Chinese and American Leadership Characteristics

Discovery and Comparison in Multi-party On-line Dialogues

Sarah M. Taylor
Sarah M. Taylor Consulting, LLC
121 South Oak St.
Falls Church, VA, USA

Ting Liu, Samira Shaikh, Tomek Strzalkowski, Aaron Broadwell, Jennifer Stromer-Galley, Umit Boz, Xiaoai Ren, and Jingsi Wu
ILS, University at Albany, SUNY
1400 Washington Ave.
Albany, NY, USA

Abstract— **Recent advances in automated analysis of on-line chat data allow us to draw conclusions about social behavior, such as leadership, in small groups previously possible only through manual methods of observation and analysis. We have applied such methods to comparable English and Chinese language data, defined a new language use called Tension Focus, and demonstrate its different effects in the data in these two languages.**

Keywords-dialogue; on-line chat; sociolinguistics; Chinese leadership

I. INTRODUCTION

Recent advances in the automated analysis of digital text are providing increasing opportunities to examine the behavior, sentiments, ideas and interactions of persons, which type of examination has previously been possible only through manual ethnographic methods of observation, surveys, and annotation. We report here on research which advances the computational capabilities necessary for understanding leadership characteristics in multiple cultures, by our transfer of leadership measures developed on English on-line chat data to Chinese on-line chat data. We report the development of a new analytic measure called Tension Focus and illustrate the power of the new capabilities with some indicative comparisons of leadership roles in small Chinese and American discussion groups.

II. PREVIOUS WORK

Broadwell [1] investigate English on-line chat multi-party dialogue and provide a method for modeling mid-level sociolinguistic behaviors, also called Languages Uses (LUs) such as topic control, task control, involvement and disagreement. Higher-level social behaviors, like task leadership or influence, can then be inferred based on the degree to which individual participants exhibit these mid-level sociolinguistic behaviors, and in what combinations. A broadly similar approach to the analysis of multiparty dialogue is being pursued currently by a number of researchers [2] [3] [4]. Our work differs from these partly due to our collection of new data, using social science practice, focused on situations with a high density of interactions between many participants, where we can interview participants to obtain their insights into their behaviors, as well as on groups where leadership behaviors can be expected to emerge clearly and with some frequency. Shaikh [5] describe our data collection methods for on-line chat dialogues, also summarized below, Section 3. Additionally, our approach has the ability to process large extents of dialogue, discovers the behavioral consistency and inconsistencies across sequences of a large number of utterances, and combines many pieces of linguistic evidence to the determination of any particular social behavior [1]. As our methods for understanding social behaviors progress, we are testing them on data where we do not control collection and which are shorter and less dense in interactions.

This developing approach to analysis of on-line dialogue contrasts with earlier work on English dialogue which generally ignored the social import of dialogue interactions and focused instead on either discovering content or on determining the structural components of the dialogue [6] [7] [8]. Additionally, we have been working with multiparty dialogues where participant behaviors and the challenges posed by the language processing differ considerably from work on two-party interactions [9]. Another branch of dialogue investigation has focused on establishing natural dialogue in machine-human interaction settings [10] for question answering from databases or similar tasks. These are, of course, also two party dialogues, with well-defined goals and a need to automatically understand the behavior of only one side of the interaction.

Recent natural language processing research for Chinese text has emphasized problems of word segmentation, POS tagging, and named entity extraction [11] [12]. Investigation of Chinese dialogue has emphasized spoken dialogue [13]. Our work differs from these both in the depth of our Chinese language processing and the data type of on-line chat.

Social science research on leadership in small groups interacting in English suggest that leadership – defined as the ability to guide a group to accomplishment of its task or

otherwise influence the group to act in accordance with the leader's direction [14] – is accomplished through a number of key behaviors: high levels of involvement in the discussion, use of directives, controlling the topics of discussion, and encouraging and excluding other participant contributions [15] [16] [17]. Our work on leadership distinguishes leadership from both influence and authority, either of which may contribute to leadership but which are not by themselves the same behavior.

Recent social science investigations of small group interactions in Chinese groups have focused on the business environment, and the role of leader as a patriarchal figure [18]. Work has emphasized the importance of maintaining harmony within the group [19] and the restraint of dissenting voices or disagreement.

III. DATA COLLECTION

Data collection methods are fully described elsewhere [5] [20]. To summarize, Chinese data and English data have both been collected from native speakers of each language associated with the University and the surrounding communities. Groups range from 4 – 10 people, were trained on the software and task, and were assigned a leader. Discussion sessions were of two types, (a) free-flowing discussion of a contemporary topic, or (b) for most, discussions around assigned tasks, such as the selection of the best US city for a summer Olympics given a set of city profiles, or selection of the best candidate to fill a position, given a job description and resumes. Groups were directed to come to a unanimous decision. Post session surveys asked for participants' assessments of their own and other's behaviors with respect to leadership and influence over the course of the discussions. Each group session lasted ninety minutes and averaged 520 utterances for English and 1189 utterances for Chinese dialogues.

The material in this paper is based upon 16.5 hours of English discussion with a total of 2581 utterances and 10.5 hours of Chinese discussions with a total of 8489 utterances. The larger number of Chinese utterances, in fewer dialogues, is the result of greater numbers of participants in each Chinese discussion group and thus a denser level of activity, for the same time period.

IV. ENGLISH AND CHINESE LANGUAGE COMPONENTS

It is not within the scope of this paper to describe the English and Chinese language processing systems in detail. Both systems are similarly configured, using POS tagging, pronoun resolution and word sense similarity. A defined set of dialogue acts and of communication links are tagged automatically, using components trained on hand-marked data. On the basis of these features, a set of Language Uses are discovered. Higher-level social constructs (or behaviors) such as leadership are determined from this set of LUs using indices and measures described in [1]. This process of calculating higher-level constructs is summarized in the next section.

V. LEADERSHIP MEASURES

Indices of four types, which we combine into the LUs shown in Table 1, are used to determine leadership. These indices are computed based on the automated tagging of the data for part-of-speech (POS), co-reference, dialogue acts, and other lower-level linguistic phenomena [1].

TABLE I.   COMPONENTS OF LEADERSHIP MEASURES

LU	Lower Level Indices
Topic Control	Local Topic Introduction (LTI); Subsequent Mentions of Local Topics (SMT); Cite Score (CSI); Turn Length (TL)
Task Control	Directive Index (DI); Process Management Index (PMI)
Involve-ment	Noun Phrase Index (NPI); Turn Index (TI); Topic Chain Index (TCI); All Subsequent Mentions (ASM); Allotopicality Index (ATP)
Disagree-ment	Disagree-Reject Index (DRX); Topical Disagreement Index (TDX)

Topic Control, for each participant, is based on the number of times a person introduces a new local topic (new noun or concept), how much these topics figure in the following discussion, and the length of turns taken by the participant. Task Control is a measure of each participant's attempts to manage the group's process, calculated from the number of directives given (including questions) and the number of times the participant addresses specific task procedures, such as moving the group to a conclusion. Participant Involvement takes into account her contributions to the content of the dialogue in terms the substantiveness of her discussion (substantive noun phrases used) and the numbers of her turns, as well as further indices capturing the degree to which the speaker contributes comments to topics the group focuses on, introduced by herself or others. Finally, Disagreement is based on the degree to which a speaker disagrees in general, with everyone, plus a pairwise measure looking at disagreement in substance between each two participants on specific topics.

These lower level indices are comparative among participants. We have summarized here only those that contribute to the measurement of leadership. However, these and other measures are also used to detect other group conditions and behaviors, besides leadership. Our development of these indices has been based on a process of working backwards from the often general insights of the social science literature on leadership to define computable linguistic indicators of the behaviors and group interactions which this literature examines.

Employing the LUs just described, composite rankings are computed for each participant to define four leadership measures: Topic Control Measure (TCM, based primarily on the Topic Control LU), Involved Topic Control Measure (ITCM, based on the Topic Control and Involvement LUs), Skilled Control Measure (SCM, based on Task Control LU) and Cumulative Disagreement Measure (CDM, based on Topic or Task Control and Disagreement LUs).

The degree of leadership for each participant (P) is then computed using these measures as in the model shown in Fig. 1.

$$Lp = \alpha 1 \times TCMp + \alpha 2 \times ITCMp + \alpha 3 \times SCMp \times \alpha 4 \ ´ CDMp$$

where TCMp, ITCMp, SCMp, and CDMp take the following values:
   1.0 – if the participant ranks first (highest) on the measure
   0.5 – if the participant ranks second on this measure and is closer to rank 1 than to rank 3
   0.0 – in all other cases
The weights, $\alpha 1$, $\alpha 2$, $\alpha 3$, and $\alpha 4$, are estimated using regression on manually assessed training data for different types of discourse:
   1. In general, for most dialogues:
        $\alpha 1 > \alpha 2$; $\alpha 1$, $\alpha 2$, $\alpha 3 > \alpha 4$
   2. In meetings and games with high incidence of directives:
        $\alpha 3 > \alpha 1$, $\alpha 2$
   3. In open discussions, with few directives:
        $\alpha 1 > \alpha 3$

Figure 1.   Leadership Model.

These four measures allow us to not only determine the leader but to also distinguish between different types or styles of leadership, as suggested in the leadership and group behavior literature [21] [22] [15] [23]. Thus a leader designated on the basis of a high TCM is a thought leader whose influence is apparent in the framing and topics of discussion; one designated on the basis of high ITCM is a leader by virtue of high engagement as well as influence over the topics of discussion; one high in SCM is a task leader focused on managing the group's interactions and progress toward a goal; and one high in either Topic Control Measure or Task Control Measure combined with a high Disagreement LU is considered an assertive leader.

The leadership measures just described were developed based on the chat data and tasks overviewed in Section 3 and described fully in [20]. Table 2 shows a typical calculation of leadership scores for a chat dialogue. Here EA has the highest score on leadership qualities, and his leadership approach is based on Task Control (SCM=0.5) although he is also clearly an assertive leader, with the highest CDM (0.263) as well.

TABLE II.   LEADERSHIP IN A CHAT DIALOGUE

Speaker	TCM	SCM	ITCM	CDM	Leader (ave)
AY	0.084	0.333	0.143	0.117	17%
DE	0.196	0	0.286	0.224	18%
EA	0.251	0.5	0.143	0.263	29%
JN	0.054	0	0.286	0.164	13%
NE	0.401	0	0	0.164	14%
TY	0.013	0.167	0.143	0.067	10%

We have also successfully applied these measures to other types of group interactions in English where we had available data. Fig. 2 shows the results of this transfer of the leadership model to a new type of group, additionally illustrating the

ranking that the leadership measures produce. This group was playing a variant of Second Life during participation in a separate project in our lab; their on-line chat was captured as part of that other experiment.
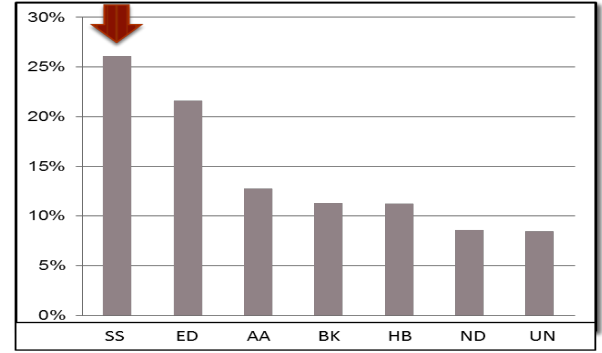

Figure 2.   Leadership in an On-line Group

The leadership model applied to English discussion groups has achieved a high level of accuracy in comparison to human judgments in predicting leadership in multiple kinds of groups. Independent evaluation of the English system applied to previously unseen data, which included both on-line chat and transcripts of lengthy formal meetings, produced an accuracy of 86%. In this evaluation, assessors were given the data (chat and meetings), in addition to the definition of leadership, and instructed to determine the leader. The system accuracy is the percentage of leadership predictions made by the system that matched those of the assessors. In our lab evaluations, we use chat and game participant judgments on a range of survey questions related to leadership behaviors. A correlation ($R^2 = .43$, $p < .001$) between the automated system performance and the average rating by participants of their fellow participants' leadership behaviors further supports the validity of the automated system performance.

VI.   LEADERSHIP IN CHINESE DIALOGUES

Our initial application of these same measures to Chinese discussion groups did not perform as well, achieving only 57% accuracy, suggesting a difference in the behavior of Chinese and American groups, which difference is also noted in the social science literature [24] [25]. In the course of our work, initially in English, on methods to detect the pursuit of power within a group, we defined an additional LU, called Tension Focus (Table 3). Tension Focus reflects the dialogue acts of other members of the group directed toward the speaker. A speaker toward whom more utterances of disagreement, disapproval, or rejection are directed by others or whose statements are questioned by others (DRT) is considered to have a higher degree of Tension Focus. Similarly, even when disagreement is not explicitly directed at the speaker, disagreement with the substance of a recent statement of the speaker (TDT) is also counted in the calculation of Tension Focus.

LU	Lower Level Indices
Tension Focus	Disagree-Reject Target Index (DRT); Topical Disagreement Target Index (TDT)

For DRT, we calculate the proportion of utterances of disagreement, disapproval and rejection, as well as confirmation requests in the discourse that are directed at each speaker as a percentage of all such utterances by all speakers in discourse. These are the utterances typically tagged (automatically) as disagree-reject or confirmation-request dialogue acts, and may include:

- Direct statements of disagreement, either strong (e.g., "I disagree" "That's not right.") or weak ("I agree, but"; "I'm not sure about that");
- Qualified responses to an assertion or opinion from another speaker using a statement containing qualifying markers such as well, but, though, however.
- Utterances in which one speaker alters a statement of another speaker by repeating one part but changing/altering another part (For example: Speaker-A: "Carla has several years of experience as a nanny." Speaker-B: "Carla has one year experience as a nanny.")
- Utterances that appear to question the statements of a prior speaker by requesting confirmation (e.g., *Are you sure?)*

For TDT, we count all unmarked statements by a speaker where there is an earlier (the most recent) statement by another speaker on the same topic but with the opposite polarity. Here, "unmarked" refers to a statement not explicitly tagged as a Disagree-Reject DA or otherwise counted under the DRT measure. The value of TDT for each speaker is the proportion of topical disagreement turns directed at this speaker by all other speakers to the sum of such statements made by all speakers in a discourse. In order to calculate TDT we automatically identify and take into account all utterances that:

- Make either positive or negative statements about a topic, e.g., "*I'm for Carla.*" or "*I would not vote for her.*"
- Offer either supportive or unsupportive information about this topic; e.g., "*She's got experience with youngsters.*" or "*If you are looking for computer skills I would cross her off.*"

We calculate Tension Focus for both English and Chinese. For English dialogues it makes no difference in selection of the leader, although we are testing its utility in the prediction of other behaviors. However, addition of the Tension Focus LU to leadership predictions in Chinese, as a negative indicator, put the performance on a par with English (86%). For Chinese dialogues, when other measures fail to determine a leader, the leader is selected from the top two contenders based on which participant has the lower Tension Focus.

VII. COMPARISONS AND FINDINGS

Because of this interesting difference in the way in which groups and leaders behave, between the Chinese and American groups, we explored further comparisons of the relationship between Tension Focus and the other leadership measures. From this we draw indicative, if highly preliminary, conclusions about the typical behaviors of the Chinese and American leaders and their groups. In the comparison reported here (Fig. 3), Chinese leaders are less likely to be a focus of disagreement than American leaders. This is illustrated in the first two columns on the left, and considers leadership calculated using all four measures of our model as discussed in Section 5. By looking inside at the components of that model, we see the differences in more detail, as shown in columns 3/4 and 5/6. In Chinese dialogues, disagreement is far less likely to be focused on the person with the highest control over the task, than in the American groups (columns 3 and 4). Task Control is also the more common mode of the Chinese leaders. On the other hand, in Chinese groups disagreement is still expressed but directed at the person
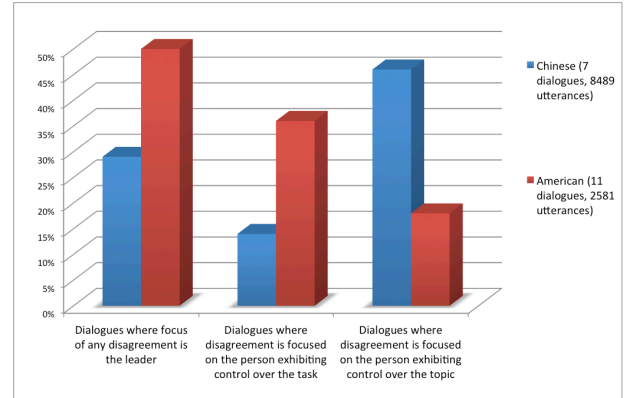

Figure 3.   Comparison of Role of Tension Focus in Chinese and American Chat Dialogues

who has taken the lead in controlling the content of the discussion. The exactly opposite behavior is shown in the American dialogues. These indications support previous social science research based upon observation suggesting high power distance for Chinese leaders [18], which would suggest unwillingness to directly challenge leadership and an expectation that the leader will provide strong guidance in the pursuit of the group agenda. However, it contrasts with previous research which suggests that disagreement is strongly muted in Chinese group interactions [26].

conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoD/ARL, or the U.S. Government.

REFERENCES

[1] G. A. Broadwell, T. Strzalkowski, J. Stromer-Galley, S. Shaikh, S. Taylor, T. Liu, U. Boz, A. Elia, L. Jiao, and N. Webb. "Modeling socio-cultural phenomena in discourse," Journal of Natural Language Engineering 2012, pp. 1-45.

[2] W. Wang, S. Yaman, K. Precoda, C. Richey, and G. Raymond. "Detection of agreement and disagreement in broadcast conversations," Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Human Language Technologies, 2011.

[3] M. Freedman, A. Baron, V. Punyakanok, and R.Weischedel. "Language use: what can it tell us?" Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Human Language Technologies, 2011.

[4] A. Maren and M. Ostendorf. "Detecting forum authority claims in online discussions," Proceedings of the HLT-NAACL 2011 Workshop on Language and Social Media, 2011.

[5] S. Shaikh, T. Strzalkowski, S. Taylor and N. Webb. "MPC: a multi-party chat corpus for modeling social phenomena in discourse," Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC2010), Valletta, Malta, 2010.

[6] S. Carberry and L. Lambert. "A process model for recognizing communicative acts and modeling negotiation dialogue," Computational Linguistics 25(1), pp. 1-53, 1999.

[7] A. Stolcke, K. Reis, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. Van Ess-Dykema, and M. Meteer. "Dialogue act modeling for automatic tagging and recognition of conversational speech," Computational Linguistics 26(3), pp. 339-73, 2000.

[8] N. Blaylock. "Managing communicative intentions in dialogue using a collaborative problem-solving model," Technical Report 774, CS Department, University of Rochester, Rochester, New York, 2002.

[9] E. Mayfield and C. P. Rose. "Recognizing authority in dialogue with an integer linear programming constrained model," Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland, Oregon, 2011, pp. 1018-1026.

[10] M. G. Core, J. D. Moore and C. Zinn. "The role of initiative in tutuorial dialogue," EACL'03, Proceedings of the Tenth Conference of the European Chapter of the Association for Computational Linguistics, Budapest, Hungary, 2003, pp. 67-74.

[11] R. Sproat and T. Emerson. "The first international Chinese word segmentation bakeoff," SIGHAN 03, Proceedings of the Second SIGHAN Workshop on Chinese Language Processing, Volume 17, Association for Computational Linguistics, 2003.

[12] O. O. Y. Kwong and H. Li. SIGHAN 08, Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing, Hyderabad, India, 11-12 January, 2008.

[13] C. H. Lee. Advances in Chinese spoken language processing, B and JO Enterprise, Singapore, 2007.

[14] M. A. Hogg and S. A. Reid. "Social identity, self categorization, and the communication of group norms," Communication Theory 16:7-30, 2006.

[15] D. G. Ellis and B. A. Fisher. Smal group decision making: Communication and the group process. McGraw-Hill, New York, 1994.

[16] S. A. Reid and S. H. Ng. "Language, power and intergroup relations," Journal of Social Issues 55(1), pp. 119-139, 1999.

[17] A. Pomerantz and P. Denvir. "Enacting the institutional role of chairperson in upper management meetings: The interactional realization of provisional authority," In F. Cooren, ed., Interacting and organizing: Analyses of a management meeting. Lawrence Erlbaum Assoc., Mahwah, NJ, 2007.

[18] B. S. Cheng, L. F. Chou, T. Wu, M. P. Huang, and J. L. Farh. "Paternalistic leadership and subordinate responses: Establishing a leadership model in Chinese organizations," Asian Journal of Social Psychology 7, pp.89-117, 2004.

[19] K. Leung, F. P. Brew, Z. X. Zhang and Y. Zhang. "Harmony and conflict: a cross- cultural investigation in China and Australia," Journal of Cross Cultural Psychology, 42(5) , pp. 495-816, 2011.

[20] T. Liu, S. Shaikh, T. Strzalkowski, A. Broadwell, J. Stromer-Galley, S. Taylor, U. Boz, X. Ren, and J. Wu. "Extending the MPC corpus to Chinese and Urdu - a multiparty multi-lingual chat corpus for modeling social phenomena in language," Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC2012), Istanbul, Turkey 2012.

[21] L. P. Bradford. Group Development, University Associates, La Jolla, California, 1978.

[22] S. A. Beebe and J. T. Masterson. Communicating in small groups: Principles and practices, Pearson/Allen and Bacon, Boston, 2006.

[23] R. Sanders, A. Pomerantz, and J. Stromer-Galley. "Some ways of taking issue with another participant during a group deliberation," Paper Presented at the Annual Meeting of the National Communication Association, San Francisco, 2010.

[24] M. T. Chen. "An interlanguage study of the speech act of disagreement made by Chinese EFL speakers in Taiwan," Thesis, National Sun Yat-Sen University, Taipei, 2006.

[25] G. Hofstede and M. H. Bond. "The Confucius connection: from cultural roots to economic growth," Organizational Dynamics 1988, pp. 5-21.

[26] D. Tjosvold, K. S. Law and H. Sun. "Effectiveness of Chinese teams: the role of conflict types and conflict management approaches," Management and Organization Review 2(2), pp. 231-252, 2006.