Chats and Chunks: Annotation and Analysis of Multiparty Long Casual Conversations

Emer Gilmartin, Carl Vogel, Nick Campbell
Trinity College Dublin
Ireland
{gilmare, vogel, nick}@tcd.ie

Abstract

Casual talk or social conversation is a fundamental form of spoken interaction. Corpora of casual talk often comprise relatively short dyadic conversations, although research into such talk has found longer multiparty interaction to be very common. This genre of spoken interaction is attracting more interest with attempts to build more friendly and natural spoken dialog systems. To study longer multiparty casual talk, we have assembled a collection of conversations from three existing corpora. We describe the collection, organization, and annotation of structural chat and chunk phases in these conversations. we then review our preliminary results, noting significant differences in the distribution of overlap, laughter and disfluency in chat and chunk phases, and finding that chunk dominates as conversations get longer. We outline our continuing work on gaining greater understanding of this genre of spoken interaction, with implications for the design of spoken dialog systems.

Keywords: casual conversation, multimodal corpus, annotation

1. Introduction

Many of the practicalities of life are managed through spoken interaction. Such interaction is task-based or practical, facilitating activities such as service encounters (buying a pizza), workplace projects (meetings), or education (tutorials, discussions). People also talk for social reasons – such interactional talk or casual conversation happens whenever people congregate, from short greetings at bus stops to interactions such as dinner party conversations lasting several hours. Although casual conversation is ubiquitous, it is not as well studied as practical talk. We investigate the structure of such talk, particularly in multiparty long (c. 1 hour) conversations. Dialog research has advanced greatly through studies of corpora of relevant speech interaction. In recent years, several high quality corpora have been made available and have underpinned a wide range of research. However, there remain very few collections of longer stretches of multiparty casual talk, and although the use of existing corpora of multiparty talk in scenarios such as meetings is convenient, it is unclear that results would generalise to casual talk. We have gathered a set of six 3 to 5 party conversations, drawn from three corpora, which have been manually segmented, transcribed, and annotated structurally into conversational phases - interactive chat phases and chunk phases where one speaker dominates. We have performed several analyses of the structure of these phases in terms of the occurrence of speech, laughter, and silence, and in the distribution of disfluencies and overlap. The motivation for this work is two-fold. First, the scientific goal of greater understanding of this fundamental human behaviour, and second, to better model conversation in human-machine interaction, particularly in dialog systems which engage in social, companionable talk. Below, we briefly review theories of casual conversation, describe the collection and annotation of the data used, and report on our current and future work towards a fuller account of the structure of multiparty casual conversation.

2. Casual Conversation

Studies of casual conversation have focussed on form and content, and on discourse and sociolinguistic functions.

2.1. Descriptions of Social Talk

Casual social conversation is described as 'talking just for the sake of talking'(Eggins and Slade, 2004), and its subgenres include smalltalk, gossip, and conversational narrative. Aimless social talk or 'phatic communion' has been described as an emergent activity of congregating people, and viewed as the most basic use of speech (Malinowski, 1936). Researchers in fields including anthropology, evolutionary psychology, and communication have theorized that such talk functions to build social bonds and avoid unfriendly or threatening silence, rather than simply to exchange information or express thought, as postulated in much linguistic theory. Instances of these views are found 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 through verbal grooming (Dunbar, 1998). It has long been speculated that the prosodic and gestural aspects of social talk carry much of its communicative load, that 'how' things are said is as important as 'what' is said (Abercrombie, 1956; Hayakawa, 1990). 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 examples of conversation between friends at a dinner party where greater intimacy allows differences of opinion. Schneider collected and analysed audio recordings of naturally occurring small talk, concentrating on the linguistic content of entire dialogues (Schneider, 1988), highlighting how the relative paucity of propositional information flow casual talk which did not seem to conform to Gricean ideas of dialogue - in particular, idling sequences of repetitions of agreeing tails such as 'Yes, of course',

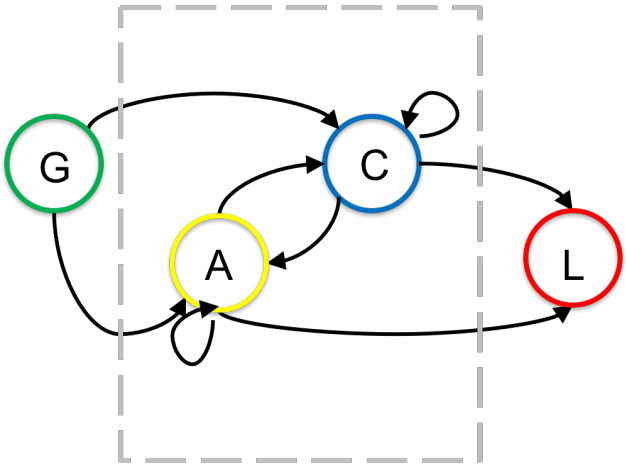

Figure 1: Greeting, Approach, Centring, and Leave-taking phases of casual conversation; a simplified version of Ventola's model.

'MmHmm' which seem to keep the conversation going rather than add any new information. He proposed a set of maxims concentrated on the importance of avoiding silence and maintaining politeness, and suggested that Grice's Co-operative Principle itself remained relevant to small talk although several of the related maxims, particularly those related to quantity and quality, did not apply as strongly as in more practical talk. Syntactical, lexical, and discourse differences between (casual) conversation and more formal spoken and written genres are described in Biber and Leech's work on the Longman Corpus of Spoken and Written English (LSWE), and particularly in their chapter on the grammar of conversation (Biber et al., 1999). Many researchers have also cited a possible text bias in linguistics as a retarding factor on the analysis of spoken interaction (Ong, 1982; Chafe and Danielewicz, 1987; Halliday, 1989).

2.2. Structure of Casual Talk

In early work on social conversation or smalltalk, Laver focussed on the edges of conversation,'psychologically crucial margins of interaction', suggesting that light smalltalk allowed transition from initial greetings to the main business of the interaction and back to closing sequences and to leave taking (Laver, 1975). Laver proposed that such talk suspended power or social differentials between interlocutors, and tended to treat subjects which were uncontroversial. Ventola described casual conversation in terms of distinct phases (Ventola, 1979); often beginning 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 leavetakings, as shown in Figure 1.

Slade and Eggins contend that casual talk can be seen as sequences of 'chat' and 'chunk' elements (Eggins and Slade, 2004, p. 230). Chunks are segments where 'one speaker takes the floor and is allowed to dominate the conversation for an extended period', and the chunk appears to move through predictable stages – that is, it is generic. 'Chat' segments, on the other hand, are described as highly interactive, appearing to be managed locally, unfolding move by move or turn by turn, and thus amenable to Conversation Analysis style study. Figure 2 shows examples of chat and chunk phases taken from the dataset described in this paper.

Text type	Percentage
Storytelling	43.4
Observation/Comment	19.75
Opinion	16.8
Gossip	13.8
Joke-telling	6.3

Table 1: Relative frequencies of chunk phase genres in Slade's workplace conversations

In a study of three hours of conversational data collected during coffee breaks in three different workplaces, 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. Excluding ridicule and chat, which are not amenable to genre analysis, Table 1 shows the relative frequency of the genres encountered by Slade in chunk phases in her conversational data.

Slade and Eggins also report that casual conversation tends to involve multiple participants rather than the dyads normally found in instrumental interactions or examples from conversation analysis. 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. Several researchers on casual conversation have noted that their analyses were limited as they were based on transcripts and thus lacked vital timing and multimodal information.

In the work described below, We focus on the chat and chunk phases in casual conversation. We examine casual conversations (c. 1 hour each in duration) to better understand chat and chunk structure in terms of speech and silence distribution, and the occurrence of laughter and disfluency.

3. Conversational Data

Multimodal corpora of spoken interaction have proven invaluable to researchers in understanding the bundle of signals present in face to face communication. Many multimodal and indeed audio corpora created in laboratory and 'real-world' conditions have been collections of performances of the same spoken task by different subjects, or of interactions specific to particular domains where lexical content was fundamental to progress towards a practical goal - such corpora include collections of information gap activities such as the HCRC MapTask corpus of dyadic information gap task-based conversations (Anderson et al., 1991). Other corpora have focussed on collecting recordings of real or staged meetings, such as the ICSI and AMI multiparty meeting corpora (Janin et al., 2003; McCowan et al., 2005), or recordings of particular genres of interaction, such as televised political interviews (Beattie, 1983).

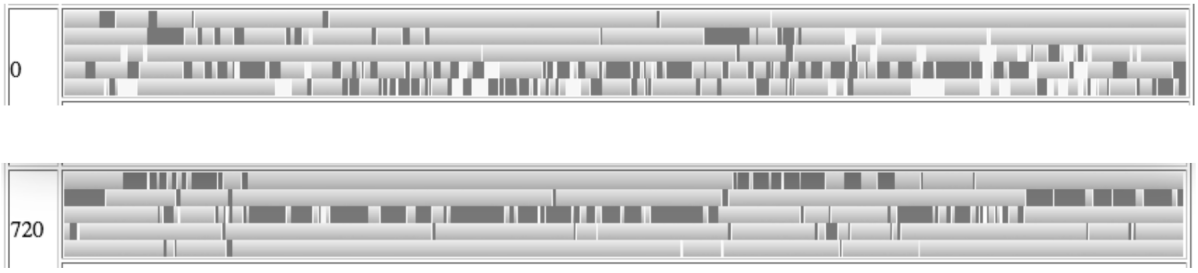

Figure 2: Examples of chat (top) and chunk (bottom) phases in two stretches from a 5-party conversation in the dataset. Each row denotes the activity of one speaker across 120 seconds. Speech is dark grey, and laughter is white on a light grey background (silence).The chat frame, taken at the beginning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

All of these corpora have contributed greatly to research into areas of spoken dialogue such as timing, turntaking, and dialogue architecture. However, the speech in these resources, while spontaneous and conversational, cannot be considered casual talk, and the results obtained from their analysis may not transfer to casual conversation.

Collections of non-task interaction include audio collections of casual talk, often telephonic, such as SWITCHBOARD (Godfrey et al., 1992) and the ESP-C collection of Japanese telephone conversations (Campbell, 2007) There are also collections of face-to-face talk as in the Santa Barbara Corpus (DuBois et al., 2000), and sections of the ICE (Greenbaum, 1991) and British National Corpus (BNC-Consortium, 2000). The Gothenburg Corpus of recordings of different types of human activity contains both audio and video recordings including casual or small talk (Allwood et al., 2000).

There is a growing number of multimodal corpora of mostly dyadic 'first encounters' where strangers were 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). These corpora are very valuable for the study of dyadic interaction, particularly at the opening and early stages of interaction. For a fuller review of available corpora and the challenges of genre in conversation, see (Gilmartin et al., 2015a). However, we are interested in the substance of longer casual conversation beyond these first encounters, and thus we have collected a number of multimodal recordings of conversations of multiparty casual speech to form a dataset for preliminary explorations.

4. Data and Annotation

We assembled a dataset of six informal conversations with three to five participants, each around an hour long. The conversations were drawn from three multimodal corpora, d64, DANS, and TableTalk (Oertel et al., 2010; Hennig et al., 2014; Campbell, 2008). Details of the dataset can be seen in Table 2, and further details of the annotation process can be found in (Gilmartin and Campbell, 2016).

In each of the corpora used, 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. All of the conversations were recorded in audio and video, using chest mounted or adjacent microphones and multiple video angles.

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

Table 2: Source corpora and details for the conversations used in dataset

4.1. Segmentation

The recordings were found to be unsuitable for automatic segmentation using voice activity detection (VAD). While VAD could handle stretches where participants were talking without overlap ('in the clear'), many turn changes involved overlap and there was significant choral production of short utterances and laughter as well as within turn overlap when listeners produced backchannels. Although recordings were made with microphones worn by participants or adjacent to them, there was considerable bleedover between the recordings. After manual synchronisation, the audio files for each speaker were segmented manually into speech and silence intervals using Praat (Boersma and Weenink, 2010) on 10 and 4-second or smaller windows as necessary, and unclear cases were resolved using Elan (Wittenburg et al., 2006) to refer to the video recordings taken at the same time. Human annotators are not completely reliable – in listening tests humans have been shown to interpret the gist of spoken language rather than what exactly they heard, resulting in listeners missing or imagining the existence of short pauses, especially when there is elongation of previous or following syllables (Martin, 1970) or having difficulties recalling disfluencies (Deese, 1980). However, the annotators here had the benefit of visual representations of the waveform and spectogram in Praat, and

thus it is hoped that segmentation is accurate. The segmentation and transcription was carried out at the intonational phrase (IP) level rather than the more commonly used interpausal unit (IPU) as IPs are a basic unit for intonation study and can easily be concatentated to the interpausal unit (IPU) and turn level as required.

4.2. Transcription, Annotation of Disfluency, Prosodic Annotation

After segmentation the data were manually transcribed and non-verbal vocalizations were annotated, using a scheme largely derived from the TRAINS transcription scheme (Heeman and Allen, 1995). Words, hesitations, filled and unfilled pauses, unfinished words, laughs and coughs were transcribed and marked. To facilitate annotation of disfluency, the transcriptions and audio for each IPU were force aligned at the word and phoneme level with the Penn Aligner (Yuan and Liberman, 2008). Sections which could not be automatically aligned, where there was significant overlap or cut off words, were manually aligned.

Symbol	Note
.	interruption point
-	unfinished word
~	unfinished utterance
^	contracted word
r	repeated word
s	substituted word
d	deleted word
f	filled pause
x	pause
o	overlap

Table 3: The annotation code used for basic disfluencies.

The word level transcription was then used with the sound files to annotate disfluencies. The scheme and procedures used were based on those outlined in Shriberg's and Eklund's respective theses (Shriberg, 1994; Eklund, 2004), and Lickley's annotation manual for the MapTask corpus (Lickley, 1998), with extra labels and conventions for recycled turn beginnings (Schegloff, 1987), disfluencies in the presence of overlap, and unfinished utterances. Complex, or nested, disfluencies were labelled following Shriberg's method (Shriberg, 1994). A fuller account of the segmentation, transcription, and disfluency annotation process can be found in (Gilmartin and Campbell, 2016)
We have also manually annotated phrase final pitch movements in a subsection of the corpus in order to investigate how turn-taking cues manifest in chat and chunk segments. Annotation of intonation contours was carried out using the IViE system (Grabe, 2001).

4.3. Chat and Chunk Annotation

All six conversations were segmented into phases by first identifying the 'chunks' using the first, structural part of Slade and Eggins' definition - 'a segment where one speaker takes the floor and is allowed to dominate the conversation for an extended period' (Eggins and Slade, 2004). All other interaction was considered chat.

Code	Details
Type	chat:o, chunk:x
Owner	speaker code/z-everybody
Chunk Genre	story:s, observation/comment:c opinion:o, gossip:g

Table 4: Labelling Scheme for Chat and Chunk Phases

The chunks were first roughly identified from the transcriptions. The type of chunk was then decided with reference to Slade and Eggins taxonomy. Temporal boundaries of each chunk marked off on a 'phases' tier in Praat. Intervals were labelled using the code shown in Table 4, marking type of phase (chat - o or chunk - x), subtype of phase (narrative, story, discussion..), name of phase (roughly equivalent to the topic under discussion), and phase 'owner' (main speaker in chunks and everyone in chat phases). As an example, the code **x_s_g_cats** would denote a chunk phase where the main speaker is **g** and the chunk, which was in story form, was about cats. A total of 213 chat and 358 chunk phases were identified across the six conversations.

5. Overview of Preliminary Results

The dataset has been used to explore chat and chunk phases in conversation, in terms of timing, distribution of speech silence, laughter,overlap, and disfluency. Below we review our preliminary results and implications for dialog system technology.

5.1. Chat and Chunk Description and Distribution

There were a total of 571 segments of chat or chunk in the dataset, comprising 213 chat segments and 358 chunk segments. The number and total durations of chat and chunk segments per conversation can be seen in Table 5.

Conv	Chat No.	Chat Dur.	Chunk No.	Chunk Dur.
A	42 (36%)	1636 (39%)	73	2527
B	38 (32%)	1371 (29%)	82	3300
C	53 (44%)	1363 (31%)	68	3014
D	18 (26%)	660 (22%)	51	2343
E	17 (41%)	909 (44%)	24	1159
F	45 (43%)	2168 (46%)	60	2571

Table 5: Number and duration of chat and chunk segments per conversation, with percentage of conversation in terms of chat phase number, and chat phase duration

It can be seen that in all conversations there were more chunk phases and the time spent overall in chunk phases is greater than that spent in chat phases. Our data differs from Slade's in terms of the proportion of conversational time devoted to chat or chunk. In the dataset examined here, chunk accounts for more conversational time in all conversations than does chat, in contrast to Slade's finding of a 50/50 split.

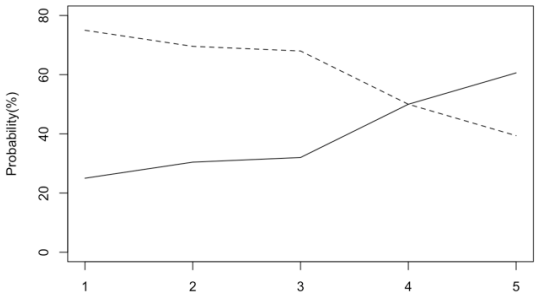

Figure 3: Probability of chunk-chunk transition (solid) and chunk-chat transition (dotted) as conversation elapses (x-axis = time) for first 30 minutes of conversation data in 6-minute bins

We also 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 alternates between single-speaker chunks interleaved with shorter chat segments. In longer conversations, the likelihood of a chunk being followed by chat decreases and chunk to chunk transitions become more common, with the conversation entering a 'swapping stories' phase, as shown in Figure 3. It seems likely that the difference in composition of talk in our data is due to the longer lengths of conversations, and the increased prevalence of chunks and indeed chunk to chunk transition with greater conversational length. From these results it would appear that existing corpora of first encounters or the initial extended chat segments found in longer conversations can be used to model 'getting to know you' interactions or brief casual talk. However, it is clear that we need to model the chunk heavy central segments of longer conversation if we want to create systems which form a longer-term dialogic relationship with users.

From our analysis of the dataset 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. Chat phase durations (Mean=28s) tend to be shorter than chunk durations (Mean=34s). These findings are not speaker specific in our preliminary experiments and seem to 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.

5.2. Speech, Overlap, Laughter and Disfluency Distribution

The commonest conversational state was a single participant speaking 'in the clear' (68%), with global silence accounting for 23% of the conversational time. The remaining time (9%) comprised overlapping speech by two or more participants, with instances dropping sharply as the number of overlapping speakers increases. The vast bulk of overlap in all conversations involved two speakers. There is significantly less overlap and more single party speech in chunk phases than in chat phases. We have also been investigating the frequency and distribution of laughter and disfluencies. Early experiments showed that laughter, and particularly shared laughter, appears more common in social talk than in meeting data, and that laughter happens more around topic endings/topic changes (Gilmartin et al., 2013a; Bonin et al., 2012). This is consistent with our current work on chat and chunk phases, as we are seeing that laughter is significantly more common in chat phases – which provide a 'buffer' between single speaker and topic chunks. In the current dataset we have found that laughter accounts for approximately 10% of vocal time in chat phases while it only accounts for 4% of chunk phases. For disfluencies, a pilot study has shown differences in the occurrence and distribution of disfluency types for chunk owners in chunks and all other speakers (Gilmartin et al., 2015b). In the chunk modality one speaker holds the floor for an extended period and this behaviour is different to that of all other speakers in chunks, to that of all speakers in chat, and indeed to that of the chunk owner when in somebody else's chunk.

6. Current and Future Work

We are studying the patterning of speaker contributions in both chat and chunk phases, particularly the length of gap or overlap in the vicinity of speaker and phase changes. We are performing prosodic analysis of the utterance final pitch movements in different contexts, using the IViE annotations, and believe the results of this work will provide information helpful in developing more finegrained 'endpointing' systems to determine *when* a system should speak; with knowledge of how turntaking occurs in different phases of talk we can work towards providing systems with turntaking behaviour appropriate to the current conversational phase. We are also currently completing dialog act annotation using the ISO 24617-2 standard in order to see how well the ISO standard covers casual conversation, and whether additional acts are necessary to reflect the goals and mechanisms of non-task conversation. In addition, we are further analysing the composition of chunk phases (narrative, gossip, etc) at different stages of conversation, in order to form a clearer picture of how longer conversations develop. This work will aid understanding of the genre, and also have useful applications in the design of artificial dialog.

7. Conclusions

We have described the segmentation, transcription and annotation of a dataset of six long multiparty casual conversations, introducing annotation of chat and chunk phases

within such conversations. The annotations have already enabled fruitful investigations into this omnipresent form of spoken interaction, in areas including laughter, overlap and disfluency (Gilmartin et al., 2013b; Gilmartin et al., 2015b), and the structure and timing of chat and chunk phases. We believe that greater understanding of casual conversation can lead to improvements in timing, dialog management, and natural language understanding in spoken dialog systems. The bulk of the recordings and the resulting annotations are available to interested researchers. Our investigations are preliminary and restricted by the lack of corpora of non task-based conversation, and particularly of long form casual or social talk. This kind of interaction is now of interest in the creation of the next generation of dialog systems - where social talk capacity will be important. We hope that these early explorations strengthen the case for creation of further collections of longer form casual talk, and encourage more investigation of aspects of this genre of spoken interaction.

8. 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 ADAPT Centre for Digital Content Technology, which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

9. Bibliographical References

Abercrombie, D. (1956). *Problems and principles: Studies in the Teaching of English as a Second Language*. Longmans, Green.

Allwood, J., Bjornberg, M., Gronqvist, L., Ahlsen, E., and Ottesjo, C. (2000). The spoken language corpus at the department of linguistics, Goteborg University. In *FQS-Forum Qualitative Social Research*, volume 1.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., et al. (1991). The HCRC map task corpus. *Language and speech*, 34(4):351–366.

Aubrey, A. J., Marshall, D., Rosin, P. L., Vandeventer, J., Cunningham, D. W., and Wallraven, C. (2013). Cardiff Conversation Database (CCDb): A Database of Natural Dyadic Conversations. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 277–282, June.

Beattie, G. (1983). *Talk: An analysis of speech and non-verbal behaviour in conversation*. Open University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., and Quirk, R. (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*.

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

Bonin, F., Campbell, N., and Vogel, C. (2012). Laughter and topic changes: Temporal distribution and information flow. In *Cognitive Infocommunications (CogInfoCom), 2012 IEEE 3rd International Conference on*, pages 53–58.

Brown, G. and Yule, G. (1983). *Teaching the Spoken Language*, volume 2. Cambridge University Press.

Campbell, N. (2007). Approaches to conversational speech rhythm: Speech activity in two-person telephone dialogues. In *Proc XVIth International Congress of the Phonetic Sciences, Saarbrucken, Germany*, pages 343–348.

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

Chafe, W. and Danielewicz, J. (1987). *Properties of spoken and written language.* Academic Press.

Deese, J. (1980). *Pauses, prosody, and the demands of production in language*. Mouton Publishers.

DuBois, J. W., Chafe, W. L., Meyer, C., and Thompson, S. A. (2000). *Santa Barbara Corpus of Spoken American English. CD-ROM. Philadelphia: Linguistic Data Consortium*.

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

Edlund, J., Beskow, J., Elenius, K., Hellmer, K., Strombergsson, S., and House, D. (2010). Spontal: A Swedish Spontaneous Dialogue Corpus of Audio, Video and Motion Capture. In *LREC*.

Eggins, S. and Slade, D. (2004). *Analysing Casual Conversation*. Equinox Publishing Ltd.

Eklund, R. (2004). *Disfluency in Swedish human-human and human-machine travel booking dialogues*. Dept. of Computer and Information Science, Linkoping Studies in Science and Technology, Linkoping, Sweden.

Gilmartin, E. and Campbell, N. (2016). Capturing Chat: Annotation and Tools for Multiparty Casual Conversation. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France. European Language Resources Association (ELRA).

Gilmartin, E., Bonin, F., Vogel, C., and Campbell, N. (2013a). Laugher and Topic Transition in Multiparty Conversation. In *Proceedings of the SIGDIAL 2013 Conference*, pages 304–308, Metz, France, August. Association for Computational Linguistics.

Gilmartin, E., Hennig, S., Chellali, R., and Campbell, N. (2013b). Exploring sounded and silent laughter in multiparty social interaction - audio, video and biometric signals. Valetta, Malta, October.

Gilmartin, E., Bonin, F., Cerrato, L., Vogel, C., and Campbell, N. (2015a). What's the Game and Who's Got the Ball? Genre in Spoken Interaction. In *2015 AAAI Spring Symposium Series*.

Gilmartin, E., Vogel, C., and Campbell, N. (2015b). Disfluency in Multiparty Social Talk. In *Proceedings of DISS 2015*, Edinburgh.

Godfrey, J. J., Holliman, E. C., and McDaniel, J. (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.

Grabe, E. (2001). The IViE labelling guide. *Journal of the Acoustical Society of America*, 101:3728–3740.

Greenbaum, S. (1991). ICE: The international corpus of English. *English Today*, 28(7.4):3–7.

Halliday, M. A. (1989). Spoken and written language.

Hayakawa, S. I. (1990). *Language in thought and action*. Houghton Mifflin Harcourt.

Heeman, P. A. and Allen, J. F. (1995). The TRAINS 93 Dialogues. Technical report, DTIC Document.

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

Jakobson, R. (1960). Linguistics and poetics. In Th. A. Sebeok, editor, *Style in language*, pages 350–377. MA: MIT Press, Cambridge.

Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., and Stolcke, A. (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.

Laver, J. (1975). Communicative Functions of Phatic Communion. In Adam Kendon, et al., editors, *Organization of behavior in face-to-face interaction*, pages 215–238. Mouton, Oxford, England.

Lickley, R. J. (1998). HCRC disfluency coding manual. *Human Communication Research Centre, University of Edinburgh*.

Malinowski, B. (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*, pages 296–336. Kegan Paul, Trench, Trübner, London, 4th ed. rev edition.

Martin, J. G. (1970). On judging pauses in spontaneous speech. *Journal of Verbal Learning and Verbal Behavior*, 9(1):75–78.

McCowan, I., Carletta, J., Kraaij, W., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., Kadlec, J., and Karaiskos, V. (2005). The AMI meeting corpus. In *Proceedings of the 5th International Conference on Methods and Techniques in Behavioral Research*, volume 88.

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

Ong, W. J. (1982). *Orality and literacy: the technologizing of the word*. New accents. Methuen, London.

Paggio, P., Allwood, J., Ahlsen, E., and Jokinen, K. (2010). The NOMCO multimodal Nordic resource - goals and characteristics.

Schegloff, E. A. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organization. *Talk and social organization*, pages 70–85.

Schneider, K. P. (1988). *Small Talk: Analysing Phatic Discourse*, volume 1. Hitzeroth Marburg.

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, University of California.

Ventola, E. (1979). The Structure of Casual Conversation in English. *Journal of Pragmatics*, 3(3):267–298.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). Elan: a professional framework for multimodality research. In *Proceedings of LREC*, volume 2006.

Yuan, J. and Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Journal of the Acoustical Society of America*, 123(5):3878.