Chunks in Multiparty Conversation - Building Blocks for Extended Social Talk

Emer Gilmartin, Benjamin R. Cowan, Carl Vogel, Nick Campbell

Abstract Building applications which can form a longer term social bond with a user or engage with a group of users calls for knowledge of how longer conversations work. This paper describes preliminary explorations of the structure of long (c. one hour) multiparty casual conversations, focusing on a binary distinction between two types of interaction phases – chat and chunk. A collection of long form conversations which provide the data for our explorations is described. The main result is that chat and chunk segments show differences in the distribution of their duration.

1 Introduction

Increasing interest in socially competent artificial spoken dialogue calls for clearer understanding of the mechanisms and form of human casual and social conversation. This knowledge can facilitate the design and implementation of applications to provide companionship, dialogic self-paced learning, entertainment and gaming, and help package information and manage interactions through natural spoken dialogue. This knowledge could also aid in machine understanding of dialogue. Dialogue technology has long focused on task-based dialogues – driven by propositional information exchange where success can be measured

Emer Gilmartin
Speech Communication Laboratory, Trinity College Dublin e-mail: gilmare@tcd.ie

Benjamin R. Cowan
Speech Communication Laboratory, Trinity College Dublin e-mail: benjamin.cowan@ucd.ie

Carl Vogel
School of Computer Science and Statistics, Trinity College Dublin e-mail: vogel@tcd.ie

Nick Campbell
Speech Communication Laboratory, Trinity College Dublin e-mail: nick@tcd.ie

by efficient arrival at a clearly defined short term goal. Casual social conversation presents a problem where dialogue success does not primarily depend on the acquisition of information by one or more participants, but also on 'buy-in' or engagement in the activity of talking itself, in addition to the construction and maintenance of a social bond. In recent years, there has been significant progress in the creation of chat applications, particularly in modelling smalltalk – short casual interactions, often in the form of 'getting to know you' dialogue activities. These efforts have been supported by the creation and analysis of corpora of relatively short and often dyadic first encounter dialogues between human participants and in Wizard of Oz scenarios. Building applications which can form a longer term social bond with a user or engage with a group of users calls for knowledge of how longer conversations work. In this paper we describe preliminary explorations of the structure of long (c. one hour) multiparty casual conversations, focussing on a binary distinction between two types of interaction phases – chat and chunk. We provide a brief review of relevant existing work, describe a collection of long form conversations which provide the data for our explorations, outline some early results that may prove useful in the design of such conversations, and finally discuss our future work.

2 The Shape of Conversation - Phases, Chat, and Chunks

Talk is ubiquitous in human life. While some spoken interaction is the medium for performance of practical or instrumental tasks such as service encounters (shops, doctor's appointments), information transfer (lectures), or planning and execution of business (meetings), much daily talk serves to build and maintain social bonds, ranging from short 'bus-stop' conversations between strangers to longer sessions where friends spend time 'hanging out' engaged in what Schelgoff described as 'a continuing state of incipient talk' [13]. In these interactions, there is no clear short-term practical task or prescribed subject of discussion. Speakers are thought to have equal rights to contribute to the talk [18] or at least not to be subject to the clearly predefined roles such as 'teacher-student' which are part of task-based or instrumental encounters [4]. The form of such talk is also different to that of task-based exchanges - there is less reliance on question-answer sequences and more on commentary [16, 18]. Instead of asking each other for information, participants seem to collaborate to fill the floor and avoid uncomfortable silence. As a simple example, a meeting has an agenda and it would be perfectly normal for the chairperson to impose the next topic for discussion. In casual conversation there is no chairperson and topics are often introduced by means of a statement or comment by a participant which may or may not be taken up by other participants.

Several researchers have noted that casual conversations develop as a sequence of phases; after initial chat where there are frequent back and forth contributions among the various participants, the structure of the talk moves to a

series of longer stretches or chunks dominated by one participant at a time, interwoven with more chatty phases.

Ventola [17] described how a non-transactional conversation may comprise several structural elements or phases, which followed one another like beads on a string, sometimes repeating. The structural elements she described are:

G	Greeting.
Ad	Address. Defines addressee (“Hello, Mary”, “Excuse me, sir”)
Id	Identification (of self) - only for strangers.
Ap	Approach. Basically smalltalk. Can be direct (ApD) – asking about interactants themselves (so usually people who already know one another), or indirect (ApI) – talking about immediate situation (weather, surroundings, so can happen between strangers or with greater social distance). In Ventola’s view, these stages allow participants to get enough knowledge about each other to enter more meaningful conversation.
C	Centring. Here participants become fully involved in a conversation, talking at length. This stage is much less predictable than the Approach stage in terms of topic, and can range over several overlapping topics for an indeterminate number of repetitions, often interspersed with further Approach phases.
Lt	Leave-taking. Signalling desire or need to end conversation.
Gb	Goodbye. Can be short or extended, in which case there are projections to further meetings.

Ventola develops a number of sequences of these elements for conversations involving different levels of social distance. She describes conversations as minimal or non-minimal, where a minimal conversation is essentially phatic, particularly in Jakobsen’s sense of maintaining channels of communication [10], or Schneider’s [14] notion of defensive smalltalk - such a conversation could simply be a greeting, or could be a chatty sequence of approach stages. Non-minimal conversations involve centring – where the focus shifts to longer bouts often fixed on a particular topic. Several of the elements are optional and omitted in particular situations. For example, friends can jump from Greeting to Centring without passing through the ‘smalltalk’ exploratory Approach stages. Strangers may not greet one another but could start with ApI (“It’s a nice day”). Many elements often only once, such as greetings (G) and goodbyes (Gb), but others can recur. Approach stages can occur recursively generating long chats without getting any deeper into centring. Centring stages can recur and are often interspersed with Approach stages in longer talks.

Another view of the structure of causal conversation has been developed by Slade and Eggins, who regard casual talk as sequences of ‘chat’ and ‘chunk’ elements [7]. Chat segments are highly interactive and appear to be managed locally, unfolding move by move or turn by turn, and are thus amenable to Conversation Analysis style study. Chunks are segments where (i) ‘one speaker takes the floor and is allowed to dominate the conversation for an extended period’, and (ii) the chunk appears to move through predictable stages, amenable to genre analysis.

In a study of three hours of conversational data collected during coffee breaks in three different workplaces involving all-male, all-female, and mixed groups, Slade found that around fifty percent of all talk could be classified as chat, while the rest comprised longer form chunks from the following genres: storytelling, observation/comment, opinion, gossip, joke-telling and ridicule.

Ventola's phases and Slade and Eggin's binary distinctions (and more detailed generic classification of chunks) could greatly aid the segmentation of conversations into phases or subroutines, which could be used in the design and management of artificial dialogues.

There has been work in the on the theory and analysis of aspects of social conversation, often covering particular phases, such as the long tradition of studies of the structure of narrative, recently applied to the Switchboard spoken corpus [5], or to the patterning of speaker turns in narratives from spoken dialogues in the British National Corpus [12]. There has also been work on creating smalltalk dialogues as part of more task-based talk [1], or alone [20], and on understanding the development of relationships between interlocutors [15, 6]. Much of this work has focused on dyadic exchanges. In the preliminary explorations described below, we focus on multiparty conversation, and take a broad view of chat and chunks rather than drilling down to specific genres within chunks.

3 Data and Binary Annotation of Chat and Chunk Phases

To aid our understanding of the conversational phases in our data, we annotated six long form conversations for chat and chunk. We then extracted descriptive statistics which we report below. We have also marked up the conversations using Ventola's phases which will form the basis for further studies.

The six conversations were drawn from three multimodal corpora of multiparty casual talk - d64, DANS, and TableTalk. In all three corpora there were no instructions to participants about what to talk about and care was taken to ensure that all participants understood that they were free to talk or not as the mood took them. The d64 corpus is a multimodal corpus of informal conversational English recorded in Dublin in 2009 in an apartment living room with 2 - 5 people on camera at all times [11]. The DANS corpus contains conversations ranging between 60-90 minutes with 2 to 4 participants in a living-room setup in the Speech Communication Lab in Dublin in 2012 [9]. The TableTalk corpus was recorded at the Advanced Telecommunications Research Institute (ATR) in Kyoto in 2007, and consists of 3 sessions of 4 or 5-party casual conversations of around 90 minutes in duration [3]. Details of the conversations are shown in Table 3

Frequent overlap and bleedover from other speakers in the audio recordings made them unsuitable for automatic segmentation, so the conversations were manually segmented into speech (including laughter) and silence using Praat [2] and Elan [19]. The segmentation, transcription and annotation of the data are more fully described in [8]

Conversation	Corpus	Participants	Gender	Duration (s)
A	D64	5	2F/3M	4164
B	DANS	3	1F/2M	4672
C	DANS	4	1F/3M	4378
D	DANS	3	2F/1M	3004
E	TableTalk	4	2F/2M	2072
F	TableTalk	5	3F/2M	4740

Table 1 Dataset for experiments

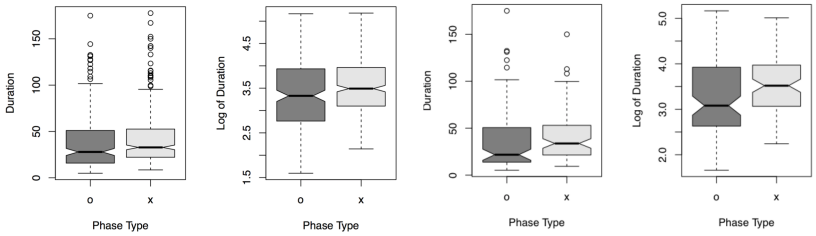

Fig. 1 Boxplots of distributions of duration and log durations of entire dataset (left) and balanced sample (right)

For an initial classification, conversations were segmented into phases by first identifying all of the 'chunks' using the first, structural part of Slade Eggins' definition - 'a segment where one speaker takes the floor and is allowed to dominate the conversation for an extended period' [7]. All other interaction was considered chat.

4 Experiments

The annotations resulted in 213 chat segments and 358 chunk segments overall. Preliminary inspection of the data showed that the distributions were unimodal but heavily right skewed, with a hard left boundary at 0. Log durations were closer to normal with skew reduced from 1.621 to 0.004 for chat and from 1.935 to .0.237 for chunks. These values are below the generally accepted 0.5 threshold for near normality. It was decided to use geometric means to describe central tendencies in the data, after removing one outlying value in the log durations(> 1.5 times IQR).

However, it should be noted that several conversations shared speakers and the number of chunk segments produced by speakers varied widely (8:68), and thus we decided to create a balanced sample to minimize bias. We took the entire sample of the speaker with the fewest chunks (8) and created random samples (n=8) of chunks for each of the other speakers. This resulted in a sample of 96

chunks. We also extracted a random sample of 96 chat segments from the data for comparison purposes. The log transform here reduced skew from 1.772 to 0.236 for chat and from 1.523 to 0.015 for chunks.

Figure 1 shows the boxplots of raw and log durations for chat and chunk segments in both the full dataset and the balanced sample.

The antilogs of geometric means for duration of chat and chunk phases in the original dataset were 28.1 seconds for chat and 34 seconds for chunks, while in the balanced sample the chat value was 25.2 and the chunk value was 33.2. These values were close to the median in all cases, in contrast to the elevated mean values seen for the untransformed data.

Mann-Whitney U/Wilcoxon Rank Su tests on both the full data set and the balanced subset showed significant differences in the distributions of the untransformed durations for chat and chunk ($p < 0.01$ for the full set, $p < 0.05$ for the smaller balanced sample). A Welch Two Sample t-test also showed significant difference in log duration distributions for both datasets.

5 Conclusions and future work

The result of interest in our preliminary explorations is that there is a difference in the distributions of chat and chunk durations – chat varies more while chunks have a stronger central tendency. This could indicate that there is a natural limit for the time one speaker should dominate a conversation and this knowledge could be used in system design. The mean duration of chunk phases was consistent between the full dataset and the balanced sample, which may indicate that chunk duration is not speaker dependent. The larger number of chunk phases in the data compared to Slade's findings on work break conversations may be due to the length of the conversations examined here - we found several instances of sequential chunks where the long turn passed directly to another speaker without intervening chat. Systems which understand and/or generate social human-machine interaction need ground truths based on relevant data in order to create accurate models. We hope that our further explorations into the architecture of longer form conversation will add to this body of knowledge. In addition to studying multiparty data, we intend to investigate dyadic conversations. Unfortunately, there is a dearth of long form conversational data avaliable for analysis. Hopefully, the proven value of past task-based data corpora and the growing importance of social human-machine spoken dialogue will encourage the collection of larger datasets of casual or social conversation open to the research community.

Acknowledgements ?This work is supported by the European Coordinated Research on Long-term Challenges in Information and Communication Sciences and Technologies ERA-NET (CHISTERA) JOKER project, JOKe and Empathy of a Robot/ECA: Towards social and affective relations with a robot, and by the Speech Communication Lab, Trinity College Dublin.?

References

[1] Bickmore T, Cassell J (2005) Social Dialongue with Embodied Conversational Agents. Advances in natural multimodal dialogue systems pp 23–54

[2] Boersma P, Weenink D (2010) Praat: doing phonetics by computer [Computer program], Version 5.1. 44

[3] Campbell N (2008) Multimodal processing of discourse information; the effect of synchrony. In: Universal Communication, 2008. ISUC'08. Second International Symposium on, pp 12–15

[4] Cheepen C (1988) The predictability of informal conversation. Pinter London

[5] Collins KJ, Traum D (2016) Towards a multi-dimensional taxonomy of stories in dialogue. In: Chair) NCC, Choukri K, Declerck T, Goggi S, Grobelnik M, Maegaard B, Mariani J, Mazo H, Moreno A, Odijk J, Piperidis S (eds) Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), European Language Resources Association (ELRA), Paris, France

[6] Devillers L, Rosset S, Duplessis GD, Sehili MA, Bechade L, Delaborde A, Gossart C, Letard V, Yang F, Yemez Y, others (????)

[7] Eggins S, Slade D (2004) Analysing casual conversation. Equinox Publishing Ltd.

[8] Gilmartin E, Campbell N (2016) Capturing Chat: Annotation and Tools for Multiparty Casual Conversation. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)

[9] Hennig S, Chellali R, Campbell N (2014) The D-ANS corpus: the Dublin-Autonomous Nervous System corpus of biosignal and multimodal recordings of conversational speech. Reykjavik, Iceland

[10] Jakobson R (1960) Closing statement: Linguistics and poetics. Style in language 350:377

[11] Oertel C, Cummins F, Edlund J, Wagner P, Campbell N (2010) D64: A corpus of richly recorded conversational interaction. Journal on Multimodal User Interfaces pp 1–10

[12] RÃŒhlemann C, Gries S (2015) Turn order and turn distribution in multiparty storytelling. Journal of Pragmatics 87

[13] Schegloff E, Sacks H (1973) Opening up closings. Semiotica 8(4):289–327

[14] Schneider KP (1988) Small talk: Analysing phatic discourse, vol 1. Hitzeroth Marburg

[15] Schulman D, Bickmore T (2010) Modeling behavioral manifestations of coordination and rapport over multiple conversations. In: Intelligent Virtual Agents, pp 132–138

[16] Thornbury S, Slade D (2006) Conversation: From description to pedagogy. Cambridge University Press

[17] Ventola E (1979) The structure of casual conversation in English. Journal of Pragmatics 3(3):267–298

[18] Wilson J (1989) On the boundaries of conversation, vol 10. Pergamon

[19] Wittenburg P, Brugman H, Russel A, Klassmann A, Sloetjes H (2006) Elan: a professional framework for multimodality research. In: Proceedings of LREC, vol 2006

[20] Yu Z, Xu Z, Black AW, Rudnicky A (2016) Strategy and policy learning for non-task-oriented conversational systems. In: Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue