, 51(7):640 – 655. Research Challenges in Speech Technology: A Special Issue in Honour of Rolf Carlson and Bjrn Granstrm.
- 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.
- J. B. Sexton and R. L. Helmreich. 1999. Analyzing cockpit communication: The links between language, performance, error, and workload. In *Proceedings of the Tenth International Symposium on Aviation Psychology*, pages 689–695.
- S. Somasundaran, J. Ruppenhofer, and J. Wiebe. 2007. Detecting arguing and sentiment in meetings. In *Proceedings of the SIGdial Workshop on Discourse and Dialogue*.