

Investigating the Relationship between Multi-Party Linguistic Entrainment, Team Characteristics, and the Perception of Team Social Outcomes

Mingzhi Yu,¹ Diane Litman,¹ Susannah Paletz²

¹University of Pittsburgh, ²University of Maryland
miy39@pitt.edu, dlitman@pitt.edu, paletz@umd.edu

Abstract

Multi-party linguistic entrainment refers to the phenomenon that speakers tend to speak more similarly during conversation. We first developed new measures of multi-party entrainment on features describing linguistic style, and then examined the relationship between entrainment and team characteristics in terms of gender composition, team size, and diversity. Next, we predicted the perception of team social outcomes using multi-party linguistic entrainment and team characteristics with a hierarchical regression model. We found that teams with greater gender diversity had higher minimum convergence than teams with less gender diversity. Entrainment contributed significantly to predicting perceived team social outcomes both alone and controlling for team characteristics.

Introduction

Linguistic entrainment is the phenomenon that speakers tend to speak similarly in conversations (Brennan and Clark 1996). It has attracted attention for its potential to improve performance when being incorporated into spoken dialogue systems. One successful example is that Lopes, Eskenazi, and Trancoso (2015) utilized entrainment to automatically choose better primes in the prompt of a dialogue system. In dyadic conversations, the existence of entrainment has been widely investigated for various linguistic features. Brennan and Clark (1996) have found that speakers tend to refer to the same object using identical lexical terms in their conversation. Levitan and Hirschberg (2011) found evidence of prosodic entrainment in intensity, pitch and voice quality. Niederhoffer and Pennebaker (2002) found that speaker pairs gradually matched their linguistic style in conversation. Besides linguistic features, researchers have also linked dyadic entrainment to other aspects of conversations, such as task success (Reitter and Moore 2007), gender factors (Namy, Nygaard, and Sauerteig 2002) and social behaviors (Levitan et al. 2012). In recent years, researchers have started studying multi-party entrainment in conversations with more than two speakers. As with dyadic entrainment, the existence of multi-party entrainment has also been found for different linguistic features. For example, Rahimi et al. (2017) found speakers in the same group entrained on high

frequency words and topic words. Litman et al. (2016) found significant group-level differences in pitch, jitter and shimmer between first and second halves of conversation. Multi-party entrainment has also been associated with other aspects of conversations such as task performance and group cohesiveness. Friedberg, Litman, and Paletz (2012) found that higher scoring teams are more likely to entrain in the use of task-related terms. Gonzales, Hancock, and Pennebaker (2010) suggested that group linguistic style matching is a significant indicator of team cohesiveness. Compared to studies demonstrating the existence of multi-party entrainment, however, fewer studies have investigated the links between multi-party entrainment and other aspects of conversations. We present two investigations of such relationships.

First, we investigate whether group or team characteristics relate to multi-party entrainment, since multiple individuals simultaneously engage in the same conversation. While dyad research has analyzed entrainment and gender composition (Levitan et al. 2012; Namy, Nygaard, and Sauerteig 2002), relationships between team characteristics and multi-party entrainment could be more complex, given the increasing number of person-to-person and person-to-team communications. We examine such relationships using cooperative game conversations, in which multiple speakers are brought together as a team. Three types of team characteristics are investigated: gender composition as in prior work, as well as team size and diversity.

Second, since relationships between multi-party entrainment and other conversational aspects are not well established, it is plausible that correlations found in prior studies (Friedberg, Litman, and Paletz 2012; Gonzales, Hancock, and Pennebaker 2010) are confounded with team characteristics such as team size and gender composition. Team characteristics have already been found to have complex impacts on team processes (O'Reilly, Caldwell, and Barnett 1989; Smith et al. 1994; Fisher et al. 2012). For instance, team size is negatively correlated with team conflict (Amason and Sapienza 1997). We hypothesize that both entrainment and team characteristics, specifically team size, gender composition and the diversity of a team, are associated with the perception of team social outcomes. We use hierarchical regression models to examine the unique contribution of multi-party entrainment in explaining perceived team social outcomes above and beyond team characteristics.

Finally, to support our studies, we have developed an innovative representation of multi-party entrainment by extending the measurement from Litman et al. (2016) and adapting it to study the feature of linguistic style from Pennebaker and King (1999). We used this measure to statistically examine relationships between multi-party entrainment and team characteristics in a set of dialogues from a freely available cooperative game corpus. We also demonstrated that multi-party entrainment is associated with team outcomes, even after controlling for team characteristics.

The Teams Corpus

The freely available Teams Corpus (Litman et al. 2016) consists of 47 hours of audio and transcriptions from 62 teams (35 three-person, 27 four-person). The audio files are manually segmented and transcribed at the level of inter-pausal units (IPUs), based on a pause length of 200 milliseconds. Each team consists of American native speakers from 18 to 67 years old who played two rounds of the cooperative board game Forbidden Island. 213 individuals (79 males, 134 females) were assigned to the teams and given one of four game roles: Engineer, Messenger, Pilot, Explorer.

The corpus also includes survey data. A pre-game survey collected personal information such as age, gender, and eight options for ethnicity. While each participant could choose multiple options, in this paper we categorize each speaker into nine exclusive categories: Caucasian (150), East Asian (12), South Asian (11), Pacific Islander (0), Black (15), Native American (0), Hispanic (3), Middle Eastern (2), and Multiple Ethnicity (20) for participants who chose more than one of the other categories. The gender data yields seven types of team gender composition: 0% female (2), 25% female (4), 33% female (7), 50% female (9), 66% female (18), 75% female (10), 100% female (12). Participants also took post-game surveys to evaluate team processes. These surveys contained a series of self-report questions on team cohesion, satisfaction, and other team social outcome constructs.

The studies presented in this paper are based only on the data related to the first of the two games, as only these transcriptions were available. Before computing entrainment, we further processed these transcripts by removing punctuation, converting all words to lower case, and removing a list of interjections, e.g., ‘hmm’, that are not discussed in linguistic style (Pennebaker and King 1999). We then concatenated all the processed IPU transcriptions for each speaker.

Features and Measures

Features of Speaker Linguistic Style

Before computing entrainment, we first extracted linguistic style features for each speaker in each transcript using LIWC2007 (Pennebaker, Booth, and Francis 2007), a computational application for text analysis that includes a dictionary mapping a list of words to 64 psychological and linguistic categories. We used this dictionary to label each word in each speaker’s concatenated IPU transcripts with potentially multiple LIWC categories. The final number of occurrences of each category was then converted into a percent.

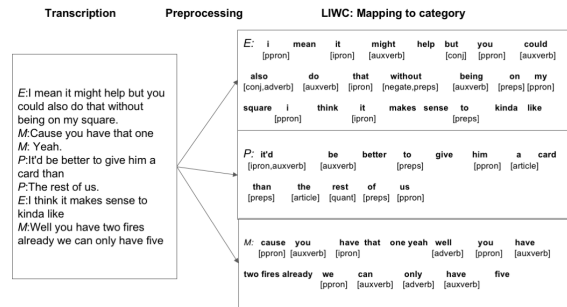


Figure 1: Each tag corresponds to a LIWC function word category. negate: negation, conj: conjunctions, preps: prepositions, ppron: personal pronouns, ipron: impersonal pronoun, article: article, adverb: adverbs, quant: quantifiers, auxverb: auxiliary verbs.

In our study we only focused on a limited subset of LIWC categories, namely function words. The first reason is that function words reflect the speaker’s psychological state and convey information about the interactive process (Chung and Pennebaker 2007). Function words represent a high-level linguistic difference in style. Second, in contrast to content words, function words do not rely on any specific task domain (Gonzales, Hancock, and Pennebaker 2010) and have a very high frequency in daily speech (Rochon et al. 2000). Using function words as features can alleviate feature sparsity. Since a considerable number of studies about linguistic style have used function words (Gonzales, Hancock, and Pennebaker 2010; Danescu-Niculescu-Mizil, Gamon, and Dumais 2011; Mukherjee and Liu 2012), we directly adopted Gonzales, Hancock, and Pennebaker’s (2010) selection of 9 LIWC categories as function words.

Figure 1 shows how we used LIWC to create function word features from a transcript excerpt. After the transcript preprocessing and speaker IPU concatenation discussed above, LIWC scored each speaker’s input text and generated the category percentages for each of the 64 categories. For instance, the Engineer uttered 24 words in this excerpt but only one word belongs to the category negate. Thus, the category percentage for negate is $1/24 = 4.20\%$. Since one word may belong to multiple categories, the sum of category percentages for the 64 categories may exceed 100.

Measures of Team Linguistic Style Entrainment

There are various methods to directly calculate multi-party entrainment using linguistic style. Some text-based studies have proposed probabilistic frameworks in linguistic style matching based on pairwise comparisons between speakers (Mukherjee and Liu 2012; Danescu-Niculescu-Mizil, Gamon, and Dumais 2011). However, compared to their data, our data has a lower density of reciprocated interactions per pair. The number of conversations between speakers is insufficient for constructing such probability models. Addressee identification to create appropriate pairs is also not straightforward. Gonzales, Hancock, and Pennebaker (2010) devel-

oped a method to perform linguistic style matching based on multi-party speech, but they only focused on a global measure rather than on the degree of change. Recently, Litman et al. (2016) proposed a method to compute multi-party entrainment on acoustic-prosodic features based on the same Teams Corpus as used here. Their method highlighted feature change over time, which is more relevant to linguistic style entrainment.

For each feature, they calculated the difference between a pair of speakers as the absolute difference of feature values, and the team difference as the average difference over all pairs. In our study, linguistic style is a single feature with multiple categories, so we converted their calculation of pair differences by summing up all the category differences. Moreover, we weighted category differences by the frequency of categories. More specifically, $TDiff_{unw}$ (unweighted team difference) converts the team difference of Litman et al. (2016) to deal with multiple feature categories. $TDiff_w$ (weighted team difference) extends $TDiff_{unw}$ by weighting the category differences similarly to Gonzales, Hancock, and Pennebaker (2010). We calculated both $TDiff_{unw}$ and $TDiff_w$ for each pair of speakers and then averaged over all pairs. The formulas are shown in Equations 1, 2, and 3, where F , K , and $|team\ size|$ respectively refer to the function word category set, an arbitrary function word category, and the team size. $KDiff_{ij}$ refers to the weighted category difference of category K between speakers i and j .

$$TDiff_{unw} = \frac{\sum_{\forall i \neq j \in team} (\sum_{K \in F} (|K_i - K_j|))}{|team\ size| * (|team\ size| - 1)} \quad (1)$$

$$TDiff_w = \frac{\sum_{\forall i \neq j \in team} (\sum_{K \in F} (|KDiff_{ij}|))}{|team\ size| * (|team\ size| - 1)} \quad (2)$$

$$KDiff_{ij} = \frac{|K_i - K_j|}{K_i + K_j}, KDiff_{ij} = 0 \text{ if } K_i, K_j = 0 \quad (3)$$

Litman et al. (2016) then define convergence, a type of entrainment measuring increase in feature similarity, by comparing the $TDiff$ of two non-overlapping temporal intervals of a game as in Equation 4. C_{ij} and $TDiff$ refer to the team's convergence and the weighted (or unweighted) team differences, respectively. Assuming the game is divided into n disjoint temporal intervals, i and j refer to two predetermined temporal intervals in chronological order.

$$C_{ij} = TDiff_i - TDiff_j, i < j \in n \quad (4)$$

However, this definition leaves two unanswered questions. First, the measure of convergence allows negative values that represent divergence, which is the tendency that team members speak differently. Second, it requires the researcher to hand pick temporal intervals that are not guaranteed to result in an optimal measurement of entrainment. Hence, we derived four new variables of convergence (see Equations 5 and 6): Max and Min calculating the maximum and minimum positive C_{ij} , and absMax and absMin calculating the absolute maximum and minimum $|C_{ij}|$.

$$Max \text{ or } Min = Max\{C_{ij} > 0\} \text{ or } Min\{C_{ij} > 0\} \quad (5)$$

$$absMax \text{ or } absMin = Max\{|C_{ij}|\} \text{ or } Min\{|C_{ij}|\} \quad (6)$$

Rather than two fixed intervals, we iterated over all two arbitrary temporal intervals in chronological order and conducted the comparison. Consequently, the Max and Min only measure maximum and minimum convergence so that they directly reflect the decrement of $TDiff$ between two intervals. The absMax and absMin measure the maximum and minimum magnitude of the change of $TDiff$ in the entire conversation. Unlike the Min and Max, the absMax and absMin are determined by the values of convergence or divergence. We added the absMax and absMin beyond Min and Max so that they reflect the overall fluctuation ranges of $TDiff$, which might also be an important aspect of entrainment. Therefore in total, we defined eight measures of team entrainment: **unweighted** and **weighted Max, Min, absMin, and absMax convergence**.

The parameter n in Equation 4 determines the length of temporal intervals being compared. Many studies defined n as two so that the conversation is evenly divided into two halves (Leviton and Hirschberg 2011; Rahimi et al. 2017). Since Litman et al. (2016) previously found that in the Teams corpus the highest acoustic-prosodic convergence occurred within the first and last three minutes, we used this finding to define our n . We evenly divided each game, which was limited to 30 minutes, into ten intervals, so each interval is less than three minutes. Since our focus is on measure development in this paper, methods for optimally tuning this temporal parameter are left for future work.

We will use Figure 1's excerpt to illustrate our calculations. Assuming n is set to two, we first divide the excerpt into two time intervals. Assuming that the temporal midpoint of the excerpt occurs after the fourth IPU, the first interval includes the first through fourth IPUs. The second interval includes the fifth through seventh IPUs. For each speaker, all IPUs in each interval are concatenated and input to LIWC. The interval division and LIWC category percentage output are shown in Figure 2. Based on Equation 1, the unweighted pair difference between the Engineer and Pilot in the first interval is calculated as the sum of the absolute differences of all categories, which is equivalent to $|0 - 11.11| + |6.25 - 0| + |12.5 - 0| + |12.5 - 22.22| + |18.75 - 11.11| + |6.25 - 0| + |12.5 - 11.1| + |0 - 0| + |25 - 22.22| = 57.64$. Similarly, the pair differences between the other two pairs (Engineer and Messenger, Pilot and Messenger) are 52.08 and 50. The unweighted team difference is the average of these pair differences, which is 53.24. The weighted team difference is calculated using Equation 2, with the pair difference now being normalized by the frequency of each category. For instance, the absolute difference between Engineer and Pilot of negate is $|6.25 - 0| = 6.25$. This number is less than the absolute difference of $|18.75 - 11.11| = 7.64$ for the category ppron. However, the occurrence of negate is less common than ppron in the speech of Engineer and Pilot. The weighted difference of negate is $|6.25 - 0| / (6.25 + 0) = 1$,

1 st interval	Engineer	i mean it might help but you could also do that without being on my square (word count = 16)								
	Pilot	it'd be better to give him a card than (word count = 9)								
	Messenger	cause you have that one yeah (word count = 6)								
2 nd interval	Engineer	i think it makes sense to kinda like (word count = 8)								
	Pilot	the rest of us (word count = 4)								
	Messenger	well you have two fires already we can only have five (word count = 11)								
Time Interval	Speaker	article	negate	conj	preps	ppron	adverb	ipron	quant	auxverb
1 st interval	Engineer	0	6.25	12.5	12.5	18.75	6.25	12.5	0	25
	Pilot	11.11	0	0	22.22	11.11	0	11.11	0	22.22
	Messenger	0	0	0	0	16.67	0	16.67	0	16.67
2 nd interval	Engineer	0	0	0	12.5	12.5	0	12.5	0	0
	Pilot	25	0	0	25	25	0	0	25	0
	Messenger	0	0	0	0	18.18	18.18	0	0	27.27

Figure 2: Top: The input text of each speaker per interval. Bottom: The corresponding LIWC category percentage.

which is now greater than the weighted difference of ppron which is $|18.75 - 11.11| / (18.75 + 11.11) = 0.26$.

Measures of Team Characteristics

This paper focuses on the following team characteristics: **team size**, **gender diversity** (Blau's index and female percentage), **ethnic diversity**, and **age diversity**. Note that the female percentage measures the numerical female dominance in a team, while gender diversity indicates the variability of gender composition. Diversity of age, which has continuous values, is measured by the population standard deviation. Diversity of ethnicity and gender with categorical values is measured by Blau's index of heterogeneity (Blau 1977) as in Equation 7, where P_k is the proportion of a specific category k .

$$\text{Ethnic/Gender Diversity}_{\text{Blau's}} = 1 - \sum P_k^2 \quad (7)$$

Measures of Perceived Team Social Outcomes

We assessed the perception of team social outcomes using the existing self-reported post-game survey responses. The survey contains scales related to team processes and team conflict. Team processes consist of the perceptions of team cohesion, general team satisfaction, potency/efficacy, and perceptions of shared cognition (Wendt, Euwema, and van Emmerik 2009; Wageman, Hackman, and Lehman 2005; Guzzo et al. 1993; Gevers, Rutte, and Van Eerde 2006). These four measures were strongly correlated with each other. Thus, we aggregated them into a single scale by averaging their z-scored scale composites, Cronbach's $\alpha = 0.78$. Team conflict consists of task, process and relationship conflict. These three types of conflict reflect the topic of the conflict, be it about the task at hand, work processes, or interpersonal values and personal relationships. Process conflict is consistently negatively related to performance, but task conflict is positively related to performance under some conditions (De Wit, Greer, and Jehn 2012). Therefore, we kept these three types of conflict as individual variables. Overall, we thus have four measures of perceived team social outcomes: **team processes**, **task conflict**, **process conflict** and **relationship conflict**.

Results

Relating Team Characteristics and Entrainment

We first tested the relationship between linguistic style entrainment and team characteristics with continuous values (gender, ethnic and age diversity) using Spearman rho correlations. There was a significant positive correlation between unweighted convergence Min and gender diversity, ($r(62) = .22, p < .05$). This correlation indicated that teams with greater gender diversity had higher minimum convergence than teams with less gender diversity.

We then performed one-way ANOVA tests between linguistic style entrainment and the categorical team characteristics, i.e., percentage of females and team size. The unweighted absMax was found to significantly vary with female percentage for the 7 conditions (see corpus section), $F(6,55) = 2.79, p = .019$. Tukey HSD post hoc tests indicated that the 25% condition ($N = 4, M = 40.15, SD = 13.263$) was significantly different with the 50% condition ($N = 9, M = 19.56, SD = 9.435$), 66% condition ($N = 18, M = 19.39, SD = 9.407$) and 75% condition ($N = 10, M = 18.92, SD = 8.117$). The mean of the 25% condition was larger than all other three conditions. This finding suggests that the maximum magnitude of the change of unweighted team differences in the 4-person team with one female was greater than other mixed-gender teams with more than one female.

Predicting Perceived Team Social Outcomes

We predicted four measures of perceived team social outcomes: team processes (MIN = -2.57, MAX = 1.51, M = 0.00, SD = 0.80); task conflict (MIN = 1.00, MAX = 3.33, M = 1.75, SD = 0.46); process conflict (MIN = 1.00, MAX = 3.00, M = 1.58, SD = 0.41) and relationship conflict (MIN = 1.00, MAX = 1.75, M = 1.15, SD = 0.20). A hierarchical linear regression (HLR) model allows us to consider the impact of team characteristics on the perception of team social outcomes, and then examine the significance of multi-party entrainment as a predictor beyond or controlling for team characteristics. Two models, Model 1 (M1) and Model 2 (M2), were constructed. Team size and team diversity (gender, ethnic and age) were entered simultaneously as independent team characteristic variables (IVs). Multi-party entrainment was entered into M2 as an IV beyond the team characteristics. Only variables of multi-party entrainment that significantly contributed to the model were selected in M2. The dependent variable (DV) of each HLR was the variable describing the perceived team social outcomes.

Significant HLR models are shown in Table 1. The M2 predicting the task and process conflict were both significant, but no M1 was significant. In the HLR predicting task conflict, no team characteristics contributed significantly to M1. Introducing variables of multi-party entrainment to M2 explained an additional 9.2% variation in task conflict and the ΔR^2 was significant, $\Delta F(1, 55) = 6.46, p < 0.05$. Unweighted absMax and team size were both significant contributors to task conflict in M2. The negative association between unweighted absMax and task conflict suggested that the higher maximum magnitude of the change of team difference signaled less team conflict. Meanwhile, the positive

DV	IV	Model 1 (β)	Model 2 (β)
Task	Age	0.04	-0.03
	Ethnic	0.19	0.19
	Gender	-0.04	-0.11
	%Female	-0.21	-0.30
	Team size	0.24	0.25*
	unw absMax		-0.31*
	R^2	0.13	0.22
F	1.60	2.54*	
Process	Age	0.06	0.04
	Ethnic	0.25	0.29*
	Gender	-0.08	-0.13
	%Female	-0.08	-0.14
	Team size	0.34**	0.31*
	w absMax		-0.36**
	R^2	0.16	0.28
F	2.10	3.57**	

Table 1: Predicting team outcomes using hierarchical regression. Age: $AgeDiversity_{SD}$, Ethnic: $EthnicityDiversity_{Blau's}$, Gender: $GenderDiversity_{Blau's}$, % Female: Percentage of female, w absMax: weighted convergence absMax, unw absMax: unweighted convergence absMax. β : standardized Beta. * if $p < 0.05$, ** if $p < 0.01$.

association between team size and task conflict in M2 added evidence to previous findings that team size is positively associated with team conflict (Amazon and Sapienza 1997; Smith et al. 1994).

In the HLR predicting process conflict, team size contributed significantly to M1. Adding multi-party entrainment (the weighted absMax) to M2 explained an additional 12.2% of the variability in process conflict, and the ΔR^2 was significant, $\Delta F(1, 55) = 9.36, p < 0.01$. Multi-party entrainment along with team size and ethnic diversity were important predictors to M2. Team size and ethnic diversity were both positively associated with process conflict. We observed a negative association between the weighted absMax and process conflict. This finding implied that the higher maximum magnitude of the change of team difference signaled less process conflict.

Overall, we found a negative association between maximum magnitude of the change of team difference and team conflict, specifically process and task conflict. Team size and ethnic diversity both had effects on team conflict. Maximum magnitude of the change of team difference was a significant predictor in team conflict.

To determine whether the team characteristics had a significant impact on the conflict variables above and beyond the effect for entrainment, we switched the IVs in M1 and M2. Variables of entrainment were entered into M1 stepwise and then the team characteristics that had shown significance in the previous HLR were entered into M2 (see Table 2). We observed similar findings in that both M1 and M2 significantly predicted task and process conflict. The maximum magnitude of the change of the team difference was significantly negatively associated with task and process conflict. Team size and, for process conflict, ethnic diversity were significantly related to conflict above and beyond entrainment.

DV	IV	Model 1 (β)	Model 2 (β)
Task	unw absMax	-0.27*	-0.27*
	Team size		0.25*
	R^2	0.07	0.13
	F	4.77*	4.55*
Process	w absMax	-0.34**	-0.34**
	Team size		0.32**
	Ethnic		0.26*
	R^2	0.12	0.26
F	7.87**	6.69**	

Table 2: Flipped HLR : Entrainment was stepwise entered in M1. Team characteristics showing significance in prior HLRs were entered in M2.

Conclusions and Future Work

We first proposed a new method for measuring multi-party linguistic style entrainment by converting and extending methods developed in prior studies of both linguistic style matching and team acoustic-prosodic entrainment. We then examined the relationship between multi-party entrainment and team characteristics. Our analysis implies that teams with greater gender diversity had greater minimum convergence than teams with less gender diversity, similarly to the findings of Levitan et al. (2012) and Namy, Nygaard, and Sauerteig (2002) that mixed-gender pairs generally entrain more in dyadic conversations. Moreover, the 4-person teams with more than one female had a higher maximum magnitude of change in team difference. Perhaps the existence of a female subgroup reconciled the team difference in these teams. In conclusion, different gender compositions affect the entraining behaviors of the overall team. These findings show that gender plays an important role for linguistic entrainment in human interactions. They also reveal a need to study the underlying process of multi-party entrainment with different granularity levels. Next, we predicted the perception of team social outcomes by team characteristics and variables of entrainment with hierarchical regression models. The experimental results indicated that the maximum magnitude of the change of the team difference was negatively associated with team conflict. Adding this variable of entrainment beyond team characteristics resulted in statistically significant improvements in model prediction. Finally, by entering entrainment variables in the first rather than second model, we showed that entrainment was significantly negatively associated with task and process conflict, both when controlling for team characteristics and when not. Although the overall models did not account for a large amount of variance, the base model of only team characteristics was improved significantly by adding entrainment. In sum, we found that entrainment is a promising feature to predict team social outcomes. In terms of broader impact, we can now possibly evaluate the success of team conversations using linguistic style entrainment. Additional interdisciplinary research building on our findings could test whether entrainment mediates the effects of team characteristics on social and task outcomes in different settings.

In future work, we will investigate different feature combinations and prediction models to improve the performance

of our statistical models. To further improve the calculation of multi-party entrainment, we intend to search for an optimal temporal window. Additionally, we plan to review the validity and accuracy of the team social outcomes, which were measured with self-reported surveys. We also plan to investigate the relationship between multi-party entrainment and individual differences, such as personality and education background.

Acknowledgements

This work is supported by the National Science Foundation under Grant Nos. 1420784 and 1420377. We thank all the transcribers, the Pitt PETAL group and the anonymous reviewers for advice in improving this paper.

References

- Amason, A. C., and Sapienza, H. J. 1997. The effects of top management team size and interaction norms on cognitive and affective conflict. *Journal of management* 23(4):495–516.
- Blau, P. M. 1977. *Inequality and heterogeneity: A primitive theory of social structure*, volume 7. Free Press New York.
- Brennan, S. E., and Clark, H. H. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(6):1482.
- Chung, C., and Pennebaker, J. W. 2007. The psychological functions of function words. *Social communication* 1:343–359.
- Danescu-Niculescu-Mizil, C.; Gamon, M.; and Dumais, S. 2011. Mark my words!: linguistic style accommodation in social media. In *Proceedings of the 20th international conference on World wide web*, 745–754. ACM.
- De Wit, F. R.; Greer, L. L.; and Jehn, K. A. 2012. The paradox of intragroup conflict: a meta-analysis. *Journal of Applied Psychology* 97(2):360.
- Fisher, D. M.; Bell, S. T.; Dierdorff, E. C.; and Belohlav, J. A. 2012. Facet personality and surface-level diversity as team mental model antecedents: implications for implicit coordination. *Journal of Applied Psychology* 97(4):825.
- Friedberg, H.; Litman, D.; and Paletz, S. B. 2012. Lexical entrainment and success in student engineering groups. In *Spoken Language Technology Workshop (SLT), 2012 IEEE*, 404–409. IEEE.
- Gevers, J. M.; Rutte, C. G.; and Van Eerde, W. 2006. Meeting deadlines in work groups: Implicit and explicit mechanisms. *Applied psychology* 55(1):52–72.
- Gonzales, A. L.; Hancock, J. T.; and Pennebaker, J. W. 2010. Language style matching as a predictor of social dynamics in small groups. *Communication Research* 37(1):3–19.
- Guzzo, R. A.; Yost, P. R.; Campbell, R. J.; and Shea, G. P. 1993. Potency in groups: Articulating a construct. *British journal of social psychology* 32(1):87–106.
- Levitan, R., and Hirschberg, J. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Twelfth Annual Conference of the International Speech Communication Association*.
- Levitan, R.; Gravano, A.; Willson, L.; Benus, S.; Hirschberg, J.; and Nenkova, A. 2012. Acoustic-prosodic entrainment and social behavior. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies*, 11–19. Association for Computational Linguistics.
- Litman, D.; Paletz, S.; Rahimi, Z.; Allegretti, S.; and Rice, C. 2016. The teams corpus and entrainment in multi-party spoken dialogues. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1421–1431.
- Lopes, J.; Eskenazi, M.; and Trancoso, I. 2015. From rule-based to data-driven lexical entrainment models in spoken dialog systems. *Computer Speech & Language* 31(1):87–112.
- Mukherjee, A., and Liu, B. 2012. Analysis of linguistic style accommodation in online debates. *Proceedings of COLING 2012* 1831–1846.
- Namy, L. L.; Nygaard, L. C.; and Sauerteig, D. 2002. Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology* 21(4):422–432.
- Niederhoffer, K. G., and Pennebaker, J. W. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology* 21(4):337–360.
- O’Reilly, C. A.; Caldwell, D. F.; and Barnett, W. P. 1989. Work group demography, social integration, and turnover. *Administrative science quarterly* 21–37.
- Pennebaker, J. W., and King, L. A. 1999. Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology* 77(6):1296.
- Pennebaker, J. W.; Booth, R. J.; and Francis, M. E. 2007. Liwc2007: Linguistic inquiry and word count. *Austin, Texas: liwc.net*.
- Rahimi, Z.; Kumar, A.; Litman, D.; Paletz, S.; and Yu, M. 2017. Entrainment in multi-party spoken dialogues at multiple linguistic levels. *Proc. Interspeech 2017* 1696–1700.
- Reitter, D., and Moore, J. D. 2007. Predicting success in dialogue. In *Proceedings of 45th Annual Meeting of the Association for Computational Linguistics*.
- Rochon, E.; Saffran, E. M.; Berndt, R. S.; and Schwartz, M. F. 2000. Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and language* 72(3):193–218.
- Smith, K. G.; Smith, K. A.; Olian, J. D.; Sims Jr, H. P.; O’Bannon, D. P.; and Scully, J. A. 1994. Top management team demography and process: The role of social integration and communication. *Administrative science quarterly* 412–438.
- Wageman, R.; Hackman, J. R.; and Lehman, E. 2005. Team diagnostic survey: Development of an instrument. *The Journal of Applied Behavioral Science* 41(4):373–398.
- Wendt, H.; Euwema, M. C.; and van Emmerik, I. H. 2009. Leadership and team cohesiveness across cultures. *The Leadership Quarterly* 20(3):358–370.