Towards A Conversation-Analytic Taxonomy of Speech Overlap

Felix Gervits and Matthias Scheutz
Human-Robot Interaction Laboratory, Tufts University
200 Boston Ave. Medford, MA 02155
{felix.gervits, matthias.scheutz}@tufts.edu

Abstract

We present a taxonomy for classifying speech overlap in natural language dialogue. The scheme classifies overlap on the basis of several features, including onset point, local dialogue history, and management behavior. We describe the various dimensions of this scheme and show how it was applied to a corpus of remote, collaborative dialogue. Moving forward, this will serve as the basis for a computational model of speech overlap, and for use in artificial agents that interact with humans in social settings.

Keywords: overlap, turn-taking, conversation analysis

1. Introduction

1.1. Background

Speech overlap is a common phenomenon found in human dialogue across the world (Schegloff, 2000). Overlap is not the same as interruption, as the former is considered to be a product of turn-taking organization while the latter a violation of conversational norms (Drew, 2009; Drummond, 1989). Much of the past work on speech overlap comes from the field of Conversation Analysis (CA). CA emphasizes talk-in-interaction and aims to study the ways in which social interaction is managed by the participants through dialogue. At the heart of CA is the model proposed by Sacks et al. (1974) (hereafter referred to as SSJ) which elegantly describes the turn-taking organization at the core of human social interaction. The SSJ model makes some important predictions about the structure of turn-taking and how it relates to speech overlap. The most important of these predictions is that while speakers exchange turns in the course of a typical dialogue, they tend to follow the "one-speaker-at-a-time" rule. Another prediction is that speaker changes occur with minimal gap or overlap, and when overlap does occur, it is usually resolved very quickly.

1.2. Speech Overlap

Much of our understanding of the structure of overlap comes from the work of Gail Jefferson (Jefferson, 1982; Jefferson, 1986; Jefferson, 2004). Jefferson showed that overlap is an orderly process characterized by precise onset times with regards to turn-taking structure. She identified several types of overlap based on these onset points (described in Section 3.2.1 below), and found that they generally coincide with the point in a turn at which a speaker change can occur. This suggests that overlap is a consequence of peoples' adherence to the one-speaker rule and the goal to minimize gaps in between turns. On this account, overlap in collaborative dialogue is not seen as rude, but rather supportive. It indicates that conversational partners are receptive to one another and attempt to make smooth and efficient turn transitions - all predictions of the SSJ model.

In addition to characterizing overlap onset, there has also been some work on understanding how people manage and recover from overlap. Jefferson (2004) addresses some aspects of overlap management by describing ways in which people can drop out or hold the turn during overlap, as well as how people deal with segments of speech in overlap that were not heard. In terms of overlap recovery, Schegloff (2000) describes an "overlap resolution device" used by participants in an interaction to recover from overlapping speech. Because of the focus on recovery from overlap, Schegloff limits his analysis to *competitive overlap*, in which there is an explicit claim for the floor that needs to be resolved. As a result, he excludes certain types of *non-competitive overlap* from his analysis, including: terminal overlap, continuers, conditional access to turn (e.g., word search), and "chordal" cases (e.g., laughter). However, these cases are very common dialogue phenomena, and should be included in any thorough account of overlap.

1.3. Present Work

The present work seeks to develop a comprehensive taxonomy of speech overlap that incorporates past work from CA as well as our own contributions. While several aspects of overlap have been studied independently, there does not exist an overall scheme that captures all the critical dimensions for classifying overlap. We seek to develop such a scheme, utilizing methods from CA as well as discourse analysis in order to balance ecological validity with experimental control (De Ruiter and Albert, 2017).

Importantly, the features in our scheme are quantifiable and computationally tractable so as to be useful not only as an explanation of empirical data but also in computational models. This necessitates an important trade off between empirical validity and computational tractability. Computational models and dialogue systems that utilize this framework must operate at the millisecond time scale, so the number of features to consider must be kept to a minimum. As a result, cues from the visual modality (e.g., gaze) are absent from our scheme due to computational complexity, even though they have been identified as important for managing turn-taking in face-to-face interactions.

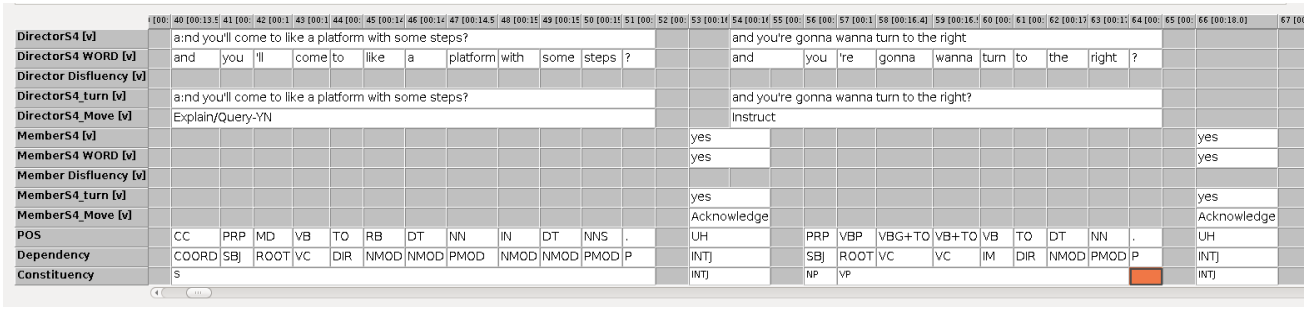

Figure 1: Corpus annotation in the EXMARaLDA Partitur Editor

2. Corpus

We used annotations of the Cooperative Remote Search Task (CReST) corpus (Eberhard et al., 2010) to develop our taxonomy. The corpus contains about 8 minutes of unscripted task-oriented dialogue from each of 10 dyads that performed the task (2712 utterances and 15194 words). The corpus was annotated using the EXMARaLDA Partitur Editor (Schmidt, 2001), and includes the following features: utterances, words, syntactic structure, part of speech, disfluencies, conversational moves, and turns.

The collaborative task at the heart of the corpus (described in more detail in Gervits et al. (2016b) involves two human teammates, a director and searcher, performing a joint search task. The two teammates communicate through remote headset and must achieve a variety of goals within a limited time of 8 minutes.

Speech overlap was relatively frequent in the corpus due to the remote communication, time pressure, and interaction demands imposed by the task. We extracted all instances of overlapping speech in the corpus and classified them according to our scheme.

3. Taxonomy of Speech Overlap

3.1. Definitions

It is important to define a few terms that readers may be unfamiliar with before moving forward. According to the SSJ model the main unit of dialogue is the turn-construction unit, or TCU. A TCU can be a word, clause, phrase, or sentence, and it represents a turn-at-talk. In between (and within) TCUs are points at which speaker change may occur - these are know as transition-relevance places, or TRPs. The TRP signifies a point of completion (grammatical, prosodic, or pragmatic) of the TCU, and is the point at which the next speaker *may* take the turn. When discussing overlap, we use the terms *first starter* and *second starter* to denote the order in which speakers initiated speech.

Another important term to define is the conversational beat. Schegloff (2000) defines a conversational beat as roughly equivalent to the average length of a syllable in spontaneous speech. This corresponds to the average gap time (silence) between speakers' turns, and has been estimated to be between 80-180 ms (Wilson and Wilson, 2005; Wilson and Zimmerman, 1986). Since this varies depending on rate of speech, we use the upper bound of 180 ms as one conversational beat.

3.2. Categories

In order to classify overlap according to our scheme, we define the following categories. These will be discussed below with example dialogues from the corpus.

Onset Point
- Transition-Space, Post-Transition, Interjacent, Last-Item

Local Dialogue History
- *Turn-Holder:*
  - Previous, Current, Next
- *Dialogue Move:*
  - Initiation, Response, Ready

Overlap Management

Non-Competitive
- Drop Turn, Single Item, Wrap Up, Finish Turn, Laughter

Competitive
- *Continue*
- *Disfluency:*
  - Prolongation, Silent Pause, Filled Pause, Combination
- *Self-Repair:*
  - Repetition, Substitution Insertion, Deletion

3.2.1. Onset Point

Perhaps the most important feature to classify overlap is the onset point at which it occurs. Our scheme includes the following types based on Jefferson (1986):

A *last-item* overlap occurs at the point immediately before a TRP (see D1[1]). They typically involve an overlap on the last word, but could also occur at the last lexical item ("cardboard box", "phone number", etc.). Sometimes a person will attempt to come in at the last-item position, but the first starter will continue their turn after the TRP. These are still treated as last-item overlaps since the second starter's entrance was at the perceived last-item position (see D2).

D1) *D: There is . one yellow block . per blue b[ox*
S: [ok]ay

[1]In the dialogue examples, brackets indicate overlap, colons indicate prolonged syllables, hyphens indicate repaired segments, periods indicate brief silent pauses of one beat, and longer pauses are indicated in parentheses.

D2) *S: There is an open do[or to my rig]ht*
D: [per:::fect p]erfect

A *transition-space* overlap (i.e., simultaneous startup) occurs in the transition space between TCUs when the previous speaker continues their prior turn at the same time that the other speaker started their new turn (see D3). The startup can be simultaneous or offset within up to one conversational beat (up to 180 ms). One special case here is when a speaker prolongs the last item of a TCU and the second starter comes in at this point. Instead of being marked as an last-item, this would actually be a transition-space since the speaker was aiming for the TRP (see D4).

D3) *S: Yes*
(0.5)
D: [So is]-
S: A[n d I] just leave that there correct?

D4) *D: Ye : :[: :s*
S: [o k]a y

A *post-transition* overlap occurs when a speaker starts their turn slightly after the current speaker started a new TCU (i.e., after the transition space). "Slightly after" is defined as within 1-2 conversational beats (180-360 ms) of the first starter. It is distinct from the transition-space overlap in that one speaker has already laid claim to the turn. This type of overlap usually occurs when the second starter refers to something that the first starter said in their previous TCU (see D5 - the TRP is between "sure" and "where").

D5) *S: Is there a time limit?*
D: I'm- I'm not sure whe[re are you?]
S: [o k a y]

An *interjacent* overlap occurs in the middle of a turn, not directly near a TRP (see D6). Thus, any overlap that does not fall within the 2-beat window of a transition-space/post-transition or on the last-item of a speaker's turn can be classified as interjacent. While these are closest to "interruption", in practice these types of interjections are usually what are known as *recognitional* overlaps. They occur when speakers seek to correct, clarify, or otherwise respond to something that the first starter said. Continuers or other acknowledgments can also occur at the interjacent point, but it is more common that they occur near TRPs (Duncan, 1972).

D6) *D: Okay maybe that was a-*
(0.5)
D: like they said th[e r e w a s]- [okay
S: [it was a pin]k b[ox

3.2.2. Local Dialogue History

Local dialogue history is a crucial element of our scheme, as people appear to be sensitive to this information when resolving overlap of different types(Schegloff, 2000). For example, if a speaker asks a question and then overlaps the recipient as s/he is responding, then the recipient is likely to drop out. This is because the first speaker violated the adjacency pair, thus creating an implicature that an expansion or clarification of the initial question will occur. Knowledge of the previous turn-holder and dialogue move is necessary to identify this type of behavior. Another common pattern is when a person drops out of competitive overlap only to restart exactly what they were attempting to say at the next available opportunity. Knowledge of the local dialogue history is necessary to classify these examples.

Overlap onset point alone is not sufficient to account for such cases, so our scheme includes information about the previous, current, and next turn-holder, as well as the corresponding dialogue moves with respect to the overall sequence organization. Dialogue move classification is based on the annotation scheme from Carletta et al. (1997), which codes dialogue moves as types of *Initiation, Response*, and *Ready* moves. Expanded Acknowledgment categories are from Eberhard et al. (2010). While the other features are straightforward to code, *current turn holder* requires slightly more consideration in cases of overlap. For interjacent and last-item overlaps, we mark the first starter as the current turn-holder since they already had a turn in progress. For post-transition overlaps, we similarly mark the first starter as current turn-holder because they have made a perceivable sound ($>$ 1 beat) to claim the turn. In transition-space overlaps, however, current turn-holder is set to "both", as both speakers have laid claim to the turn simultaneously.

3.2.3. Overlap Management

Jefferson (2004) describes the following types of general behaviors that can occur to manage overlap, and return the dialogue back to a single speaker: First starter drops - second starter begins; Second starter drops out after false start; Both parties continue simultaneously; both parties drop out simultaneously.

We expand on this preliminary scheme to capture both competitive and non-competitive overlap management behaviors. Though we use the term *competitive* to describe a fight for the turn, it is important to note that such "fights" are very brief and are typically not contentious (barring political debates). They are used as a means to quickly establish who will take the next turn. In our scheme, these behaviors are only considered as overlap management mechanisms if they occurred within two beats of the end of the overlap (following Schegloff (2000)).

The non-competitive categories denote ways in which people come in during overlap with no attempt to take (if second starter) or hold (if first starter) the turn. One such type is *Single Item*, in which the speaker utters a single word (or lexical item) TCU in overlap. Oftentimes these are continuers such as "okay", "right", etc. Another type is *Wrap Up*, which we define as finishing up a turn when overlap is detected. The first starter continues their turn just enough to get to the next TRP, and then allows the second starter to take the floor. This is in contrast to *Finish Turn* which involves completing the remaining item and relinquishing the turn, as in last-item overlap. *Drop Turn* is when a speaker drops out before a TRP, abandoning their utterance. *Laughter* is the final category here. It is a non-competitive activity that typically elicits a similar response from the recipient.

The competitive overlap management categories include several behaviors which participants use to take or maintain the turn during overlap. One such category includes disfluencies such as *prolongations* ($>$ 180 ms/syllable), *silent pauses*, and *non-lexical filled pauses* (uh/um). *Combina-*

Table 1: Frequency of overlap onset points.

Overlap onset	Frequency
Transition-Space	35%
Post-Transition	15%
Interjacent	15%
Last-Item	35%

Table 2: Frequency of overlap management behaviors for asynchronous onset cases (post-transition, interjacent, last-item).

Overlap Management (asynchronous)	First Starter	Second starter
Non-Competitive	**38%**	**26%**
Drop Turn	2.6%	1.5%
Single Item	8.8%	21%
Wrap Up	3.4%	1.6%
Finish Turn	20%	<1%
Laughter	3.4%	1.5%
Competitive	**12%**	**24%**
Continue	6.5%	16%
-Disfluency-	4.3%	4.7%
Prolongation	1.9%	2.6%
Silent Pause	2%	<1%
Filled Pause	<1%	<1%
Combination	<1%	<1%
-Self-Repair-	1.1%	3%
Repetition	<1%	1.8%
Substitution	<1%	<1%
Insertion	0%	<1%
Deletion	0%	<1%

tions of the above behaviors can occur, such as a filled pause followed by a silent pause. The other category includes self-repairs from the HCRC map task coding scheme - repetitions, substitutions, insertions, and deletions (Lickley, 1998). *Repetitions* can be a restart from the beginning of the turn, or can involve repeated syllables or fragments, as in "recycled turn beginnings" (Schegloff, 1987). *Substitutions* occur when a word/item is replaced in the TCU, and *Insertions* occur when a new word/item is added. *Deletions* are abandoned utterances followed by a restart. If a disfluency (pause or prolongation) occurs within a self-repair then the self-repair is given priority for purposes of annotation. This is done to resolve ambiguity in coding overly complex repair combinations which sometimes arise. Finally, the scheme includes a category called *Continue* which indicates that the speaker continued to talk through overlap with no disfluent behavior. They also did not stop at the next TRP, as in the *Wrap Up* case.

4. Corpus Annotation Results

4.1. Summary and Interpretation of Results

As a demonstration of the present scheme, we extracted the above categories from the annotated CReST corpus. While a complete analysis of the corpus is a work in progress, here we report on some observed frequencies in the data.

Table 3: Frequency of overlap management behaviors for synchronous onset cases (transition-space).

Overlap Management (synchronous)	Frequency
Non-Competitive	**55%**
Drop Turn	7%
Single Item	39%
Wrap Up	5.8%
Finish Turn	<1%
Laughter	2.1%
Competitive	**45%**
Continue	24%
-Disfluency-	15.6%
Prolongation	9.6%
Silent Pause	3%
Filled Pause	1.2%
Combination	1.8%
-Self-Repair-	5.8%
Repetition	3%
Substitution	1.2%
Insertion	<1%
Deletion	1.5%

There were a total of 541 overlaps in the 10 teams we analyzed. Table 1 shows the frequency of each type of overlap based on the onset point. Transition-space and last-item overlaps accounted for 70% of all overlap in the corpus. While this may seem surprising given that these overlaps have the smallest window of classification (less than a beat in most cases), this finding highlights the orderly nature of overlap.

We also looked at overlap management behaviors from our scheme. There were 1082 cases here (twice the number of overlaps) because we tracked both speakers' responses. Table 2 shows the distribution of behaviors for post-transition, interjacent and last-item overlaps (total: 741), while Table 3 shows the distribution of transition-space overlaps (total: 341). The data were divided in this way because transition-space overlaps involve a synchronous startup of both speakers, and do not have a first or second starter.

Overall, there was a numerically higher rate of non-competitive overlap in all cases. This supports the SSJ model in that most overlap is not competitive and is resolved quickly. For asynchronous cases, the most frequent behavior for first starters was *Finish Turn* (20%), with most of these coming from last-item onsets. For second starters, the most frequent behavior was *Single Item* (21%). Interjacent overlaps had a high proportion of these, which suggests that they served as verbal acknowledgments, i.e., "continuers". A total of 33% of turns in which a speaker was overlapped at the interjacent point contained a *Single Item* utterance by the second starter. *Continues* were also frequent for second starters (16%), and often occurred at last-item positions. Despite being classified as competitive, when *Continues* occur at the last-item point there is often no competition for the turn; the first starter typically finishes the turn immediately. For the synchronous onset cases, *Single Items* were by far the most common (39%), and often involved both speakers producing them simulta-

neously (e.g., "OK", "OK") in the transition space between turns. *Continues* (24%) and *Disfluencies* (15.6%) were also relatively common, and here they were often used to hold the floor. This suggests that transition-space overlaps may have led to more competition for the floor than the other types. This is not surprising given the fast paced nature of the task and the fact that teammates could not rely on visual cues to predict turn completion.

4.2. Future Work

As we move forward, we will evaluate our taxonomy on additional corpora. The challenge with using traditional data sets is that they often lack the kind of fine-grained turn annotations (e.g., TCU/TRP) necessary to apply our scheme. Corpora that do have turn annotations are typically from natural open-ended interactions (with no particular task), and so may not inform behavior in the kinds of task-oriented settings that are of interest to us. To address these concerns, we will construct a new corpus of task-oriented dialogue with annotations of various turn-taking features.

The long-term goal of this work is to extract the taxonomic features automatically from a dataset. This automated extraction will be a necessary step towards the development of a computational model of speech overlap, as these features will need to be identified in real time. Such a model will be useful both as a theoretical testbed and also in dialogue systems in order to improve communication and make interaction more natural. This latter goal will allow artificial agents to interpret and utilize disfluent segments of dialogue (such as those used to manage overlap) to improve coordination and to serve as better teammates (Gervits et al., 2016a; Gervits, 2017).

5. Conclusion

We have introduced a novel taxonomy of speech overlap that extends prior work in CA and discourse analysis. The scheme classifies overlap on the basis of onset point, local dialogue history, and management behavior. We applied our scheme to a corpus of collaborative, task-oriented dialogue and reported the distribution of the various features of interest. Moving forward, we plan to implement a computational model based on our scheme, with the goal of identifying a minimum set of computationally tractable features that can be used for real-time overlap classification in dialogue systems and artificial agents.

6. Acknowledgements

This work was in part funded by a NASA Space Technology Research Fellowship under award 80NSSC17K0184.

7. Bibliographical References

Carletta, J., Isard, S., Doherty-Sneddon, G., Isard, A., Kowtko, J. C., and Anderson, A. H. (1997). The reliability of a dialogue structure coding scheme. *Computational linguistics*, 23(1):13–31.

De Ruiter, J. and Albert, S. (2017). An appeal for a methodological fusion of conversation analysis and experimental psychology. *Research on Language and Social Interaction*, 50(1):90–107.

Drew, P. (2009). Quit talking while I'm interrupting: a comparison between positions of overlap onset in conversation. In *Talk in Interaction: Comparative Dimensions*, pages 70–93, 01.

Drummond, K. (1989). A backward glance at interruptions. *Western Journal of Communication (includes Communication Reports)*, 53(2):150–166.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, 23(2):283.

Eberhard, K. M., Nicholson, H., Kübler, S., Gundersen, S., and Scheutz, M. (2010). The indiana "cooperative remote search task"(crest) corpus. In *Proceedings of LREC 2010*.

Gervits, F., Eberhard, K., and Scheutz, M. (2016a). Disfluent but effective? a quantitative study of disfluencies and conversational moves in team discourse. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3359–3369.

Gervits, F., Eberhard, K., and Scheutz, M. (2016b). Team communication as a collaborative process. *Frontiers in Robotics and AI*, 3:62.

Gervits, F. (2017). Disfluency handling for robot teammates. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 341–342. ACM.

Jefferson, G. (1982). Two explorations of the organization of overlapping talk in conversation. In *Tilburg Papers in Language and Literature 28*. University of Tilburg.

Jefferson, G. (1986). Notes on 'latency' in overlap onset. *Human Studies*, 9(2-3):153–183.

Jefferson, G. (2004). A sketch of some orderly aspects of overlap in natural conversation. *Pragmatics and Beyond New Series*, 125:43–62.

Lickley, R. J. (1998). HCRC disfluency coding manual. In *Technical Report HCRC/TR-100*. Human Communication Research Centre, University of Edinburgh.

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *language*, pages 696–735.

Schegloff, E. A. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organization. *Talk and social organisation*, 1:70–85.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in society*, 29(1):1–63.

Schmidt, T. (2001). The transcription system exmaralda: An application of the annotation graph formalism as the basis of a database of multilingual spoken discourse. In *Proceedings of the IRCS Workshop On Linguistic Databases*, pages 219–227.

Wilson, M. and Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic bulletin & review*, 12(6):957–968.

Wilson, T. P. and Zimmerman, D. H. (1986). The structure of silence between turns in two-party conversation. *Discourse Processes*, 9(4):375–390.