

Just Talking - Modelling Casual Conversation

Emer Gilmartin
ADAPT Centre
Trinity College Dublin
gilmare@tcd.ie

Carl Vogel
Computational Linguistics Group
Trinity College Dublin
vogel@tcd.ie

Christian Saam
ADAPT Centre
Trinity College Dublin
saamc@cs.tcd.ie

Nick Campbell
SCL
Trinity College Dublin
nick@tcd.ie

Vincent Wade
ADAPT Centre
Trinity College Dublin
vwade@adaptcentre.ie

Abstract

Casual conversation has become a focus for dialogue applications. Such talk is ubiquitous and its structure differs from that found in the task-based interactions that have been the focus of dialogue system design for many years. It is unlikely that such conversations can be modelled as an extension of task-based talk. We review theories of casual conversation, report on our studies of the structure of casual dialogue, and outline challenges we see for the development of spoken dialog systems capable of carrying on casual friendly conversation in addition to performing well-defined tasks.

1 Introduction

People talk. Human society depends on spoken (or written) interaction. Instrumental or task-based conversation is the medium for practical activities such as service encounters (shops, doctor's appointments), information transfer (lectures), or planning and execution of business (meetings). Much daily talk does not seem to contribute to a clear short-term task, but builds and maintains social bonds, and is described as 'interactional', social, or casual conversation. Casual conversation happens in a wide variety of settings, including 'bus-stop' conversations between strangers, gossip tea break chats between workmates, family and friends 'hanging out' at home or in cafes and bars engaged in Schegloff's 'continuing state of incipient talk' (Schegloff and Sacks, 1973), or indeed in stretches of smalltalk and chat preceding or punctuating business interactions. Much research is focused on dyadic task based dialogue interactions. Early dialogue system researchers recognised the complexity of dealing with social

talk (Allen et al., 2000), and initial prototypes concentrated on practical tasks such as travel bookings or logistics (Walker et al., 2001; Allen et al., 1995). Implementation of artificial task-based dialogues is facilitated by a number of factors. In these tasks, the lexical content of utterances drives successful completion of the task, conversation length is governed by task-completion, and participants are aware of the goals of the interaction. Such dialogues have been modelled as finite state and later slot-based systems, first using hand-written rules and later depending on data-driven stochastic methods to decide the next action. Task-based systems have proven invaluable in many practical domains. However, dialog technology is quickly moving beyond short task-based interactions, and interest is focussing on realistic artificial dialog for roles such as social companions, educators, and helpmates. To model and generate a wider variety of social talk and indeed to improve the quality and user engagement of task-oriented interactions, there is a need for understanding of social conversation. Stochastic models require appropriate data. This paper provides an overview of our recent work in this area, based on corpus studies of casual conversation. Below we describe the concept of social talk and previous work in the area. We then describe our dataset, annotation and the results of our preliminary analyses, discussing how these may aid the design of conversational agents.

2 Casual Conversation

Social talk or casual conversation, 'talk for the sake of talking', or 'phatic communion' has been described as an emergent behaviour whenever humans gather (Malinowski, 1936), and there are theories which posit that such talk is an 'unmarked case' or base form for human spoken interaction

(Dunbar, 1998). Examples of such talk include short conversations when people meet, intermittent talk between workers on topics unrelated to the job in hand throughout the workday, or longer dinner table or pub conversations. Subgenres of casual conversation include smalltalk, gossip, and conversational narrative. The duration of such interactions can vary from short ‘bus stop’ conversations to ongoing interactions which lapse and start again over the course of several hours. Researchers have theorized that such talk functions to build social bonds and avoid unfriendly or threatening silence, as in the phatic component in Jakobson’s model of communication (Jakobson, 1960), distinctions between interactional and instrumental language (Brown and Yule, 1983), and theories that language evolved to maintain social cohesion (Dunbar, 1998). Social talk differs in many ways from task-based conversations. A chat between a concierge of an apartment building and a tenant about football differs in many respects from a customer ordering pizza from an employee. In the chat there is no important information exchanged which is vital to the success of a short-term task, the topic could be the weather or football. In the pizza ordering scenario, information on the type of pizza and the price are vital to a successful transaction, and the goal – sale of a pizza – is short-term, achievable within the conversation, and known to both parties. In the chat, the goal could be described as the maintenance of a social relationship – fulfillment of this goal is a process which extends past the temporal boundaries of the current conversation. Casual conversation seems to be based on avoidance of silence and engagement in unthreatening but entertaining verbal display and interaction, as observed by Schneider (Schneider, 1988), who noted ‘idling’ – sequences of repetitions of agreeing tails such as ‘Yes, of course’ or ‘MmHm’, which seem to keep the conversation going rather than add any new information. He proposed a set of maxims peculiar to this genre, concentrated on the importance of avoiding silence and maintaining politeness. While instrumental talk is often dyadic, casual conversation is very often multiparty. In terms of function, Slade and Eggins view casual conversation as the space in which people form and refine their social reality (Eggins and Slade, 2004) citing gossip between workmates, where participants reaffirm their solidarity, and dinner table talk between friends. In

task-based encounters, participants have clear pre-defined roles (‘customer-salesperson’, ‘teacher-student’) which can strongly influence the timing and content of their contributions to the exchange. However, in casual talk, all participants have equal speaker rights and can contribute at any time (Wilson, 1989) (Cheepen, 1988). 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, storytelling, and discussion (Thornbury and Slade, 2006; Wilson, 1989). Instead of asking each other for information, participants seem to collaborate to fill the floor and avoid uncomfortable silence. Topics are managed locally – a meeting has an agenda and chairperson to impose the next topic, while casual 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. Instrumental and interactional exchanges differ in duration; task-based conversations are bounded by task completion and tend to be short, while casual conversation can go on indefinitely. There are a number of syntactical, lexical, and discourse differences between (casual) conversation and more formal spoken and written genres (Biber et al., 1999). Our work explores the architecture of casual talk.

3 The Architecture of Casual Talk

Casual conversation is not a simple sequence of adjacency pairs, but proceeds in distinct phases. Laver concentrated on the ‘psychologically crucial margins of interaction’, conversational openings and closings in particular, suggesting that small talk performs a transitional function from initial silence through stages of greeting, to the business or ‘meat’ of the interaction, and back to closing sequences and to leave taking (Laver, 1975). Ventola concentrated on longer conversations, identifying distinct phases. Such conversations often begin with ritualised opening greetings, followed by approach segments of light uncontroversial small talk, and in longer conversations leading to more informative centre phases (consisting of sequential but overlapping topics), and then back to ritualised leave-takings (Ventola, 1979). Ventola described several structural elements or phases (listed below), which could be combined to form conversations ranging from minimal exchanges of greetings to long group interactions such as dinner party conversations.

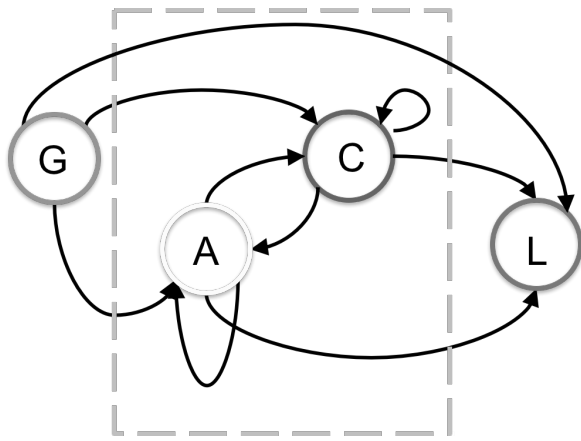


Figure 1: A simplified view of the phases of casual talk described by Ventola - Greeting, Approach, Centre, and Leavetaking.

G	Greeting.
Ad	Address. (“Hello, Mary ”)
Id	Identification (of self)
Ap	Approach. Smalltalk. Direct (ApD) – asking about interactants themselves, or indirect (ApI) – talking about immediate situation (weather, surroundings).
C	Centring. Participants fully involved in conversation, talking at length.
Lt	Leave-taking. Signalling desire or need to end conversation.
Gb	Goodbye. Can be short or extended.

In this model, lighter talk in the form of Approach phases occurs not only at the extremes of conversations, but can recur between Centring phases throughout a longer conversation. Figure 1 shows a simplified schematic of the main phases described by Ventola.

Another model is provided by Slade and Eggins, who contend that casual talk can be seen as sequences of ‘chat’ and ‘chunk’ elements (Eggins and Slade, 2004, p. 230). 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 – that is, it is generic. ‘Chat’ segments, on the other hand, are highly interactive and appear to be managed locally, unfolding move by move or turn by turn. In a study

of three hours of conversational data collected during work coffee breaks, Slade found that around fifty percent of all talk was chat, while the rest comprised longer form chunks from the following genres: storytelling, observation/comment, opinion, gossip, joke-telling and ridicule. In chat phases, several participants contribute utterances with many questions and short comments. Chat is highly interactive with frequent turn changes, and often occurs at the start of an interaction. The conversational floor is shared among the participants and no single participant dominates for extended periods. Chat is often used to ‘break the ice’ among strangers involved in casual talk (Laver, 1975). As the conversation progresses, chat phases are interspersed with chunk phases. The ‘ownership’ of chunks seems to pass around the participants in the talk, with chat linking one chunk to the next (Eggins and Slade, 2004). Figure 2 shows examples drawn from our data of typical chat and chunk phases in a 5-party conversation.

Both Ventola’s and Slade and Eggins’ models treat conversation as composed of phases, with parallels between Ventola’s approach phases and Slade and Eggins’ chat phases. It is likely that the various conversational phases are subject to different norms of turntaking and that phenomena such as laughter or disfluency may appear in different distributions in different phases. Although Ventola’s and Slade and Eggins’ respective work is based on real dialogue in the form of orthographic transcripts, analyses of longer casual talk have been largely theoretical or based on qualitative descriptions. Our work aims to expand our knowledge of the form of these phases so that they can be modelled for artificial dialogue. In our investigations, we first segmented our data into chat and chunk phases to analyse the characteristics of these two types of talk, and in later work plan to refine our analysis by further segmenting our data into Ventola’s phases. Below we outline the limitations of available corpora for work on longer form multiparty casual talk, describe our dataset, annotation, and experiments.

4 Corpora used for Casual Conversation Research

Relevant corpora of human interaction are essential to understanding different genres of spoken dialogue. Dialog corpora have been created of the same spoken task by different subjects, or of inter-

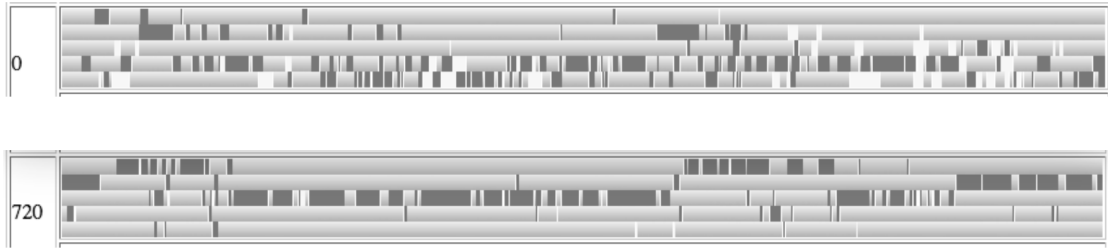


Figure 2: Examples of chat (top) and chunk (bottom) phases in two stretches from a 5-party conversation. Each row denotes the activity of one speaker across 120 seconds. Speech is dark grey, and laughter is white on a light grey background (silence). The chat frame, taken at the beginning of the conversation, can be seen to involve shorter contributions from all participants with frequent laughter. The chunk frame shows longer single speaker stretches.

actions specific to particular domains where lexical content was fundamental to achievement of a practical goal. Such corpora include information gap dialogs such as the HCRC MapTask corpus of dyadic information gap task-based conversations (Anderson et al., 1991) or the LUCID Di-aPix corpus of ‘spot the difference’ games (Baker and Hazan, 2011), as well as real or staged meetings (e.g., ICSI and AMI multiparty meeting corpora (Janin et al., 2003; McCowan et al., 2005)) or genres such as televised political interviews (Beattie, 1983). Because of their task-focused nature, these data, while spontaneous and conversational, cannot be considered true casual talk, and results obtained from their analysis may not generalize to casual conversations.

There are some corpora of casual talk, including telephonic corpora (SWITCHBOARD (Godfrey et al., 1992) and the ESP-C collection of Japanese telephone conversations (Campbell, 2007)), and face-to-face talk datasets (e.g., Santa Barbara Corpus (DuBois et al., 2000), and sections of the ICE corpora (Greenbaum, 1991) and British National Corpus (BNC-Consortium, 2000)). These corpora are audio only and thus cannot be used to inform research on facial expression, gestural or postural research.

Several multimodal corpora of mostly dyadic ‘first encounters’ have appeared recently, where strangers are recorded engaged in casual conversation for periods of 5 to 20 minutes or so (Edlund et al., 2010; Aubrey et al., 2013; Paggio et al., 2010) in several languages including Swedish, Danish, Finnish, and English. These corpora are very valuable for the study of dyadic interaction, particularly at the opening and early stages of in-

teraction. However, the substance of longer casual conversation beyond these first encounters or approach stages has not been focused on in the field.

5 Dataset and Annotation

We compiled a dataset of six informal multiparty conversations, each around an hour long. The requirements for the data were that participants could speak freely, that there was no task or topic imposed by the experimenter, and that recordings were multimodal so that analyses of visual cues could be carried out on the same data and used to build a more comprehensive understanding of multimodal face-to-face interaction. Suitable conversations were drawn from three multimodal corpora, d64, DANS, and TableTalk (Oertel et al., 2010; Hennig et al., 2014; Campbell, 2008). In each of these, participants were recorded in casual conversation in a living room setting or around a table, with no instructions on topic of type of conversation to be carried out - participants were also clearly informed that they could speak or stay silent as the mood took them. Table 1 shows details of participant numbers, gender, and conversation duration for each of the six conversations.

5.1 Data Preparation

The audio recordings included near-field chest or adjacent microphone recordings for each speaker. These were found to be unsuitable for automatic segmentation as there were frequent overlaps and bleedover from other speakers. The audio files were segmented manually into speech and silence intervals using Praat (Boersma and Weenink, 2010). The segmentation was carried out at the intonational phrase level (IP), rather than a more

Table 1: Source corpora and details for the conversations used in dataset

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

coarse and theory dependent utterance or inter-pausal unit (IPU) level. Labels covered speech (SP), silence (SL), coughs (CG), breaths (BR), and laughter (LG). The speech label was applied to verbal and non-verbal vocal sounds (except laughter) to include contributions such as filled pauses, short utterances such as ‘oh’ or ‘mmhmm’, and sighs. Laughter was annotated inline with speech. Annotators worked on 10 second and four-second Praat windows of the audio. Doubtful cases were resolved using Elan (Wittenburg et al., 2006) with the video recordings. Manual segmentation into speech and silence can be problematic, as humans listening to speech can miss or indeed imagine the existence of objectively measured silences of short duration (Martin, 1970), and are known to have difficulty recalling disfluencies from audio they have heard (Deese, 1980). However these results were based on speakers timing pauses with a stopwatch in a single hearing. In the current work, using Praat and Elan, speech could be slowed down and replayed and, by using the four-second window, annotators could see silences or more accurately differences in amplitude on the speech waveform and spectrogram. Although breath is extremely interesting as a feature of conversation (Włodarczak et al., 2015), it was not possible to annotate breath accurately for all participants and thus the breath intervals annotated were converted to silence for the purposes of this study. Similarly, coughs were relabelled as silence for the current work. After segmentation, the data were transcribed, and marked into chat and chunk phases as described below.

5.2 Annotation of Chat and Chunk Phases

Chat and chunk phases were marked using an annotation scheme devised from the definitions of chat and chunk phases given in Slade and Eggins work (Eggins and Slade, 2004; Slade, 2007).

For an initial classification, conversations were divided by first identifying the chunks and considering everything else chat. In the first instance, this was done using the first, structural part of Slade and Eggins’ definition of a chunk as ‘a segment where one speaker takes the floor and is allowed to dominate the conversation for an extended period’ (Eggins and Slade, 2004). The following guidelines were created to aid in the placing of chat/chunk boundaries.

- Start** A chunk starts when a speaker has established himself as leading the chunk.
- Stop** To avoid orphaned sections, a chunk is ended at the moment the next element (chunk or chat) starts.
- Aborted** In cases where a chunk is attempted, but aborted before it is established, this is left as chat. In cases where there is a diversion to another element mid-chunk and a return later, all three elements are annotated as though they were single chunks/stretches of chat.
- Overlap** When a new chunk begins where a previous chunk is still tailing off, the new chunk onset is the marker of interest and the old chunk is finished at the onset of the new one.

Once the chunk was identified, it could be classified by genre. For annotation, a set of codes for the various types of chunk and chat was created. Each code is a hyphen-separated string containing at least a Type signifier for chat or chunk, an Ownership label, and optional sub-elements further classifying the chunks with reference to Slade and Eggins taxonomy. A total of 213 chat and 358 chunk phases were identified across the six conversations.

6 Results

Our analysis of social talk focuses on a number of dimensions; chat and chunk duration, laughter and overlap in chat and chunk phases, distribution of chat and chunk phases across conversations, and turntaking/utterance characteristics.

6.1 Chat and Chunk Duration

Preliminary inspection of chat and chunk duration data showed that the distributions were unimodal

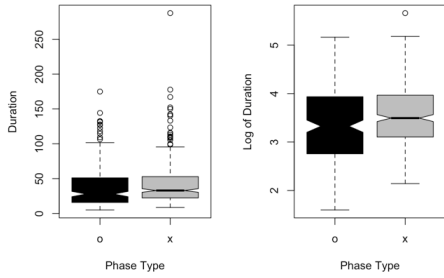


Figure 3: Boxplots of phase duration in Chat (grey) vs Chunk (black) in raw and log transformed data

but heavily right skewed. It was decided to use geometric means to describe central tendencies in the data. The antilogs of geometric means for duration of chat and chunk phases in the dataset were 28.1 seconds for chat and 34 seconds for chunks.

The chat and chunk phase durations (raw and log) are contrasted in the boxplots in Fig 3, where it can be seen that there is considerably more variance in chat durations.

6.2 Speaker, Gender, and Conversation Effects

The raw chunk data were checked for speaker dependency using the Kruskal-Wallis rank sum test, a non-parametric alternative to a one-way analysis of variance (ANOVA), and no significant difference in means due to speaker was found (Kruskal-Wallis chi-squared = 36.467, $df = 24$, p -value = 0.04941). Wilcoxon Rank Sum tests on chunk duration data showed no significant difference between duration distributions for chunks owned by male or female participants ($W = 17495$, p -value = 0.1073). Kruskal-Wallis rank sum tests on chunk duration showed no significant difference between duration distributions for chunks from different conversations (Kruskal-Wallis chi-squared = 9.2077, $df = 5$, p -value = 0.1011). However, the Kruskal-Wallis rank sum tests applied to chat duration showed significant differences between duration distributions for chats from different conversations (Kruskal-Wallis chi-squared = 15.801, $df = 5$, p -value = 0.007436).

6.3 Laughter Distribution in Chat and Chunk phases

Comparing the production by all participants in all conversations, where a participant may produce either laughter or speech, laughter accounts for ap-

proximately 9.5% of total duration of speech and laughter production in chat phases and 4.9% of total duration of speech and laughter production in chunk phases.

6.4 Chunk owner vs Others in Chunk

In the chunks overall, the dominant speakers or chunk owners produced 81.81% (10753.12s) of total speech and laughter, while non-owners produced 18.19% (2390.7s).

6.5 Overlap

There is considerable overlapping of speech in the corpora. For the purposes of this analysis laughter was treated as silence and overlap considered as overlapping speech only. Table 2 and Fig 4 show the occupancy of the conversational floor for all conversations in chat and chunk phases. The number of speakers ranges from 0 (global silence), 1 (single speaker), 2 (2 speakers in overlap) to 3+ (3 or more speakers in overlap).

No. Speaking	Chat	Chunk
0	25.75	22.14
1	61.58	72.27
2	11.88	5.25
3+	0.73	0.42

Table 2: Floor occupancy (%) in chat and chunk for all conversations

It can be seen that overlap is twice as common in chat as in chunk phases, and that silence is slightly more common in chat phases.

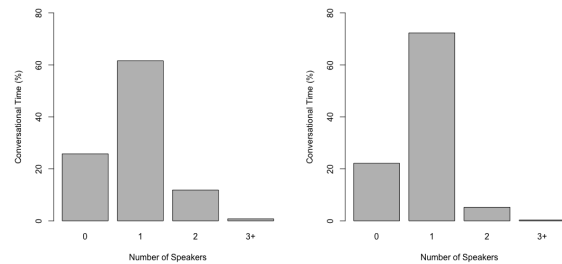


Figure 4: Distribution of the floor in terms of % duration in chat (left) and in chunk (right) phases. X-axis shows number of speakers (0,1,2,3+) speaking concurrently.

6.6 Chat and Chunk Position

Chat predominates for the first 8-10 minutes of conversations. However, as the conversation de-

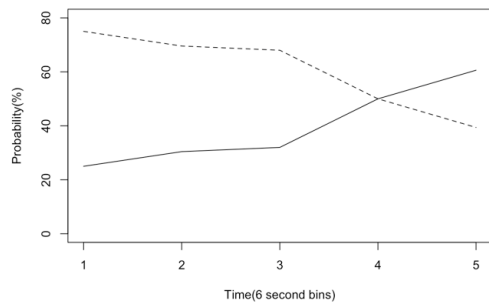


Figure 5: Probability of chunk-chunk transition (solid) and chunk-chat transition (dotted) as conversation elapses (x-axis = time) for the first 30 minutes of conversation

velops, chunks start to occur much more frequently, and the structure is an alternation of single-speaker chunks interleaved with shorter chat segments. Figure 5 shows the probability of a chunk phase being followed by chat or by chunk as conversation continues. 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.

6.7 Utterances and Turntaking

We are studying the patterning of speaker contributions in both phases. Overall we have found that utterances cluster into two groups: short utterances with a mean of around 300ms and longer utterances with mean around 1.4s. In chunk owner speech, utterance mean is higher than utterance means in chat.

We performed a prosodic analysis of phrase final intonation in a subset of the data using the IViE annotation system, finding that falling nuclei ($H^*+L\%$, $!H^*+L\%$) dominated across the data, and particularly in chunks, with relatively few fall-rise tones ($H^*+LH\%$) and small numbers of other tunes.

7 Discussion

We have found differences in the distributions of durations of chat and chunk phases, with chat durations varying more while chunk durations have a more consistent clustering around the mean. Chat phase durations tend to be shorter than chunk durations. These findings are not speaker or gender specific in our preliminary experiments and may indicate a natural limit for the time one speaker should dominate a conversation. The dimensions

of chat and chunk durations observed would indicate that social talk should ‘dose’ or package information to fit chat and chunk segments of roughly these lengths. In particular, the tendency towards chunks of around half a minute could help in the design of narrative or education-delivering speech applications, by allowing designers to partition content optimally. Both laughter and overlap are far more prevalent in chat than in chunk phases, reflecting their light and interactive nature. Interestingly, the rarity of more than two speakers talking concurrently was noted in recent work on turn distribution in multiparty storytelling (Rühlemann and Gries, 2015) – our results would seem to show the same phenomenon in casual conversation, where it much more likely for a speaker to be overlapped by one other speaker than by two or more others. Laughter has previously been shown to appear more often in social talk than in meeting data, and to happen more around topic endings/topic changes [self]. This is consistent our with observations on chat and chunk phases – laughter is more common in chat phases – which provide a ‘buffer’ between single speaker (and topic) chunks.

Chat is more common at the start of multiparty 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. Chunk phases become more prominent later. 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, perhaps reflecting ‘story swapping’ directly without need for chat as the conversation evolves. While the initial extended chat segments can be used to model ‘getting to know you’ sessions, and will therefore be useful for familiarisation with a digital companion, it is clear that we need to model the chunk heavy central segments of conversation if we want to create systems which form a longer-term dialogic relationship with users. As chunks are generic (narrative, gossip..), it may be fruitful to consider modelling extended casual talk as a series of ‘mini-dialogs’ of different types modelled on different corpora – how to convincingly join these sections is an interesting research ques-

tion.

We have noted that many between speaker silences (pauses) during chunk owner speech in chunks are shorter than between speaker silences in chat, probably due to backchannelling in chunks, this would pose a problem for endpointing in dialog systems which relied simply on speaking at a certain delay after detection of silence, as the system would butt in during chat or wait too long during chunks depending on the time delay set. The majority of phrase final intonation curves are the same for chat and chunk reflecting the nature of casual conversation where utterances are predominantly comments or statements rather than question/answer pairs, exacerbating the endpointing/turntaking problem. Knowledge of the type of phase the dialog is in would allow systems to use more nuanced endpointing and turntaking mechanisms. A major limitation of the current work is the scarcity of data. Data for casual conversations which are longer than 15 minutes are hard to find. We hope that the current study will encourage the production of corpora of longer form casual conversation. We are currently extending our explorations to dyadic conversations, and also working on a dialog act annotation scheme for non-task based talk.

8 Conclusions

There is increasing interest in spoken dialogue systems that act naturally and perform functions beyond information search and narrow task-based exchanges. The design of these new systems needs to be informed by relevant data and analysis of human spoken interaction in the domains of interest. Many of the available multiparty data are based on meetings or first encounters. While first encounters are very relevant to the design of human machine first encounters, there is a lack of data on longer human conversations. We hope that the encouraging results of our analysis of casual social talk will help make the case for the creation and analysis of corpora of longer social dialogues. We also hope that our further explorations into the architecture of longer form conversation will add to this body of knowledge.

Acknowledgments

This work is supported by the European Co-ordinated 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 Science Foundation Ireland (Grant 13/RC/2106) and the ADAPT Centre (www.adaptcentre.ie) at Trinity College, Dublin.

References

- J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent. 2000. An architecture for a generic dialogue shell. *Natural Language Engineering*, 6(3&4):213–228.
- James F. Allen, Lenhart K. Schubert, George Ferguson, Peter Heeman, Chung Hee Hwang, Tsuneaki Kato, Marc Light, Nathaniel Martin, Bradford Miller, Massimo Poesio, and David R. Traum. 1995. The trains project: a case study in building a conversational planning agent. *Journal of Experimental & Theoretical Artificial Intelligence*, 7(1):7–48.
- A.H. Anderson, M. Bader, E.G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, et al. 1991. The HCRC map task corpus. *Language and Speech*, 34(4):351–366.
- A. J. Aubrey, D. Marshall, P. L. Rosin, J. Vandeventer, D. W. Cunningham, and C. Wallraven. 2013. Cardiff Conversation Database (CCDb): A Database of Natural Dyadic Conversations. In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2013 IEEE, pages 277–282.
- Rachel Baker and Valerie Hazan. 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior research methods*, 43(3):761–770.
- Geoffrey Beattie. 1983. *Talk: An analysis of speech and non-verbal behaviour in conversation*. Open University Press.
- Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan, and Randolph Quirk. 1999. *Longman Grammar of Spoken and Written English*, volume 2. Longman London.
- BNC-Consortium. 2000. British national corpus. URL <http://www.hcu.ox.ac.uk/BNC>.
- Paul Boersma and David Weenink. 2010. *Praat: doing phonetics by computer [Computer program]*, Version 5.1. 44.
- Gillian Brown and George Yule. 1983. *Teaching the Spoken Language*, volume 2. Cambridge University Press.
- N. Campbell. 2008. Multimodal processing of discourse information; the effect of synchrony. In *Universal Communication, 2008. ISUC'08. Second International Symposium on*, pages 12–15.

- Nick Campbell. 2007. Approaches to conversational speech rhythm: Speech activity in two-person telephone dialogues. In *Proc XVIth International Congress of the Phonetic Sciences, Saarbrücken, Germany*, pages 343–348.
- Christine Cheepen. 1988. *The predictability of informal conversation*. Pinter London.
- James Deese. 1980. *Pauses, prosody, and the demands of production in language*. Mouton Publishers.
- John W. DuBois, W. L. Chafe, C. Meyer, and S. A. Thompson. 2000. *Santa Barbara Corpus of Spoken American English. CD-ROM. Philadelphia: Linguistic Data Consortium*.
- R. Dunbar. 1998. *Grooming, gossip, and the evolution of language*. Harvard Univ Press.
- Jens Edlund, Jonas Beskow, Kjell Elenius, Kahl Hellmer, Sofia Strömbergsson, and David House. 2010. Spontal: A Swedish Spontaneous Dialogue Corpus of Audio, Video and Motion Capture. In *LREC*.
- S. Eggins and D. Slade. 2004. *Analysing Casual Conversation*. Equinox Publishing Ltd.
- John J. Godfrey, Edward C. Holliman, and Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, volume 1, pages 517–520.
- Sidney Greenbaum. 1991. ICE: The international corpus of English. *English Today*, 28(7.4):3–7.
- Shannon Hennig, Ryad Chellali, and Nick Campbell. 2014. The D-ANS corpus: the Dublin-Autonomous Nervous System corpus of biosignal and multimodal recordings of conversational speech. Reykjavik, Iceland.
- R. Jakobson. 1960. Linguistics and poetics. In Th. A. Sebeok, editor, *Style in language*, pages 350–377. MA: MIT Press, Cambridge.
- Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, and Andreas Stolcke. 2003. The ICSI meeting corpus. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 1, pages I–364.
- John Laver. 1975. Communicative Functions of Phatic Communion. In Adam Kendon, Richard M. Harris, and Mary R. Key, editors, *Organization of behavior in face-to-face interaction*, pages 215–238. Mouton, Oxford, England.
- Bronislaw Malinowski. 1936. The Problem of Meaning in Primitive Languages. In *The meaning of meaning: a study of the influence of language upon thought and of the science of symbolism*, 4th ed. rev edition, pages 296–336. Kegan Paul, Trench, Trübner, London.
- James G. Martin. 1970. On judging pauses in spontaneous speech. *Journal of Verbal Learning and Verbal Behavior*, 9(1):75–78.
- Iain McCowan, Jean Carletta, W. Kraaij, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, and V. Karaiskos. 2005. The AMI Meeting Corpus. In *Proceedings of the 5th International Conference on Methods and Techniques in Behavioral Research*, volume 88.
- Catharine Oertel, Fred Cummins, Jens Edlund, Petra Wagner, and Nick Campbell. 2010. D64: A corpus of richly recorded conversational interaction. *Journal on Multimodal User Interfaces*, pages 1–10.
- Patrizia Paggio, Jens Allwood, Elisabeth Ahlsén, and Kristiina Jokinen. 2010. The NOMCO multimodal Nordic resource—goals and characteristics.
- Christoph Rühlemann and Stefan Gries. 2015. Turn order and turn distribution in multi-party storytelling. *Journal of Pragmatics*, 87:171–191.
- E.A. Schegloff and H. Sacks. 1973. Opening up closings. *Semiotica*, 8(4):289–327.
- Klaus P. Schneider. 1988. *Small Talk: Analysing Phatic Discourse*, volume 1. Hitzeroth Marburg.
- Diana Slade. 2007. *The texture of casual conversation: A multidimensional interpretation*. Equinox.
- Scott Thornbury and Diana Slade. 2006. *Conversation: From description to pedagogy*. Cambridge University Press.
- Eija Ventola. 1979. The Structure of Casual Conversation in English. *Journal of Pragmatics*, 3(3):267–298.
- Marilyn A. Walker, Rebecca Passonneau, and Julie E. Boland. 2001. Quantitative and qualitative evaluation of darpa communicator spoken dialogue systems. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, ACL '01*, pages 515–522, Stroudsburg, PA, USA. Association for Computational Linguistics.
- John Wilson. 1989. *On the boundaries of conversation*, volume 10. Pergamon.
- Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. Elan: a professional framework for multimodality research. In *Proceedings of LREC*, volume 2006.
- Marcin Włodarczak, Mattias Heldner, and Jens Edlund. 2015. Communicative needs and respiratory constraints. ISCA.