Stitching together the conversation - considerations in the design of extended social talk

Emer Gilmartin, Brendan Spillane, Christian Saam, Carl Vogel, Nick Campbell, Vincent Wade

Abstract Spoken interaction mediates much human social and practical activity. Talk is not monolithic in form but rather weaves in and out of different genres. Practical tasks are peppered with lubricating social talk, while casual conversation proceeds in phases of interactive chat and longer almost monologue chunks. There is increasing interest in building applications which enable convincing human-machine spoken or text interactions, not only to facilitate immediate practical tasks but also to build a longer term relationships within which conversation can take place in order to entertain, provide companionship and care, and build a user model which will facilitate future tasks through an 'always on' conversational interface. Such applications will require modelling of the different subgenres of talk, and of how these can be convincingly joined to form a coherent ongoing conversation. In this paper we describe our work towards modelling such talk, focussing on theories of casual talk, insights gleaned from human-human corpora, and implications for dialog system design.

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

Brendan Spillane
ADAPT Centre, Trinity College Dublin e-mail: brendan.spillane@adaptcentre.ie

Christian Saam
ADAPT Centre, Trinity College Dublin e-mail: christrian.saam@adaptcentre.ie

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

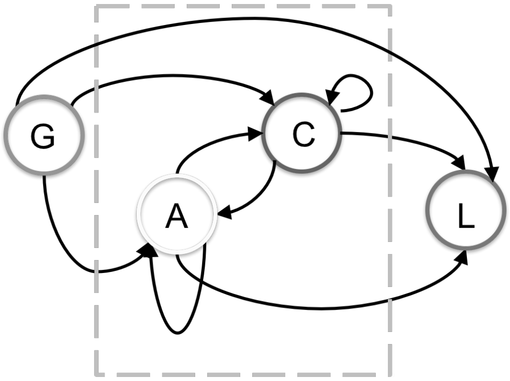
Nick Campbell
Speech Communication Laboratory, Trinity College Dublin e-mail: nick@tcd.ie

Vincent Wade
ADAPT Centre, Trinity College Dublin e-mail: vwade@adaptcentre.ie

1 Introduction

Much dialog system technology has focused on practical (instrumental or task-based) exchanges, and modelling less focussed casual or social talk has been considered less tractable than modelling a task-based exchange [1]. Task-based exchanges can often be formalized and predicted as the progress of dialog is heavily dependent on the exchange of recognisable information (prices, options), the task is known to interlocutors, goals are short-term, and the form of the dialog is largely question-answer. Such dialogs can be modelled as slot-filling paradigms with rules or stochastic methods used to manage the dialog flow. However, much human interaction is partially or wholly composed of social and casual talk, where the ultimate goal seems to be maintaining social bonds, achieved by keeping an entertaining or interesting flow of smalltalk, gossip, conversational narratives or anecdotes, or discussion of topics of mutual interest [15, 16, 9]. It has also been postulated that keeping this channel of dialog open facilitates discussion and performance of practical tasks as and when they arise [3, 12]. There has been much progress in the creation of systems which perform simple tasks or maintain short chats, and now attention is shifting to include longer relational dialogs. Longer casual talk has been shown to occur in phases. Ventola modelled conversation as a sequence of phases including greeting, light conversational approach phases, more detailed centring phases, and forumulaic leavetakings, as shown in Figure 1 [17]. 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 [12], or Schneider's [15] notion of defensive smalltalk - such a conversation could simply be a greeting, a greeting followed by a short approach phase and goodbye, 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. Conversations between people who know each other well (low social distance) can progress directly from greetings to centring, 'dispensing with formalities', while conversations between strangers or more distant acquaintances incorporate more 'smalltalk' in the form of approach phases. Some conversational elements occur only once, as in the case of greetings (G) and goodbyes (Gb), while 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. In their work on casual conversation, Slade and Eggins have observed a structure of alternating phases of often light interactive 'chat', and longer more monologic 'chunks' where one participant tells a story or discusses a topic [4]. Some parallels can be drawn between Ventola's approach phases and Slade and Eggin's chat phases, as some centring phases would match chunk phases, although centring could also include stretches of task based dialog in the case of business encounters wrapped in casual or social talk (as noted by Laver [13]). It is clear that conversation is not monolithic, and may be better modelled as a sequence of phases, where each phase might well involve different dynamics. This has implications for natural language generation, endpointing

Fig. 1 A simplified version of Ventola's conversational phases - Greeting(G) and Leavetaking(L) occur at most once in a conversation, while the Approach(A) and Centring(C) phases may repeat and alternate indefinitely in a longer conversation. Different conversational sequences may be generated from the graph.



and turntaking management, and for the optimization of information flow in a range of applications. Most work on casual conversation to date has been theoretical or qualitative, with little quantitative analysis. The lack of large datasets of substantial casual conversations is an ongoing difficulty in the field, and work has often been limited to the study of corpora of short interactions. We have assembled a collection of casual talk interactions, and are carrying out quantitative analysis on the phases of talk in long multiparty casual conversations (chat and chunk) and shorter dyadic text exchanges (greeting and leavetaking). Below we briefly describe progress made in further understanding casual conversational phases in our recent work, and discuss where this knowledge can be exploited in the design of artificial conversations.

2 Modelling Phases

Our investigations are based on a number of corpora of multiparty casual spoken interaction, d64 [14], DANS [10], and TableTalk [2], and on the ADELE corpus of text conversations [8].

2.1 Greetings and Leavetaking

Our recent work on these phases is based on the ADELE Corpus, a collection of text exchanges where participants were asked to discover biographical information and preferences about their partners through friendly chatty conversation. Using the 'ISO standard 24617-2 Semantic annotation framework, Part 2: Dialogue acts' [11], we annotated 196 conversations from the corpus. From these, we learned that the greeting sequences typically involved 4-6 turns, while leavetaking and goodbyes involved 6-8 turns. We found that the existing labels for greeting and leavetaking insufficient to cover the dialog act patterns we encountered and created a number of new dialog act labels. The sequence of dialog acts comprising greeting and leave-taking phases of dialog were found to vary very little. A full description of the collection, annotation and analysis of the ADELE corpus can be found in [8]. We conclude that modelling such sequences must involve awareness of their multiturn nature and of the acts involved.

2.2 Approach/Chat and Centring/Chunk

To aid our understanding of the structure of the 'meat' of conversation phases, we segmented, transcribed and annotated six long form conversations for chat and chunk phases. The segmentation, transcription and annotation of the data, and the methodology for annotating chunks are more fully described in [5]. The annotations resulted in 213 chat segments and 358 chunk segments overall. Further details of these experiments can be found in [7, 6]. We found a number of differences in chat and chunk phases:

2.2.1 Length of phase:

We have found that the distributions of durations of chat and chunk phases are different, with chat phases durations varying more while chunk durations have a more consistent clustering around the mean (34 seconds). Chat phase durations tend to be shorter than chunk durations. We found no significant differences in chunk duration due to gender, speaker, or conversation.These findings seem to indicate a natural limit for the time one speaker should dominate a conversation. However, we did find that chat phase durations were conversation dependent.

2.2.2 Laughter Distribution in Chat and Chunk phases

We found laughter occurring more frequently in chat phases. Comparing the production by all participants in all conversations, laughter accounts for approximately 9.5% of total duration of speech and laughter production in chat phases and 4.9% in chunk phases.

2.2.3 Overlap and Silence

Overlap is more than twice as common in chat phases (12.6%) as in chunk phases (5.3%), with chunk overlap generally occurring as backchannels, while in chat there is often competition for turns and the talk is more interactive, resulting in longer stretches of overlap. Silence was marginally more common in chat phases. As would be expected, between speaker silences (gaps) predominated in chat, while chunks contained more within speaker silences (pauses) than gaps. The mean lengths of silences were quite similar, but silence duration in chat varied considerably more than in chunk phases.

2.2.4 Chat and Chunk Position

We observed more chat at conversation beginnings, with chat predominating for the first 8-10 minutes of conversations. Although our sample size is small, this observation conforms to descriptions of casual talk in the literature, and reflects the structure of 'first encounter' recordings. However, as the conversation develops, chunks start to occur much more frequently, and the structure is an alternation of single-speaker chunks interleaved with shorter chat segments. Figure 2 shows the probability of a chunk phase being followed by chat or by chunk for the first 30 minutes of conversation. It can be seen that there is a greater tendency for the conversation to go directly from chunk to chunk the longer the conversation continues, resulting in 'story swapping'.

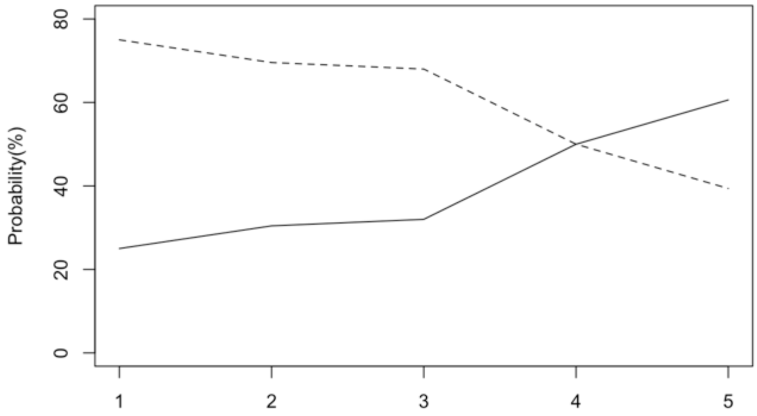


Fig. 2 Probability of chunk-chunk transition (solid) and chunk-chat transition (dotted) as conversation elapses (x-axis = time) for first 30 minutes of conversation data in 6-minute bins

3 Discussion

We have identified several points of interest on phases in talk. The knowledge gained from studying greeting and leavetaking phases can be directly applied to scripting or generating these rather standard sequences. For mid conversation, we have found significant differences in chat and chunk phases. The rather stable duration of chunk phases could be useful in designing the length of stretches where the system shares information with a user – for example for a companion application summarising news articles for an elderly user, or indeed in educational applications. It could also aid by imposing a limit for uninterrupted speech - in prompt generation for system speech and also to inform backchannel models. The distribution of laughter and overlap could inform the behaviour of an agent in different phases of interaction. Differences in overlap and silence distribution could prove very valuable in creating more accurate endpointing and turntaking management depending on the phase of conversation, thus avoiding inappropriate interruptions or uncomfortable silences due to miscalculation by the system - for example, a system taking turns based on an elapsed silence trained on chat could be infelicitous in a chunk phase and vice versa. The positioning of chat and chunk phases also has implications. Chat is more common towards the beginning of phases while the likelihood of chunk to chunk transitions grows over the course of the conversation. This knowledge could help with design of dialog flow, particularly where dialog is to entertain or keep the user company. The modular nature of conversation, with phases differing in their dynamics, makes modelling of the dialog as a monolith rather impractical, particularly in light of the extreme shortage of conversational data beyond collections of task-based interactions and short first encounter casual conversation. However, this very aspect could allow for modelling using diverse corpora for different conversational phases, in a manner similar to unit selection in speech synthesis.

4 Conclusions

Creation of more human-like conversational agents entails understanding of how humans behave in interactions. Casual talk is an essential part of human-human interaction, contributing to bonding, allowing for experiences and opinions to be swapped, and providing entertainment. Casual talk can also act as a matrix for embedded task-oriented talk, keeping conversational channels open in readiness for practical needs. Systems aiming to provide companionship or a persistent presence with a user will need to generate realistic casual conversation. We believe that our ongoing work will help inform the design of such systems - agents which can provide companionship or better serve the user by performing as a human would at the basic dynamics of conversation. We are currently using these insights to build a companion/coaching system for the elderly, ADELE, and exploring different data sources to model different phases. We are also exploring the dynamics of conversational phases in dyadic talk, to see if our multiparty results generalise. The major limitation of the current work is the scarcity of data as corpora of for casual conversations longer than 15 minutes are hard to find and difficult to collect without significant resources. We hope that the current study will encourage the production of corpora of longer form casual conversation, to facilitate the design of convincing artificial interlocutors.

Acknowledgements This work is supported by the European Coordinated Research on Long-term Challenges in Information and Communication Sciences and Technologies ERA-NET (CHIS-TERA) JOKER project, JOKe and Empathy of a Robot/ECA: Towards social and affective relations with a robot, and by Science Foundation Ireland (Grant 13/RC/2106) and the ADAPT Centre (www.adaptcentre.ie) at Trinity College, Dublin.

References

[1] Allen J, Byron D, Dzikovska M, Ferguson G, Galescu L, Stent A (2000) An architecture for a generic dialogue shell. Natural Language Engineering 6(3&4):213–228

[2] 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

[3] Dunbar R (1998) Grooming, gossip, and the evolution of language. Harvard Univ Press

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

[5] 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)

[6] Gilmartin E, Campbell N, Cowan B, Vogel C (2017) Chunks in multiparty conversation-building blocks for extended social talk. Proceedings of IWSDS 2017 pp 6–9

[7] Gilmartin E, Cowan BR, Vogel C, Campbell N (2017) Exploring multiparty casual talk for social human-machine dialogue. In: Karpov A, Potapova R, Mporas I (eds) Speech and Computer, Springer International Publishing, Cham, pp 370–378

[8] Gilmartin E, Spillane B, O'Reilly M, Saam C, Su K, Cowan BR, Levacher K, Devesa AC, Cerrato L, Campbell N, et al (2017) Annotation of greeting, introduction, and leavetaking in dialogues. In: Proceedings of the 13th Joint ISO-ACL Workshop on Interoperable Semantic Annotation (ISA-13)

[9] Hayakawa SI (1990) Language in thought and action. Houghton Mifflin Harcourt

[10] 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

[11] ISO (2012) ISO 24617-2:2012 - Language resource management – Semantic annotation framework (SemAF) – Part 2: Dialogue acts. International Organization for Standardization, Geneva, Switzerland

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

[13] Laver J (1975) Communicative functions of phatic communion pp 215–238

[14] 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

[15] Schneider KP (1988) Small talk, vol 1. Hitzeroth Marburg

[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