MPC: A Multi-Party Chat Corpus For Modeling Social Phenomena In Discourse

Samira Shaikh[1], Tomek Strzalkowski[1, 2], Aaron Broadwell[1], Jennifer Stromer-Galley[1], Sarah Taylor[3], and Nick Webb[1]

[1]ILS Institute, University at Albany, State University of New York
[2]Institute of Computer Science, Polish Academy of Sciences
[3]Advancded Technology Office, Lockheed Martin IS&GS
E-mail: ss578726@albany.edu, tomek@albany.edu

Abstract

In this paper, we describe our experience with collecting and creating an annotated corpus of multi-party online conversations in a chat-room environment. This effort is part of a larger project to develop computational models of social phenomena such as agenda control, influence, and leadership in on-line interactions. Such models will help capturing the dialogue dynamics that are essential for developing, among others, realistic human-machine dialogue systems, including autonomous virtual chat agents. In this paper we describe data collection method used and the characteristics of the initial dataset of English chat. We have devised a multi-tiered collection process in which the subjects start from simple, free-flowing conversations and progress towards more complex and structured interactions. In this paper, we report on the first two stages of this process, which were recently completed. The third, large-scale collection effort is currently being conducted. All English dialogue has been annotated at four levels: communication links, dialogue acts, local topics and meso-topics.

1. Introduction

Multi-party online conversation has become a pervasive form of communication within virtual communities. The popularity of social networking sites has made such communication ubiquitous across all age groups. This phenomenon creates a vast amount of conversational data, which may be utilized for studying a wide spectrum of linguistic and social phenomena and could be exploited in support of various NLP tasks. In particular, a great amount of communication within an online community occurs in virtual chat-rooms, where users log in and converse with other users who may be online at that time. Conversations are typically conducted using free form, highly informal text dialect.

While chat data is plentiful on-line, its adaptation for research purposes presents a number of challenges that include users' privacy issues on the one hand, and their complete anonymity on the other. Furthermore, most data that may be obtained from public chat-rooms is of limited value for the type of modeling tasks we are interested in due to its high-level of noise, lack of focus, and rapidly shifting, chaotic nature, which makes any longitudinal studies virtually impossible. Public chat-rooms may be excellent sources of data for studies involving on-line language usage (e.g., novel uses of vocabulary, syntax), general conversational etiquette, and related issues. However, for deriving more complex models of conversational behavior, we need the interaction to be reasonably focused on a task and/or social objectives within a group.

In order to obtain a suitable dataset we designed a series of experiments in which recruited subjects were invited to participate in a series of on-line chat sessions in a specially designed secure chat-room. Participants were selected from among current and past University students and staff based on their general level of experience with chat communication, but otherwise representing fairly diverse demographics and backgrounds. Whenever possible, we interviewed the candidates to make sure they would be comfortable with various roles we envisioned for them, including the nominal conversation lead, as well as with assuming any emergent and opportunistic roles, such as a challenger, a supporter, a disruptor, etc.

The purpose of this collection was two-fold: (1) understanding how certain social behaviors are reflected in language, and (2) building an automated chat agent that could effectively achieve certain (initially limited) social objectives in the chat-room.

This required a careful design of the experiments around topics, tasks, and games for the participants to engage in so that appropriate types of behavior, e.g., disagreement, power play, persuasion, etc. may emerge spontaneously.

Obtaining high-quality conversational corpora with such complex characteristics is inherently difficult. The foremost consideration here is to make sure that the conversation appears as natural as possible given that the entire setup is, in fact, artificial and that the subjects are well aware that their interactions are recorded. Moreover, there is a distinct lack of motivation or incentives for the subjects to engage in more effective but risky behaviors.

In order to mitigate these concerns we have devised a multi-tiered collection process in which the subjects start from simple, free-flowing conversations and progress towards more complex and structured interactions. In this paper, we report on the first two stages of this process, which were recently completed. The third, large-scale collection effort is currently being conducted. Details of the experimental design are discussed in section 3.

The initial two stages of data collection comprised of 14 sessions of English chat dialogue conducted in groups ranging in size from 3 to 8. We have also conducted 14 sessions of chat with participants conversing in Roman Urdu, which constitutes a part of our Urdu collection. In this paper we discuss English chat data only.

All English dialogue has been annotated at four levels: communication links, dialogue acts, local topics and meso-topics. Some details of these annotations will be discussed later in this paper, although a full description is impossible within the scope of this article. It is important to note that the annotation has been developed to support the objectives of our project and does not necessarily conform to other similar annotation systems used in the past, for example dialogue act tagging.

2. Related Work

Much research has been undertaken to create corpora in support of dialogue research; however, most available collections are spoken language conversations involving two participants. Few data collections exist covering multi-party dialogue, and even fewer with on-line chat. Moreover, the few collections that exist were built primarily for the purpose of training dialogue act tagging and similar linguistic phenomena; few if any of these corpora are suitable for deriving pragmatic models of conversation, including socio-linguistic phenomena.

Previous work on the study of dialogue phenomena concentrated on two person interactions, both task-focused, such as Map Task (Anderson et al., 1991) and open conversation, as in the annotated portion of the Switchboard corpus (Jurafsky et al., 1997), as well as languages other than English, such as Spanish CALLHOME (Levin et al., 1998) or the NESPOLE speech to speech translation corpus of German, French, Italian and English (Levin et al., 2003).

Recently, work has expanded to include multi-person meetings (such as the ICSI-MRDA corpus) and include a wider range of modalities. For example, the AMI corpus stems from a European research project centered on multi-modal meeting room technology. The AMI Meeting Corpus (Carletta, 2007) contains 100 hours of meetings captured using many synchronized recording devices. All of these resources look at spoken language. There is a parallel interest in the online chat environment, although the development of useful resources has progressed less. The NPS chat corpus (Forsyth and Martell, 2007) is a corpus of around 10,000 postings from age specific chat rooms on the internet, which have been hand-anonymized and labeled with part-of-speech tags and dialogue act labels. The NPS corpus is freely distributed as part of the Natural Language Toolkit (Bird et al., 2009). The StrikeCom corpus (Twitchell et al., 2007) is a corpus of 32 multi-person chat dialogues between players of a strategic game, where in 50% of the dialogues one participant has been asked to behave 'deceptively'.

It is more typical that those interested in the study of Internet chat compile their own corpus on an as needed basis, such as the work of Wu et al. (2002) and Khan et al. (2002) on IRC chat rooms, the work of Kim et al. (2007) on student discussion boards or the study of online conversations between two people, a customer and a shopping assistant, used in the dialogue act annotation effort of Ivanovic (2005).

3. Experiment Design

We collected approximately 20 hours of chat dialogue spread out over 14 sessions of 90 minutes each, amounting to a total of 7317 individual utterances. There were, on average, 5 participants present in each session. In this section we describe how the data collection process was accomplished. In the next section we discuss the characteristics of the dataset collected.

3.1 Subjects

Subjects were recruited from within the University community, and consisted of students and alumni, as well as research staff, including junior faculty. We sent out an email recruiting messages on the mailing lists for a few departments including Computer Science, Information Science and Communication, following the guidelines set out by the University IRB protocol regarding human subject experiments. For the purposes of our research, we wanted to have a minimum of 4 participants for each chat session. We started with a pool of 13 respondents to our initial recruitment message. We asked the participants to fill out a simple demographic questionnaire, to allow us to make hypotheses about the correlation of socio-linguistic phenomena with, for instance, participant age. Participant age varied from 22 years to 55 years with average age being 34 years, with 7 males and 6 females. Participants were compensated for their time.

3.2 The Chat-room Setup

We developed a chat server and client for the purpose of this data collection. The communication between the server and the client is over simple HTTP protocol. Both programs are written in Java, where the chat client is a Java applet, with an interface that is similar to popular chat clients, and can be accessed using any web browser. We have since replaced this chat environment with an XMPP based client-server setup, which we are using for further data collection. Participants were assigned unique nicknames and given secure login access to the chat server, which could be accessed via the web from any remote location.

3.3 Chat Sessions

We conducted a series of 14 chat sessions divided into two phases. Each phase consisted of 7 sessions, or approximately 10 hours of discourse. We posted a chat session schedule and participants would sign up for as many sessions as were convenient to them. For each session we had an average of 5 users online at the same time, including a nominal "leader" who was responsible for keeping the discussion on a particular topic for the duration of the session (90 minutes).

During the first phase of collection, we let the participants to volunteer as leaders and to choose any topic they wished to discuss, but beyond that they were free to converse in any way that felt most comfortable. This phase produced some good, lively conversations that helped to establish initial relationships between the participants, and set the ground for more structured discourse in the second phase.

During the second phase, we gave the participants a specific topic to discuss or a task to perform. For example, the topic could be "Should Dick Cheney and others be prosecuted for their role in using torture?" or "What is your opinion of the government bailout of the Ameri-

can auto-industry?" For a specific task based dialogue, we had the participants form a search committee and select the best candidate for a job from a list of fictional resumes. As in Phase 1, there was a nominal leader whose job was to keep conversation on the topic, but now this assignment required significantly more skill. In Phase 2 dialogues, we observed a marked increase of social phenomena including disagreements, agenda control, and varying degrees of involvement among the participants. In some sessions, alliances formed and discussion leaders emerged quite separate from the nominal chat leads. We are currently analyzing this data towards a formal assessment on how frequently and in which manner these social phenomena occur.

One specific phenomenon we wanted to model was an effective change of conversation topic, when a participant or a group of participants deliberately (if perhaps only temporarily) shift the discussion to a different, possibly related topic. Both success and failure of this action were of interest because the outcome depended upon the choice of utterance, the persons to whom it was addressed, their reaction, and the time when it was produced. In a few dialogues, we gave selected participants "hidden" roles. One role, which we may call a "disruptor" was to opportunistically introduce a secondary topic, somewhat related to the discussion topic, but not directly. Another possible hidden role was that of a consensus breaker where the purpose was to split the group into camps.

We gave the participants who were selected for the particular roles such as leader, disruptor, and consensus breaker only a general outline of what these roles should accomplish. The participants were free to play out these roles in any manner they wished, and only when a suitable opportunity presented itself.

4. Data Statistics

The basic statistical information about the collected data set, which we shall refer to as Multi-Party Chat (MPC) corpus, is given in Table 1 and Table 2. As already indicated earlier, the current data set represents only a fraction of a larger corpus currently under development.

Total turns in chat corpus	Total words in corpus	Average Words/Turn	Total Emoticons/Misspellings/ Abbreviations
7317	58175	8	241/683/1362

Table 1: Turn total statistics from 14 sessions

Avg. Participants per session	Average Turns per session	Average Turns Per User	Max/Min Turns per session
5	520	100	165/47

Table 2: Turn average statistics from 14 sessions

In Table 1, we use emoticons to mean a sequence of characters commonly used to signify emotions in chat, such as a smiley face. Misspellings are different from chat-speak, and can be a result of typing errors and non-standard abbreviations.

While we are analyzing the data in detail for the kinds of social phenomena reflected in language use, we collected various statistics on the linguistic, syntactic and grammatical properties of the utterances in our corpus. In particular, we were interested in the rates of use of emoticons, chat-speak (words that are part of chat room jargon such as 'imho' or 'lol'), punctuation, as well as presence of misspellings, ratio of content words to stop words, and number of words per utterance. Figure 1 shows some of these features for each participant averaged across all sessions. Figure 1 shows an interesting trend of how the use of emoticons may be related to the occurrence of misspellings for a user. This trend holds true for all participants except participant P6.

We also compared some simple characteristics of nominal conversation leaders against those of other participants in the discussion. For example, we measured the leader verbosity, as the number of turns multiplied by the amount of words in a turn. The chart in Figure 2 indicates that there may be a correlation between verbosity and a leading role in a discussion. We use the term verbosity to be a measure of turn length times the number of turns for that user.

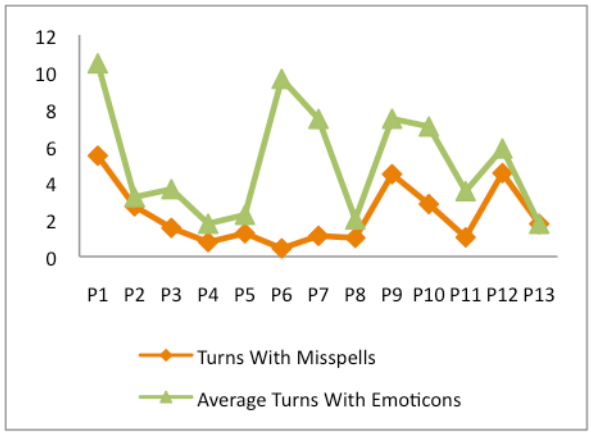

Figure 1: Turns with misspelling versus turns with emoticons

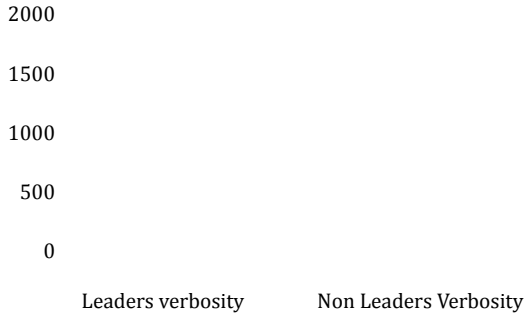

Figure 2: Leader vs. Non-Leader Verbosity

While these superficial statistics are comparatively easy to compute, we are interested in assessing their correlation with more advanced language use factors, such dialogue acts, communicative links and topic and focus changes that are known to be predictive of the types of social phenomena we wish to detect. In the next section we briefly outline the corpus annotation process applied

to the MPC corpus.

5. Modeling Social Phenomena in Dialogue

We are interested in modeling the social phenomena of Leadership and Power in discourse. These high-level phenomena (or Social Roles, *SR*) will be detected and attributed to discourse participants based on their deployment of selected *Language Uses* (LU) in multi-party dialogue. Language Uses are mid-level socio-linguistic phenomena that link linguistic constructs deployed in discourse (from lexical to pragmatic) to social relations obtaining between the participants. Examples of such language uses that we are currently studying are: *Agenda Control*, *Disagreement*, and *Involvement* (Broadwell et al., 2010).

Our research so far is focused on the analysis of English-language synchronous chat, and we are looking for correlations between various metrics that can be used to detect LU in multiparty dialogue. We are well aware that some of these correlations may be culture-specific or language-specific, and we are also looking for changing patterns as we move into the analysis of Urdu and Mandarin discourse in later phases of this project.

5.1 Agenda Control in Dialogue

Agenda Control is defined as efforts by a member or members of the group to advance the group's task or goal. This is a complex LU that we will model along two dimensions (which may be viewed as distinct behaviors, or language uses): (1) *Topic Control* and (2) *Task Control*. Topic Control refers to attempts by any discourse participants to impose the topic of conversation. Task Control, on the other hand, is an effort by some members of the group to define the group's project or goal and/or steer the group towards that goal. We believe that both behaviors can be detected using scalar measures per participant based on certain linguistic features in their utterances.

For example, one hypothesis is that topic control is indicated by the rate of *local topic introductions* (LTI) per participant (Givon, 1983). Local topics may be defined quite simply as noun phrases introduced into discourse, which are subsequently mentioned again via repetition, synonym, pronoun, or other form of co-reference. Thus, one measure of topic control is the number of local topics introduced by each participant as percentage of all local topics in a discourse. Similarly, we are testing the hypothesis that Task Control can be measured by the rate of directive and process management speech acts that participants deploy in discourse.

Using LTI index we can construct assertions about topic control in a discourse. For example, based on the following information about speaker LE in a multi-party discussion:

1. LE introduces 23/90 (25.6%) of local topics in a dialogue.
2. The mean rate of local topic introductions is this dialogue is 14.29%, and standard deviation is 8.01.
3. LE is in the top quintile of participants for introducing new local topics

We can claim the following, with a degree of confidence (to be determined experimentally):

TopicControl (*LE, 5, dialogue-1*)

In addition to LTI, we have defined several other metrics for topic control and for task control (which we don't have space to explain in detail here). Each of these provide an additional source of evidence that the targeted language use is present in the discourse. We are currently working on how these different metrics correlate to each other and how they should be weighted to maximize accuracy of making LU claims.

5.2 Disagreement in Dialogue

There are two ways in which disagreement is expressed: *expressive disagreement* and *topical disagreement* (Stomer-Galley, 2007; Price, 2002). Both can be detected using scalar measures applied to subsets of participants, typically any two participants. In addition, we can also measure for each participant the rate of generating disagreement (with any and all other speakers). *Expressive Disagreement* is normally understood at the level of dialogue acts, i.e., when discourse participants make explicit utterances of disagreement, disapproval, or rejection in response to a prior speaker's utterance. Here is an example (KI and KA are two speakers in a multiparty dialogue):

KA: *CARLA... women are always better with kids*
KI: *That's not true!*
KI: *Men can be good with kids too*

Our hypothesis is that one measure of Expressive Disagreement is the number of Disagree-Reject dialogue acts between any two speakers as a percentage of all utterances between these two speakers.

Topical disagreement is defined as a difference in referential *valence* in utterances (statements, opinions, questions, etc.) made on a topic. Referential valence of an utterance is determined by the type of statement made about the topic in question, which can be: positive (+), negative (–), or neutral (0). A positive statement is one in favor of (*express advocacy*) or in support of (*supporting information*) the topic being discussed. A negative statement is one that is against or negative on the topic being discussed. A neutral statement is one that does not indicate the speaker's position on the topic. Here is an example of opposing polarity statements about the same topic in discourse:

Sp-1: *I like that he mentions "Volunteerism and Leadership"*
Sp-2: *but if they're looking for someone who is experienced then I'd cross him off*

Detecting topical disagreement in discourse is more complicated because its strength may vary from one topic in a conversation to the next. A reasonable approach is thus to measure the degree of disagreement between two speakers on one topic first, and then extrapolate over the entire discourse. Accordingly, our first hypothesis is that a measure of topical disagreement between two speakers is valuation differential between these speakers as expressed in their utterances about a topic. Here, the topic (or an "issue") is understood more narrowly than the local topic defined in the previous subsection, and may be assumed to cover only the most

persistent local topics, i.e., topics with the largest number of references in dialogue. The resulting *Topical Disagreement Metric* (TDM) captures the degree to which these two speakers advocate the opposite sides of a topic. TDM is computed as an average of *P*-valuation differential for one speaker (advocating *for* a topic) and (*–P*)-valuation differential for the other speaker (advocating *against* the topic). This metric is then extended by averaging it over all relevant topics in discourse into the *Averaged Topical Disagreement Metric* (ATDM).

Using TDM index we can construct assertions related to disagreement in a given multiparty dialogue of sufficient duration (exactly what constitutes a sufficient duration is being researched). Here is an example based on a 90-minute chat dialogue about several job candidates for a YMCA youth counselor. The discussion involved 7 participants, including KI and KA. Topical disagreement is measured on 5 points scale (corresponding to quintiles in normal distribution):

TopDisAgree (*KI, KA, "Carla", 4, YMCA*)

This may be read as follows: speakers KI and KA *topically disagree to degree 4* on topic [job candidate] "Carla" in YMCA discussion. In order to calculate this we compute the value of TDM index between these two speakers. We find that KA makes 30% of all positive utterances made by anyone about Carla (40), while KI makes 45% of all negative utterances against Carla. This places these two speakers in the top quintiles in the "for Carla" valuation distribution and "against Carla" valuation distribution, respectively. Taking into account all opposing polarity statements by KA against Carla and by KI made for Carla, we calculate the level of topical disagreement between KA and KI to be 4 on the scale 1 to 5.

6. Annotations

We wish to annotate the data we collected in order to derive models for language use related to disagreement, involvement, agenda control, and eventually for social roles such as leadership. All of the above represent complex pragmatic concepts that are difficult to annotate directly, let alone detect automatically. Our approach is thus to build a multi-level annotation scheme, where each lower (component) level annotation supplies evidence that supports a claim that some higher-level phenomenon is present. The indices described in the previous section are examples of mappings between linguistic features and social phenomena in discourse. In this paper we briefly outline only basic component level annotation that consists of four interleaved layers: communicative links, dialogue acts, local topic tracking, and meso-topic valences. A more detailed description of the annotation scheme is available in (Shaikh et al., 2010).

6.1 Communicative Links

One of the challenges in multi-party dialogue is to establish which user an utterance is directed towards. Users do not typically add addressing information in their utterances, which leads to ambiguity while creating a communication link between users. With this annotation level, we asked the annotators to determine whether each utterance was addressed to some user, in which case they were asked to mark which specific user it was addressed to; was in response to another prior utterance by a different user which required marking the specific utterance responded to; or a continuation of the user's own prior utterance.

Communicative link annotation allows for accurate mapping of dialogue dynamics in the multiparty setting, and is a critical component of tracking such social phenomena as disagreement and speaker power.

6.2 Dialogue Acts

We developed a hierarchy of 21 dialogue acts for annotating the functional aspect of the utterance in discussion. The tagset we adopted is loosely based on DAMSL (Allen & Core, 1997) and SWBD (Jurafsky et al., 1997), but greatly reduced and also tuned significantly towards dialogue pragmatics and away from more surface characteristics of utterances. In particular, we ask our annotators what is the pragmatic function of each utterance within the dialogue, a decision that often depends upon how earlier utterances were classified. Thus augmented, DA tags become an important source of evidence for detecting language uses and such social phenomena as leadership. Examples of dialogue act tags include Assertion-Opinion, Acknowledge, Information-Request, Confirmation-Request.

Using the augmented DA tagset also presents a fairly challenging task to our annotators, who need to be trained for many hours before an acceptable rate of inter-annotator agreement is achieved. For this reason, we consider our current DA tagging as a work in progress.

6.3 Local Topics

Local topics are defined as nouns or noun phrases introduced into discourse that are subsequently mentioned again via repetition, synonym, or pronoun. Any content-bearing noun or noun phrase can be used to introduce a new local topic, and there may be one of more local topics introduced in each dialogue turn. Tracking local topics and their subsequent mentions is constructive in detecting such social language uses as Topic Control and Involvement

For this kind of annotation, we are paying attention to noun phrases in the dialogue, and we would like to know when speakers refer back to a previously mention item. The annotator is asked to consider each noun phrase in the dialogue and decide whether this noun phrase is new to the dialogue, or whether it is a subsequent mention of some previously mentioned item. We are excluding 1st and 2nd person pronouns (I, me, my, we, us, our, you, your) and names of the participants in the dialogue from this coding. So if the participants in the chat are named *Bob, Joe*, and *Fred,* we are not marking them as local topics. After a local topic has been introduced into discourse, it can be referred to again in subsequent mentions. Consequently, we have two main tags in this category: *New Local Topic* and *Subsequent Mention of the Local Topic*.

6.3.1 New Local Topics

Any content-bearing noun or noun phrase can be used to introduce a new local topic, and there may be one of more local topics introduced in each dialogue turn.

6.3.2 Subsequent Mentions of Local Topics

Local topics that are subsequently mentioned in dialogue through repetition are tagged as Subsequent Mentions. A local topic may be mentioned again by repeating the same noun phrase that was used to introduce it or its part, as long as such a reference is unambiguous. A local topic may also be subsequently mentioned by using a synonymous expression, or by using a pronoun. We code each of these cases separately by marking them as either Subsequent Mention by Repetition or Synonym; or as Subsequent Mention by Pronoun.

6.4 Meso-Topics and their Valences

While most local topics have low granularity and they tend to come and go as the discourse progresses, some topics, which we call *meso-topics*, will persist through a number of turns and become focus of a part of conversation. A selection of meso-topics is closely associated with the task in which the discourse participants are engaged. For example, when the task is to select a candidate for a job, the name of each applicant becomes a meso-topic. Meso-topics can be distinguished from the local topics because the speakers often make polarized statements about them. An utterance is polarized if it expresses sentiment or valence that a speaker assigns to the meso-topic. Valence can be positive or negative, or in absence of any obvious polarity, it may be neutral. A *positive polarity* tag is used when an utterance is expressly in favor of the meso-topic, or if it supplies favorable or supporting information about it. A *negative polarity* tag is used when an utterance is expressly against the meso-topic, or if it supplies unfavorable or negative information about it. If an utterance is neither positive nor negative the *neutral polarity* tag is used.

sessions /total utterances annotated	Most frequent CL per session	Most frequent dialogue act per session	Meso topic annotations per session
8/4640	Response-to (256)	Assertion-opinion (222)	220

Table 3: Annotation statistics for MPC corpus.

Communication Link	Total
Addressed-to	5160
Response-to	7618
Continuation-of	3371

Table 4: Total frequency of communicative links annotated in MPC corpus

Dialogue Act	Total occurrences
Assertion-Opinion	5346
Acknowledge	1315
Information-Request	984
Agree-Accept	966
Positive-Answer	944
Explanation	741
Confirmation-Request	593
Communication-Management	508

Table 5: Total frequency of dialogue acts annotated in MPC corpus

Meso-Topic Valence	Counts
Positive	596
Negative	435
Neutral	512

Table 6: Valence of meso-topics annotated in MPC corpus

Local Topics Identified	Number of Annotators
410	3

Table 7: Unique local topics identified in a single session

Table 3 summarizes the statistics of the annotated MPC corpus. The first column is the number of sessions we have annotated so far, and the total number of utterances in those sessions. The most frequent communicative link assigned by the annotators was a 'response-to' CL, as listed in the second column (256 times per session). The third column shows that of the 20 dialogue act tags we have developed for this corpus, the most frequently assigned tag was the 'Assertion-Opinion' tag (222 average frequency). The fourth column in Table 3 shows the average number of meso-topics identified in a session by the annotators. Table 4 shows distribution of communicative links assigned by annotators. Table 5 shows DA tag distribution; besides Assertion-Opinion, other frequent tags include 'Acknowledge', 'Agree-Accept', 'Information-Request' that were assigned to utterances on average 50 times per session. Table 6 shows how frequently positive, negative and neutral valence have been assigned for meso-topics in annotated corpus. In Table 7, the number of unique local topics identified in one selected session annotated by 3 annotators is shown.

7. Discussion

In this paper we described the first two phases of building a chat corpus for specific research goals. This is a work in progress. While we continue to refine the experimental design for data collection we are encouraged by the properties of the emerging corpus. Our intention is to make this corpus available to the research community once the collection and annotation process is complete.

8. Acknowledgements

This work has been supported in part by the Intelligence Advanced Research Projects Activity (IARPA) via Army Research Laboratory (ARL). The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Also, development of the MPC corpus has been supported in part by grants from the National Institute of Justice and Lockheed Martin Corporation. The views and 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, ARL, or the U.S. Government.

References

Allen, J. M. Core. (1997). Draft of DAMSL: Dialog Act Markup in Several Layers. http://www.cs.rochester.edu/research/cisd/resources/damsl/

Anderson, A., M. Bader, E. Bard, E.Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson and R. Weinert. (1991). The HCRC Map Task Corpus. Language and Speech 34(4), 351--366.

Bird, Steven, Klien, Ewan and Loper, Edward. (2009). Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media.

Broadwell, Aaron; Jennifer Stromer-Galley, Tomek Strzalkowski, Sarah Taylor, Umit Boz, Alana Elia, Samira Shaikh, and Nick Webb. (2010). Social Phenomena and Language Use. Technical Report, ILS Institute, SUNY Albany.

Carletta, J. (2007). Unleashing the killer corpus: experiences in creating the multi-everything AMI Meeting Corpus. Language Resources and Evaluation Journal 41(2): 181-190

Eric N. Forsyth and Craig H. Martell. (September 2007). Lexical and Discourse Analysis of Online Chat Dialog. In Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007), pp. 19-26.

Givon, Talmy. (1983). Topic continuity in discourse: A quantitative cross-language study. Amsterdam: John Benjamins.

Ivanovic, Edward. (2005). Dialogue Act Tagging for Instant Messaging Chat Sessions. In Proceedings of the ACL Student Research Workshop. 79–84. Ann Arbor, Michigan.

Jurafsky, Dan, Elizabeth Shriberg, and Debra Biasca. (1997). Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual. http://stripe.colorado.edu/~jurafsky/manual.august1.html

Jurafsky, D., R. Bates, N. Coccaro, R. Martin, M. Meteer, K. Ries, E. Shriberg, A. Stolcke, P. Taylor, and C. Van Ess-Dykema. (1997). Automatic detection of discourse structure for speech recognition and understanding. In Proc. of IEEE Workshop on Speech Recognition and Understanding, Santa Barbara.

Khan, Faisal M., Todd A. Fisher, Lori Shuler, Tianhao Wu and William M. Pottenger. (2002). Mining Chat-room Conversations for Social and Semantic Interactions. Computer Science and Engineering, Lehigh University.

Kim, Jihie., Erin Shaw, Grace Chern, and Donghui Feng. (2007) An Intelligent Discussion-Bot for Guiding Student Interactions in Threaded Discussions. In Proceedings of the AAAI Spring Symposium on Interaction Challenges for Intelligent Assistants

Levin, L., A. Thyme-Gobbel, A. Lavie, K. Ries, and K. Zechner. (1998). A discourse coding scheme for conversational Spanish. In Proceedings of the International Conference on Speech and Language Processing.

Levin, L., C. Langley, A. Lavie, D. Gates, and D. Wallace. (2003). Domain specific speech acts for spoken language translation. In Proceedings of 4th SIGdial Workshop on Discourse and Dialogue.

Price, V., Capella, J. N., & Nir, L. (2002). Does disagreement contribute to more deliberative opinion? Political Communication, 19, 95-112.

Shaikh, S.; T. Strzalkowski, A. Broadwell, J. Stromer-Galley, N. Webb. U. Boz, A. Elia, K. Stahl. (2010). DSARMD Annotation Guidelines, V. 2.5. Technical Report, ILS Institute, SUNY Albany.

Stromer-Galley, J. (2007). Measuring deliberation's content: A coding scheme. Journal of Public Deliberation, 3(1).

Tianhao Wu, Faisal M. Khan, Todd A. Fisher, Lori A. Shuler and William M. Pottenger. (2002). Posting Act Tagging Using Transformation-Based Learning. In the Proceedings of the Workshop on Foundations of Data Mining and Discovery, IEEE International Conference on Data Mining

Twitchell, Douglas P., Jay F. Nunamaker Jr., and Judee K. Burgoon. (2004). Using Speech Act Profiling for Deception Detection. In Proceedings of Intelligence and Security Informatics, Lecture Notes in Computer Science, Vol. 3073