

To cite, please refer to: Fox Tree, J. E. (2000). Coordinating spontaneous talk. In L. R. Wheeldon (Ed.), *Aspects of Language Production* (pp. 375-406). Philadelphia: Psychology Press.

## Coordinating Spontaneous Talk

Jean E. Fox Tree

University of California, Santa Cruz

### CONTENTS:

Chapter Title: Coordinating Spontaneous Talk

Prelude

Introduction

Grounding in conversation

Coordinating conversational turns

Coordinating ideas

Coordinating repairs

Conclusion

**KEYWORDS:** discourse markers, grounding, grounding criterion, um and uh, turn-taking, overlapped speech, monitoring, pausing, linearization problem, bridging inferences, disfluency, repair (self-repair, other-repair, pragmatic repair), repetitions, replacements, restarts, editing expressions, principle of least collaborative effort, referential card tasks, semantic verification, word monitoring

<ch>Coordinating spontaneous talk

<au>Jean E. Fox Tree

<a>Prelude

Traditional accounts of language production rarely reach the level of describing coordinated talk among people engaged in conversations. Many models explain various aspects of what goes on in an individual's mind as speaking occurs. But what happens when two minds are involved? A distinguishing characteristic of two minds is that talk cannot be planned in advance the way it might be with the production of preset words, phrases, or sentences, and that the talk needs to be coordinated between conversational participants. Both these elements -- spontaneity and coordination -- have been relatively under-researched in comparison to single-speaker, planned production. Correspondingly, in the current chapter I will focus on different issues from other, more often explored aspects of production. I will lay out the problems that conversations present and some of the tools participants have to overcome them. My hope is that the information provided here will aid researchers in expanding current production models to include the elements necessary for real-life, real-time conversations.

<a> Introduction

The way we talk to old friends over a cup of coffee is different from the way we talk in front of the mirror as we rehearse colloquium addresses. Compare the following spontaneously told story with the segment of a political speech after it (adapted from Svartvik & Quirk 1980: 1.3.788 - 1.3.801 and from a Martin Luther King, Jr. speech transcribed by Atkinson 1984: 109; asterisks indicate overlapped speech, periods indicate brief pauses, and dashes indicate longer pauses):

(1) A: and conversation . went like this . this sort of conversation um - - - have you noticed president . that . um - - the boiled eggs at Sunday \*breakfast\* - always hard - -

B: \*( - laughs)\*

A: and president said - ah well - the simple truth is that . if you're going to boil eggs . communally - they must be hard \*( - - - laughs)\*

B: \*( - - - laugh)\*

(2) King: and he's allowed me to go up to the mountain .

Audience: go a\*head\*

King: \*and\* I've looked over and I've seen . the promised \*land\*

Audience: \*holy\* holy \*holy holy\*

King: \*I may\* not get there with you but I want you to know tonight that we as a \*people\*

Audience: \*yeah\*

King: will get to the promised land

The two forms of speech share some important features; in both, speakers plan for particular audiences, decide what to say, and decide how to say it. But spontaneous and prepared talk differ in several key ways. Conversations are not planned in advance, they are not rehearsed, and they always involve the mutual cooperation of two or more people. These factors influence the way we express ideas spontaneously.

One of the most noticeable ways spontaneity affects talk is by causing an increased number of repetitions of words or phrases, ums, restarted ideas, words like you know and well, and long pauses. Though these phenomena are sometimes found in prepared speech, such as

dialogue in novels, they are most common in everyday talk. A quick mental comparison of t.v. commercials provides a ready example of the naturalness of imperfect speech. Those advertisements featuring slick and rehearsed actors are recognised as staged. But those that contain a variety of naturalistic speech phenomena such as disfluencies, wells, and ums create the impression of spontaneous enthusiasm.

The phenomena that are the hallmark of spontaneous talk have often been thought of as unwanted elements of speech, unfortunate byproducts of speaking on the fly. However, another way of viewing these phenomena is as an integral part of the communicative enterprise. In this chapter, I will show how these elements are used to get around some of the problems inherent to the communication medium of spontaneous speech.

In all speaking situations, we have one overarching goal that governs everything that we say: making sure that we have been understood by our listeners. Participants in conversations are engaged in a continual process of achieving mutual understanding, which includes collaborating on what's said and checking that intentions have been understood. In the first part of the chapter, Grounding in conversation, I will discuss how speakers and listeners achieve grounding by monitoring each other's states of understanding. I will also discuss how spontaneous speech differs from other communication mediums because it lacks the ability to be revised in private before production and it lacks the ability to be reviewed as a whole after production. The lack of these features has implications for how grounding proceeds in spontaneous talk.

In the second section, Coordinating conversational turns, I will discuss how conversations are structured, how conversational turns are coordinated, and what happens when turns are not coordinated. The need to carefully time turns has implications for how speech is produced. Speakers do not have the luxury to carefully prepare their utterances before taking a

turn, to privately revise them after they've started speaking, or to review a record of the conversation up to that point.

In the third section, Coordinating ideas, I will discuss devices that developed in spontaneous speech to handle the lack of reviewability. Without reviewability, speakers face a challenge in coherently organizing their hierarchical ideas to conform to the linear nature of speech. Discourse markers are one tool speakers use to help them represent a multilayered discourse. Discourse markers are words such as well, you know, like, oh, and some uses of connectives such as and and but anyway, which are found frequently in spontaneous speech but not in prepared speech. They are a kind of vocal activity that doesn't contribute to the propositional content of utterances but instead relates ideas at the level of organizing talk (Schiffrin 1987). Addressees also contribute to communicative success by building bridging inferences between utterances. I will discuss both the use of discourse markers and the building of bridging inferences, and show how speakers' uses of discourse markers can control the bridging inferences that are built.

In the fourth and final section, Coordinating repairs, I will discuss how the lack of pre-production reviseability is handled in spontaneous speech. Turn-taking constraints and problems in speech production lead all speakers to make speech disfluencies or lexical or pragmatic errors at some time or other. But a number of devices exist in spontaneous speech that help minimise the impact of disfluencies and speed the process of revision. So, though speakers cannot fully revise their talk before they speak, they can and do speak in such a way that their spontaneous revisions lead to the least possible comprehension difficulty for their addressees.

Participants in conversations are engaged in a continual process of achieving mutual understanding, which includes collaborating on what they say, checking that their utterances and

intentions have been understood, and minimizing the impact of errors and revisions. The way we talk in spontaneous conversation helps us fulfill these communicative goals.

#### <a>Grounding in conversations

Spontaneous face-to-face talk is a universal communication medium. Many cultures are not literate and have no access to telephones or more recent communication mediums such as email. But in every culture people use face-to-face conversation as their primary means of communication. Spontaneous talk can come in many forms, including storytelling, where one speaker has the floor and speaks in a monologue, verbal instruction, where a speaker gives directions that are confirmed by addressees, and ritualised teasing, where people take turns ribbing each other. The most common form of talk, however, is conversation. Conversations share certain underlying principles with all forms of communication, but they also have their own particular problems and tools for resolving problems.

No matter how varied the medium or format of communication, all communication has in common the ultimate goal of achieving understanding between communicative participants. In conversations, we check that we have been understood by noting whether our interlocutors nod their heads, say mhm, or reply appropriately to what's been said. Even when the discourse is a one way street, as in monologues or lectures, we still check for understanding. In giving a speech, we monitor understanding by checking for alert faces, head nods, or quizzical expressions. When we can't make out faces, as in a large crowd or on a stage in a darkened room, we monitor audience comprehension by noticing whether people laugh or clap at appropriate moments. If our addressees make no reply to what we've said, or our audience fails to respond, we may be facing communicative failure. Communicative failure in conversations can arise for two reasons: one is that something necessary for communicative success in general

is missing, and the other is that something necessary for communicative success in conversations is missing. I'll first discuss what's necessary for communicative success in general, and then I'll discuss the special case of conversations.

What does it take to be understood? To achieve understanding, speakers cannot just vocalise what is on their minds. Saying I'd like a cup of coffee communicates nothing if the statement isn't heard. Clark and Schaefer identified four states that an utterance must go through before understanding is reached. They are listed below, with S representing the speaker, A the addressee, and u the utterance (adapted from Clark & Schaefer 1987: 22):

State 0. A didn't notice that S uttered any u.

State 1. A noticed that S uttered some u (but wasn't in state 2).

State 2. A correctly heard u (but wasn't in state 3).

State 3. A understood what S meant by u.

In understanding every utterance, addressees need to go through each of these states. They need to notice that they are being spoken to, they need to hear the utterance correctly, and they need to interpret the utterance correctly. Likewise, speakers need to check that addressees are paying attention, that they have heard the utterances, and that the words heard have conveyed a meaning appropriate to the conversational goals. Without feedback from addressees, speakers would not know if the communication succeeded or failed, and ultimately whether or not their conversational goals would be met.

How do speakers gauge addressees' states of understanding? One way they could do this would be to assume that the addressees have attended, heard, and understood unless the addressees say otherwise. That is, speakers could rely on negative evidence of mishearing and misunderstanding, such as an addressee's request for clarification, or an addressee's saying I

didn't hear you (Clark & Brennan 1991). But relying on negative evidence is unlikely to be the norm for two reasons. First, it won't work for assessing whether addressees are in State 1. If addressees don't know that they are being spoken to, they will not be able to respond that they haven't been attending. Second, it won't work in many cases of misunderstanding where the misunderstanding is not recognised by the addressee. It turns out that speakers do not consider negative evidence a sufficient gauge of understanding (Clark & Brennan 1991). Instead, they seek positive evidence of understanding, in the form of positive feedback from addressees that they are attending, hearing, and understanding what's being said.

Let's illustrate the seeking of positive evidence with the shift from State 0 to State 1, or from not noticing that someone has spoken, to noticing that they've said something. How do speakers know that they are being attended to? The most direct positive evidence an interlocutor could supply would be responding to the speaker's utterance with I'm listening, go ahead, a precise description of the addressee's being in State 1. But I'm listening, go ahead leaves the exact state ambiguous; the addressee may or may not be in State 2, because they may or may not have understood. To supply precise information about their state of understanding, addressees would have to say something like What did you say? This question shows not only that addressees are attending and have heard something, but they do not know what it is; that is, they are in State 1 but not State 2.

In addition to direct evidence of attention, interlocutors can also use indirect means to demonstrate that they've heard that something was said. They can turn their head in the direction of the speaker, meet the speaker's gaze, or lay open their hands in a gesture of offering the floor to the speaker. These movements indicate that the interlocutor is listening without directly saying so. Another indirect means that interlocutors can use to demonstrate that they are attending is by



supplying information that they are in a higher state. If they show that they are in State 2, then they must have been in all earlier states. Addressees can nod or utter backchannels like mhm, providing evidence that the speaker has been heard, and also, logically, that the speaker has been attended to. Note that backchannels do not imply that the speaker has been heard correctly; to do that the addressee would have to say I heard you say such and such or to reply in a way that the speaker could judge whether the hearing was correct. Addressees on their own cannot determine their states. It takes both participants in the conversation to assess whether the hearing and understanding were correct.

Direct and indirect evidence can also be used by speakers in checking that their interlocutors have reached the other states. The action of getting up and closing a door in response to the utterance it's cold in here implies that the door-closer has reached each component state leading to understanding: attending, hearing, and understanding. The action, if appropriate to the speaker's intentions, is indirect evidence of the addressee's having reached State 3 in interpreting the speaker's utterance. But the action might also demonstrate that the addressee has not reached State 3: if it's cold in here were meant merely as an opening line to a story about surviving a New England winter (...but it's not as cold as it was back in '76), then the addressee who closed the door would only have achieved State 2.

To achieve understanding, each interlocutor participates in a constant process of checking that the other interlocutor is attending, hearing what's been said, and understanding what's been said. Seeking negative evidence with the passive process of waiting for an addressee's signal of confusion does not do the job. By seeking positive evidence, speakers increase the amount of work they need to do to in the conversation, but they also ensure more accuracy in evaluating comprehension. They have a moment-by-moment picture of exactly where the addressees are in

their understanding process.

But what counts as understanding in the final state? Do addressees need to have a complete picture of exactly what the speaker is thinking, or only of the direction the speaker is headed towards? Clark and Schaefer describe the final state as the point where conversational participants have achieved the grounding criterion: that “the contributor and the partners mutually believe that the partners have understood what the contributor meant, to a criterion sufficient for current purposes” (1989: 262). The necessary level of understanding is that at which the conversational goals can be met. These goals could be any desired outcome of a conversation, such as conveying directions, explaining a decision, or sharing gossip. So it isn't necessary to exhaustively analyze every word a speaker utters, just to understand enough to realise the goals of the conversation.

Both speakers and addressees are responsible for checking that the grounding criterion is reached with every utterance. This mutual responsibility leads to a system of tradeoffs between the amount of effort required to formulate an utterance compared to the amount of effort required to understand the utterance. These tradeoffs have been formalised as the principle of least collaborative effort (Clark & Wilkes-Gibbs 1986), which has been summarised as “[i]n conversation, the participants try to minimise their collaborative effort -- the work that both do from the initiation of each contribution to its mutual acceptance” (Clark & Brennan 1991: 135).

Together, the grounding process and the principle of least collaborative effort lead to certain predictions about language use. One prediction is that as interlocutors become accustomed to each other's way of thinking, they will cut down on the words used to convey an idea. Because they know each other, fewer words are needed to reach criterion. Another prediction is that people will be very good at determining just how much information is needed

to reach the necessary level of understanding. Because the grounding process is necessary for every linguistic interaction, people have a lot of practice at achieving the grounding criterion with diverse conversational partners.

Both phenomena have been observed in the speech produced in referential card tasks. In one version of this task, one person, the director, describes a set of ordered abstract figures to another person, the matcher, who tries to put an identical set of scattered figures in the same order. People cannot see each other, so success or failure of placement can confidently be related to verbal communication. When describing a figure, people start off with long descriptions, but then shorten them upon recurrent references to the figure. What started out as “a person who’s ice skating, except they’re sticking two arms out in front” became on later reference “the person ice skating, with two arms” and then became simply “the ice skater” (Clark & Wilkes-Gibbs 1986: 12). At first, a long explanation was necessary to describe exactly which figure the director meant, but afterwards the same figure could be referred to by a shorter label. When interlocutors have different levels of expertise, they are very good at determining exactly how much information is needed to achieve the grounding criterion. In a referential card task using postcards of New York City, directors adjusted for the expertise of the matchers by the time the first quarter of the cards had been described (Isaacs & Clark 1987). If people did not try to reach a grounding criterion or to minimise their collaborative effort, each referential description would be equally long no matter how much the interlocutors had discussed the figure, and interlocutors would speak in the same way to each other no matter what their expertise.

As with the referential card task, conversational participants monitor each other’s states of understanding to ensure that the grounding criterion is reached with each utterance, while at the same time following the principle of least collaborative effort. A similar process exists for all

other communicative mediums as well, such as telephone conversations, scripted interviews, and email. In all settings, the ultimate goal of grounding is the same: understanding sufficient for current purposes. But how grounding is achieved differs across mediums.

To illustrate these differences, let us consider Clark and Brennan's sample of eight factors that differ across different mediums (adapted from 1991: 141):

- (a) Copresence: A and B share the same physical environment.
- (b) Visibility: A and B are visible to each other.
- (c) Audibility: A and B communicate by speaking.
- (d) Cotemporality: B receives at roughly the same time as A produces.
- (e) Simultaneity: A and B can send and receive at once and simultaneously.
- (f) Sequentiality: A's and B's turns cannot get out of sequence.
- (g) Reviewability: B can review A's messages.
- (h) Reviseability: A can revise messages for B.

As Clark and Brennan discuss, the presence and absence of any of these factors will influence how grounding takes place.

Consider letter writing. The fact that there's no copresence or cotemporality means that letter writers cannot get immediate positive evidence from their addressees that they have been understood. Consequently, they put more time into expressing their ideas carefully, and take care to review and revise their messages before sending them off. More thought goes into each letterwriter's turn than into each turn of a conversational participant. Furthermore, the lack of copresence and cotemporality affects the type of information supplied in a turn. Letter writers might need to remind each other of what was said in prior letters. This isn't necessary in conversations because speaking turns are not separated by days or weeks and the same discourse

record is immediately available to all participants, assuming everyone has roughly the same memory capacity.

The absence of grounding can be thought of as incurring costs for the communicative process, the goal of which is to minimise the costs across communicative partners. For example, the absence of cotemporality can incur the letter writer a cost of not knowing how much is enough when writing down information. But the presence of reviseability allows the writer to make up for this absence by writing more clearly, and correcting what was written until it expresses just what was meant. At the same time, the fact that letter writers have the opportunity to revise means that if writers make errors that do not get corrected before the letters are sent, they will be held more accountable for their mistakes than they would have been for errors in talk. The politician who declares that calling an opponent a drunk was a slip of the tongue would have much more serious problems were the word the result of a slip of the pen. Written errors are harder to correct than spoken errors and consequently repair is more costly in writing.

Spontaneous speech has its own cost tradeoffs. Because spontaneous speech is cotemporal, it doesn't incur the cost of not knowing how much is enough. The interlocutor can readily provide feedback about when to stop talking or when to elaborate. However, cotemporality does incur a cost for delaying. Because the other interlocutor is actively waiting for the speaker to continue, if the speaker delays speech, this has immediate ramifications. This cost doesn't exist with letter writing.

But there is one special kind of writing that does incur a waiting cost: writing with the talk facility on unix. With talk, two people can write to each other on a computer with a split screen where one person's writing appears on the top of the screen and the other's writing appears on the bottom. The writing takes place in real time. Like letter writing, talk writers are

not copresent, not visible to each other, not audible to each other, and they have some amount of reviewability (at least a half-screen full). But like speaking, the talk writers are cotemporal, can send simultaneous messages, take sequential turns, and cannot revise their writing in private. Correspondingly, talk writers incur the costs of cotemporality and have been observed to use the same devices that speakers use in speech (as will be discussed below) to minimise these costs, such as writing um.

In spontaneous speech, interlocutors have the first six of Clark and Brennan's factors, which are exactly the factors letter writers do not have. Interlocutors talk to each other while in the same environment, while visible and audible to each other, and in sequence. Each speaker also knows that they are being heard at the same time as they are speaking, without a time delay between production and receiving of the production, and that they can overlap in speech with their interlocutors. What spontaneous speech does not have is reviewability and (private) reviseability. The effervescence of talk means that there isn't a record available for both people to review after something has been said. Short term memory can only briefly store the most recent speech in both interlocutors' minds. The cotemporality of talk also means that it cannot be revised after it is spoken without all participants being aware of the revision. This awareness makes public revision different from private revision. With public revision, other speakers hear what's being revised and can jump in to help or even change the revision. Private revision, where speakers revise their speech plan mentally before uttering anything, is difficult to accomplish in spontaneous talk because it requires long moments of silence that disrupt the conversation by allowing gaps where interlocutors could take the floor.

This doesn't mean that reviewability and reviseability are impossible in spontaneous speech. We can imagine situations where these factors are present. It's possible to record

conversations and then play them back as the discussion develops, and it's also possible to set an accepted standard where people could be silent for longer periods without risking losing their turns. But both possibilities would entail different grounding procedures. In the first case, speakers might take longer to say something that they knew was being taped. In the second, there would need to be some way other than the existence of a long pause to indicate a turn exchange, such as a system where the person holding a stick gets to talk, or the words I'm done must be said before someone else can take the floor.

In spontaneous speech, the lack of reviewability and reviseability affects grounding in predictable ways. The lack of reviewability means that listeners incur a cost in not being able to check exactly what had been said up to that point in the conversation. This may lead speakers to repeat information, or to argue about who said what in a debate. This wouldn't happen with letter writers; indeed, a letter writer who repeated information several times would seem strange. Likewise, the lack of repetition of instructions in a letter would not be noticed as bizarre, but a single list of verbal instructions without repetition would seem strange, as if the speaker assumed the addressee had a taperecorder for a brain. To balance this cost of not having reviewability, conversations are structured so that participants can keep a good representation of the discourse in memory. This tradeoff in the grounding process helps maintain interlocutors' least collaborative effort in communicating.

The lack of reviseability means that speakers may end up saying something they will later need to adjust. The reasons people speak before they are ready is that it is costly to delay a message in order to revise it. A delay could open the floor for another person to start talking, or could send social signals to the other conversational participants implying a lack of interest in the conversation (Clark & Brennan 1991; Smith & Clark 1993) There might also be a cost to

addressees in waiting. Addressees might be less able to connect ideas to those preceding after a long delay, or their attention may wander in a long break. These waiting costs to speakers and addressees don't exist with other mediums like letter writing, where revision time has no implications for either communicator.

But once again, the costs of not being able to carefully prepare talk has its own benefits in comparison to other mediums. Revising on the fly has fewer ramifications in speech than in other mediums because people expect formulation problems in speech. Readers do not expect unprepared sentences in letters, so a revision is much more difficult to achieve. It is easier to correct a misspoken word than a miswritten one.

The communicative tradeoffs between grounding process factors have been demonstrated in studies of conversations. In interviews, people will pause longer and talk more when the interviewer is absent rather than when the interviewer is present (Siegman & Reynolds 1983). This is considered to be a direct result of the tradeoff between the time pressure posed by a ready and waiting listener and people's ability to organise their thoughts and decide what to say. Without the cotemporal constraint, interviewees can think more about what to say, and review messages for a longer period of time in their heads before talking. The absence of simultaneity had marked effects on the amount of speech people needed to complete a referential card task. When the dyads couldn't communicate in both directions, the amount of speech used to convey the card positions varied greatly (372 words to 1830 words); when they could communicate, speech was much more constrained (641 words to 1280 words; Fox Tree, unpublished research). Like in letter writing, without feedback from matchers, directors could not tell how much information was enough.

Conversational participants are constantly working towards achieving mutual



understanding. They check that their utterances have been attended to, heard, recognised, and understood well enough to carry out the goals of the communication. They also deal with a number of problems that are not present in other forms of communication. In this chapter, I will focus on the lack of reviewability, the lack of reviseability, and the devices that interlocutors use to achieve grounding despite these handicaps. But before I discuss how spontaneous speech is designed to help fulfill the overarching goal of being understood despite the inherent constraints of conversations, I will first discuss the structure of conversations themselves.

#### <a>Coordinating conversational turns

Conversations can be described as an incremental buildup of units called contributions (Clark & Schaefer 1987). Contributions consist of a presentation phase and an acceptance phase. In the presentation phase, speakers present an assertion, request, or question to their interlocutors. In the acceptance phase, the interlocutors either accept the presentation or work to revise the presentation so that it is acceptable. Each phase must achieve the grounding criterion before it is accepted. As described in the previous section, interlocutors check that they have achieved the highest state of understanding by monitoring each other's acceptances.

To make sure that the presentation phase has been accepted and that participants have grounded their contributions requires either nonverbal feedback or taking turns speaking. If people said important things simultaneously, it would be hard to determine what each contributor thought about each other's presentation because their own simultaneous utterances would not necessarily be taken as replies, and as such could not be taken as indirect evidence of acceptance. The grounding process would grind to a halt.

Though it is true that one or more conversational participants can speak at the same time, overlapped speech is usually different in quality from nonoverlapped speech, with overlapped

speech being less informative (Atkinson 1984; Clark 1996; Sacks, Schegloff, & Jefferson 1974). In fact, one person's overlapping another can be purposefully done to demonstrate that additional information is unnecessary (Jefferson 1973). In cases of nonintentional overlap, interlocutors choose to treat overlapped speech either such that one person's contribution were not said, or such that both people's contributions were said sequentially, with the first person to talk after the overlap determining which utterance would be taken as having been said first (Jefferson 1973). When information needs to be conveyed, only one speaker has the floor. The regularity in turns is essential to ensure that each partner in a conversation has the opportunity to both make a presentation and to accept the interlocutor's presentation so that grounding can be achieved.

Turn-taking is a precisely timed activity. People begin to speak just as their interlocutors finish (Sacks et al. 1974). Precision timing is perhaps best demonstrated by the fact that addressees can complete speakers' sentences without any intervening pause (Jefferson 1973). How do people coordinate their utterances so precisely?

One way is by using gaze, facial expressions, and body language to identify turn units. Turning gaze away and putting on a "thinking face" is one way speakers can signal to listeners when to enter a conversation in the situation where speakers need help completing their thoughts (Goodwin & Goodwin 1986: 57). People can also use other nonverbal cues. Addressees can indicate their desire to take the floor by turning their heads away from the speaker or using hand gestures (Duncan & Niederehe 1974). Likewise, speakers can signal the desire to end a turn by stopping gesturing (Duncan 1972). When conversational participants can see either each other's eyes or each other's bodies, they can synchronise their talk better (Argyle, Lalljee, & Cook 1968).

Another way to synchronise turn units is by using special words and phrases to signal

either a desire to gain access to the floor or a willingness to give up a turn (Erman 1987; Erman & Kotsinas 1993; Holmes 1986; Jefferson 1973; Schegloff 1987; Schourup 1985; Schiffrin 1987; Stenström 1990). The evidence for these claims, as well as for the claims in the following sections, comes by and large from corpora analyses. This empirical technique starts with a written-out transcript of spontaneous talk including every vocalization made by the conversational participants. Transcriptions vary in detail, such as how pronunciations and prosody are represented (Edwards 1993). Often with the help of a computer, researchers search their corpora for every example of a particular phenomenon, such as the word oh. They then identify ways the phenomenon is used, sometimes tallying the number of times the use occurred in their data source. Phenomena can be analyzed in a myriad of ways, including where they fall in an utterance or a turn, where they fall syntactically, how they are pronounced, what type of other phenomena precede or follow them, and how they are used pragmatically. Researchers have gained a lot of mileage with this technique, and I will review some of the findings. At the end of the next section, I will also review the few experimental tests that have been done.

To start in on a turn, speakers can use words like well, but, and or so. These words can be used to orient backwards to what was already said in the discourse and can indicate a desire to start in on a turn even before the turn itself has been planned (Clark 1983; example adapted from Svartvik & Quirk 1980: 3.5.914 - 3.5.916):

(3) A: it's necessary for the best kind of tragedy I think,

B: so you think that uhw u:h Romeo and Juliet, is an inferior kind of tragedy

Here speaker B uses so to take over the floor, initiating the turn before the utterance is completely ready, as evidenced by the stumbling before the words Romeo and Juliet. Speakers can also grab the floor with a word like now, which shifts attention to upcoming talk (Schiffrin

1987; example adapted from Svartvik & Quirk 1980: 3.5.999 - 3.5.1003).

(4) A that u:h i it is . related . to the . image . that he's chosen - I think it - - -

B: now, let's go back to Hamlet then

Here speaker B's now helps to both change turns and to shift attention away from what speaker A said and towards what speaker B is going to say.

In order to end a turn, speakers can use tag questions like don't you think so? or didn't it?. These devices indicate that the turn is finished, and orient forwards towards choosing the next speaker (Clark 1983; Sacks et al. 1974; example adapted from Svartvik & Quirk 1980: 1.3.1190 - 1.3.1195):

(5) A: m - - but she . at the same time she seems, . unusual, . doesn't she,

B: . yes, - yes, and everybody notices that she's unusual

Here speaker B pauses for a brief moment before picking up on the cue and taking over the floor. Speakers can also use other phrases like but uh and or something to show that they've completed their turns (Duncan 1972).

Still other words, like then, can be used to hold the floor and prevent interlocutor interruption (Redeker 1991; example adapted from Svartvik & Quirk 1980: 1.3.557 - 1.3.571).

(6) A: I mean I just insisted very firmly, on calling her Miss Tillman, but one should really call her president. - - and . um then, . a bell rang, - - and - millions of feet, . ran, . along corridors, you know, and then they . it all died away, it was like like sound effects from the Goon Show, . you \*know,\* - and then there was a, - tap on the door, -

B: \*m\*

Here the speaker is relating a long story and uses then to maintain her turn while she tells it.

Anyway can also help speakers keep their turns. One researcher found that addressees don't take

over the floor after a speaker says anyway even when anyway is followed by a pause (Bublitz 1988). Using a signal of desire to keep the floor is particularly important in cases where a long pause might run the risk of signalling the end of a turn; to counteract this effect of the pause, the speaker can say then to block the interlocutor's taking the floor.

The more nonverbal and verbal cues for turn exchange, the more likely a successful turn exchange will occur (Duncan 1972). Overlapped speech can be viewed as a result of interlocutor error; either the speakers inadvertently supplied turn-yielding signals, or the addressees took the floor without signals, which is perceived as an interruption (Duncan 1972). Turn taking cues are also important in resolving overlap that occurs when two interlocutors try to take the floor at the same time; whoever has provided the most positive cues for taking the floor will be the one to win the turn (Duncan & Niederehe 1974).

So conversational participants have a lot of tools available to them to estimate what their conversational partners want to do. They can use gaze, gesture, and special words and phrases to carefully interleave their conversational turns with minimal overlap. But why are turns so carefully timed? Consider what happens when they aren't, as in the following two constructed exchanges:

(7) A: Did you have a good time in Boston?

B: It was great.

(8) A: Did you have a good time in Boston?

B: (3 second delay) It was great.

In the second exchange, speaker A will infer that the question posed was not an easy one to answer, that there were some misgivings about the trip to Boston, or that something happened that B did not want to say to A. B would be expected to give some explanation for the long

pause, if only to say I don't want to talk about it. Turns are carefully timed because responses that are too quick or too slow carry weight and have meaning for the interlocutors. A slow response to a question can cause a speaker to be seen as “uncooperative, ignorant, poor in judgment, or slow-witted” (Smith & Clark 1993: 36). Pauses of three or more seconds are particularly undesirable, and there is some evidence that frequently producing such lapses is interpreted as having poor social skills (McLaughlin & Cody 1982).

These judgements are made because at every moment in the conversation, each interlocutor is monitoring what the other is saying (Clark 1996; Clark & Brennan 1991; Smith & Clark 1993). If a pause occurs, it is a contribution to the discourse and will be interpreted. Pauses have been shown to not only be used differently, but to actually be interpreted differently. For example, pauses can reflect the state of understanding that people are in. In answering questions, people pause longer when they think they know the answer but can't think of the word than when they know they don't know the answer (Smith & Clark 1993), and listeners interpret these pauses accordingly (Brennan & Williams 1995). Of course, not all pauses are interpreted in a negative light as a sign of uncooperativeness or of lack of knowledge. If speakers are talking while driving, abrupt stops in conversations might be expected, especially coinciding with difficult maneuvering of the car. Interlocutors take pragmatic circumstances into account when estimating acceptable pause length or inter-turn intervals. But in every situation, there is only so much pausing that is acceptable given the circumstance. When left without a situational excuse in a conversation, such as visible evidence of doing something else that precludes talking, interlocutors come up with other explanations for too-long pauses. But because nonsituational pauses are sometimes necessary to organise ideas or decide what to say next, there are several devices speakers can use to control the interpretations that are made.

To prevent their conversational partners from interpreting the wrong things, people can supply information about what is causing time delays in their responses. Sometimes this information is a direct explanation, such as saying let me get my thoughts straight, or, in answering questions, saying “shoot hang on a minute (1 sec pause) this one has potential” (Smith & Clark 1993: 36). Speakers can also signal the length of their upcoming pause by using either an um or an uh, with ums signalling longer pauses and uhs shorter (Smith & Clark 1993). By using one of these words, speakers can control the interpretations interlocutors make and save face. In answering questions, speakers can punctuate a pause with an um or an uh to show that they are not generally ignorant, they just can’t recall the answer at that time (Smith & Clark 1993). In general, not more than one second goes by before speakers do some activity like guess at a sought after word or utter an um or an uh (Jefferson 1989). There is also some evidence that if a pause does become longer than a second, a correspondingly more noteworthy expression will be used, such as oh gee (Jefferson 1989).

Conversational participants need to coordinate turns in order to achieve the grounding criterion. Conversations are structured as a series of presentation and acceptance phases which together make up a conversational unit. Interlocutors take turns presenting and accepting information in their accumulating discourse representation. The timing of turns has implications for how talk is interpreted. By saying well, but, didn’t it?, um, or oh gee speakers can help to coordinate turns and they can provide information about how pauses should be interpreted. We will now turn to how interlocutors handle the absence of the two grounding factors, reviewability and reviseability. I’ll discuss reviewability first.

<a>Coordinating ideas

Planning speech while speaking poses a number of problems for conversational

participants. One is to be able to utter a smooth passage of speech while at the same time monitoring the way the thoughts are being expressed, maintaining interlocutor's interest, checking for understanding, and other constraints, like making sure there is time to say what needs to be said in the time available for the conversation to take place. How are all these complexities of the everyday speaking situation maintained in a coherent way?

Because speakers need to keep several layers of information in mind as they're speaking, they sometimes experience difficulty in saying what they mean to say. Speakers start talking before they are fully prepared, and then later revise or qualify what they've said, as in the following example (adapted from Svartvik & Quirk 1980: 1.5.416 - 1.5.421):

(9) but what functions, do people variously fill, I mean are you . all members of a research . project, or just a group, I mean is . Marilyn, . uh: uh assistant le uh I mean is she a lecturer?

In this example, the speaker paused, said uh, and restarted several times before finally getting to the question Is Marilyn a lecturer? Speakers also sometimes mention information out of order, and have to return to an earlier part of their discourse to clarify, as in the following example from a corpus collected by Herbert Clark:

(10) uh I believe there are two people talking Susan and Kevin she is a tv broadcaster or something and he's a journalist and they're discussing Gary Hart and the Miami Herald story about his- his supposed affair with Donna Rice and um she asks him where he was when he heard about it and he says he was in Denver oh he was apparently canvassing for Hart and um she asked him if feel- if he felt angry  
about it

In this example, the speaker is prepared to relate the information about the t.v. broadcaster's



question about whether the journalist felt angry, but realises that a crucial piece of information is missing, namely that the journalist was also canvassing for Hart. So the speaker says oh and retraces the temporal order of her narrative to add the missing information.

The difficulty in ordering ideas in a discourse arises because of the inherent problem in having sequential or linear speech and yet nonlinear ideas. Ideas frequently have many levels, such as a primary goal with several subgoals. Despite their layering, the constraints of speech are such that the ideas have to be expressed one at a time. This has been called the linearization problem (Bestgen & Costermans 1994; Levelt 1989). On top of this problem there is the additional problem of expressing the relationship between the ordered ideas. Some ideas might be more or less closely related to each other. For example, collections of ideas might work together as a unit, and we might want to express the different relationship between the ideas in the collection versus between the units (Bestgen & Costermans 1994). In dealing with these two problems, speakers often find themselves making asides or shifting focus as they talk, which further reduces the local coherency of adjacent utterances (Jucker 1993).

Understanding conversations involves more than a sentence by sentence analysis of incoming speech. In the face of the linearization problem, the problem of expressing relationships between ideas, and the potential inconsistencies between utterances, how do conversational participants manage to build an integrated picture, while at the same time coordinating their speech with each other? One way they do this is by using discourse markers and other devices found frequently in spontaneous speech but not in prepared speech or written text. In the rest of this section, I will discuss the general inference making process, the types of problems that arise in spontaneous speech to thwart the inference making process, and how speakers deal with these problems in such a way that they preserve maximum coherency.

One problem that addressees face in all communicative mediums is how to connect one sentence to its following sentence, or how to build bridging inferences (Clark & Haviland 1975). The difficulty of building a bridging inference between two sentences varies depending on how much information needs to be inferred. Clark and Haviland (1975) found that The beer was warm took less time to read after Horace got some beer out of the trunk than after Horace was especially fond of beer. This is because in the first case, some particular beer was in mind to connect with the concept warm, but in the second case, there was no particular beer in mind. In the second case a larger bridging inference had to be constructed between Horace's generally liking beer and a particular beer that was warm.

In spontaneous speech, inferences between successive utterances can often require a large leap. Speakers do not always have enough time to put their thoughts in order and to check that each idea follows from the previous idea. Speakers can also change their speech plans on the spur of the moment to add asides or to switch topics. But speakers do have tools available to them to help listeners make the correct bridging inferences at these difficult transition points. We'll discuss how interlocutors deal with the problem of time pressure first, and then we'll talk about the problem of change of plans.

One way of dealing with the lack of time to prepare orderly utterances is by using techniques to gain time while still holding the floor. One technique is the use of stock phrases such as "this is the point" (Schiffrin 1987: 328) or "They're away and racing" in a racetrack commentary (Kuiper 1996: 17). Although it's not always possible to determine what is a stock phrase and what is not, most speech has some more or less formulaic aspects (Kuiper 1996). Formulaic phrases may aid speech production by providing the opportunity to access larger than one-word chunks directly from memory, and thereby freeing processing resources (Kuiper

1996). They may also aid comprehension by “making speech more predictable” and spreading out the resources necessary to understand a stretch of speech (Kuiper 1996: 98). Single words such as bueno (good in Spanish), well, then, or um are also thought to free processing resources by maintaining the floor while speakers plan their upcoming utterance (Brody 1987; Jucker 1993; Redeker 1991). The here-and-now nature of talk puts pressure on the language processing system to operate quickly while at the same time conforming to the short term memory constraints that arise out of the lack of reviewability of speech.

Gaining time is one technique speakers can use to get their ideas straight, but it is not always effective. Time pressure can cause speakers to speak before they’re fully prepared, which can result in speakers’ belated realization that they’ve omitted some crucial information. When ideas get out of order, interlocutors face the problem of determining how to correct the discourse to represent the accurate order of information, given linear utterances and turn taking. One way speakers do this is by using a speech signal such as oh or explicit phrases like oops I forgot something to indicate to listeners that the upcoming speech won’t fit in to what’s just been said, but rather fits in to an earlier stretch of discourse (Fox Tree & Schrock 1999; Redeker 1991; Schiffrin 1987). That is, the signals inform listeners not to build bridging inferences between these out-of-order utterances. People’s natural inclination is to assume continuity between sentences in the absence of evidence to the contrary (Segal, Duchan, & Scott 1991). Without a signal to halt bridge building, people will always try to bridge adjacent ideas.

Even when speakers do have enough time to organise their ideas and are able to correct organization errors, they can still suddenly decide to alter their speech plans. That is, speakers can have said what they needed to without omitting background information but decide spontaneously and intentionally to change their conversational focus. But as with the time-

crunch problems, speakers have tools available to them to change speech plans on the spur of the moment while maintaining continuity for the listener. The main way they do this is with discourse markers.

Discourse markers have been described as unwanted particles of speech that interrupt the speakers' message (Adams 1982; Johnson 1961; Levin, Silverman, & Ford 1967). One reason that they might have been viewed as extraneous is because people don't need to have them in order to understand language. Think of any newspaper article or slick radio advertisement: written words and rehearsed orations pose no problem for the comprehension system. So in some sense, discourse markers are superfluous to the understanding of language. But at the same time, discourse markers are not empty words. They cannot occur at any point in a conversation, as pointed out rhetorically by Schourup (1985: 162): "Did anyone, except the critics, ever really say 'Like Hi!'" James (1972: 164-165) offers many other examples, including "\*With a hammer . . . well . . . Bill hit Fred" and "\*I'll throw the dinner I just ate . . . oh . . . up." Discourse markers can also not be substituted for each other (Redeker 1991: 1165, adapted from Schiffrin 1987: 93):

(11) Henry: Do you know where Abe's is?

Debby: Yeh I know where Abe's is.

Henry: Right across the street.

Debby: Oh (\*Well) it's that way.

They only occur at certain points in a conversation and for certain reasons.

Analysts have identified a wide array of functions discourse markers might serve, such as to show politeness, to make a conversational setting less formal and more intimate, to play down interpersonal difficulty, and to identify with a social group (Bernstein 1962; Brody 1987; Jucker

1993; Kotthoff 1993; Maschler 1994; Östman 1981; Stubbe & Holmes 1995). We will be concerned with the use of discourse markers in facilitating the grounding process. Discourse markers help interlocutors get around the lack of reviewability that can cause disorganised ideas and thwart the inference making process, and, as will be discussed in the next section, they also help interlocutors get around the problem of the inevitable occurrence of speech disfluencies, a result of the lack of reviseability.

Discourse markers like oh, then, actually, now, and well can help listeners deal with speakers' shifts of topic and focus by indicating when a topic shift will occur (Aijmer 1988; Bestgen & Costermans 1994; Heritage 1984; Schiffrin 1987), as in the following two examples (adapted from Svartvik & Quirk: 2.7.187 - 2.7.199 and 2.8.431 - 2.8.435):

(12) A: I think it's a snotty place the the Academy - .

B: oh, . from the point of view of non-smoking, I find it marvelous, says she, strikes a match - - . actually they've got a film, . on now, that . was on at dhi: festival, - just opened at the Academy

(13) A: the reason it was quiet before, nineteen sixty-eight, was because . you can argue, is because . the British - didn't . didn't stir up the Northern aiuhr uh um the Ulster Protestants,

B: well it wasn't entirely quiet

Because grounding is a continual process of updating a discourse with each new contribution, it will sometimes happen that speakers have built up a background that doesn't match what they next want to say (Jucker 1993). Discourse markers can signal the need to alter prior assumptions, helping listeners to build bridging inferences between utterances.

Other markers can be used to show the return to a prior topic of conversation. Anyway

and variants like but anyway and anyway be that as it may can be used to mark the end of a digression and the return to the prior topic, as can be seen in the following example after the brief digression you know what getting up Sunday's like (Bublitz 1988: 118; Takahara 1988; example adapted from Svartvik & Quirk 1980: 2.7.41 - 2.7.48):

(14) A: I had some people to lunch on Sunday, and . they turned up half an hour early,

B: really

A: I mean you know what g getting up Sunday's like, anyway, and - . I'd - I was behind in any case, . and I'd said to them one o'clock

Stock phrases like as I was saying can also be used to signal the return to an earlier topic (Bublitz 1988).

Two superficially similar discourse markers, and and then, can be used in contrasting ways to indicate either topic continuity or topic shift in text comprehension, and might serve similar functions in listening. And is used to show continuity and then is used to show that a new idea will come up (Bestgen & Costermans 1994). The choice between one or the other provides very specific information to the reader, and likely listener as well, about what to expect next. Removing either of them from the discourse would eliminate the forewarning of a new idea, which is potentially useful information for building bridging inferences.

So, to deal with the two main ways inference making can be thwarted, by having disorganised ideas or by suddenly shifting focus, speakers can use a number of carefully placed discourse markers. But even when inference making is not thwarted, there is still the problem of expressing the relationship between ideas so that it is clear to listeners which ideas go together and how groups of ideas relate to one another. Here too speakers can use discourse markers to move between different layers of talk and to indicate different kinds of perspective shifts.

Speakers can indicate shifts at a register level, such as moving from joking back to seriousness, with the marker I mean (Redeker 1991). They can indicate shifts between speakers' stances, such as moving from statements to evaluation of those statements, with the word well (Schiffrin 1987). And they can indicate shifts between the speakers themselves and the characters in the speakers' narratives, as in the following where the critical discourse markers are italicised:

(15) and I said well what does she want it for, he said . oh I I don't know, but Ella needs it, she called for it (adapted from Svartvik & Quirk 1980: 2.13.1244 - 2.13.1247)

(16) Both sides of the street can hear her yelling at us and she's like "Come in here and have a beer" you know? (adapted from Schourup 1985: 43)

(17) But the twins in the family say well they were so surprised that of all the people, that she had the twins (adapted from Schiffrin 1987: 125)

Well, like, and oh have been thought to mark shifts between the speakers' talk and the talk of the character in a story the speaker is telling, introducing both direct and indirect quotes (Jucker 1993; Redeker 1991; Schourup 1985; Schiffrin 1987). In addition to forewarning a shift between the perspective of the speaker versus that of the character, these discourse markers help limit the confusion about who said what in a storytelling situation, so that it is clear whether a speaker's I refers to the speaker or to the character; in the above example, the I of he said oh I don't know refers to the he who is talking, not to the narrator of the story.

Discourse markers might also be used to indicate shifts between major and minor idea breaks. The different kinds of breaks are associated with different temporal markers in writing (Bestgen & Costermans 1994), and these markers might be used in the same way in conversations. Anchorage markers, like in the afternoon, are markers of precise time and tell exactly when an event occurred for later events to refer back to. Sequence markers, such as

afterwards, show the relative organization of events. In written text, anchorage markers are used at major idea breaks and sequence markers at minor breaks (Bestgen & Costermans 1994). These discourse markers keep the flow of ideas coherent by providing information about how to relate sentences.

In addition to forewarning perspective shifts, another way speakers help listeners interpret the relationship among ideas in a discourse is by providing information about what they are thinking. Because speakers are driven to express ideas in a brief amount of time and with limited elaboration to comply with the constraints of turn-taking and least collaborative effort, their contributions might end up vague or not well thought out; it is to both speakers' and addressees' advantage if addressees can predict where speakers' ideas are headed.

Discourse markers can serve this function is by indicating a mismatch between what's said and what's intended or by indicating that speakers have not completely filled out their intentions, inviting addressees to complete the ideas (Jucker & Smith 1998; Schourup 1985). Researchers argue that like, I mean, sort of and kind of indicate what's said is not exactly what's intended, that you know indicates that what's said is incomplete, and that well and oh indicate that speakers are choosing between alternatives for what to say next, as in "another guy comes in, a little more heavy-set guy, and uh mmh *ohh* a banker of some sort" (Redeker 1991: 1154; Jucker 1993; Schourup 1985; Stubbe & Holmes 1995). Speakers can also use um and uh to indicate that their answers to questions might not be correct; with an um or uh, answers are perceived as less likely to be correct than had the speaker used a pause of the same length (Brennan & Williams 1995). By marking their utterances as equivocal, speakers can inform listeners that they may need to do more work to interpret the speakers' meaning. Having information about the speakers' knowledge states and being able to predict upcoming utterances



aids in the grounding process and helps maintain least collaborative effort in an exchange.

Addressees can use the information in evaluating the speakers' contributions and planning their own.

The multifunctionality of some discourse markers may seem to threaten a functional account. After all, if they can do so many things, how can a listener know which use to apply in a given instance? One reassuring observation is that markers can be uniformly interpreted enough to be used deceptively. Well can be used to get listeners to believe that the next statement is relevant when it may not be (Schiffrin 1982). You know's use in getting addressees to take what's said to be common ground (Östman 1981) can be used to elicit agreement from addressees in an argument, or "as a subtle means of getting the hearers to admit to the validity of a premise" (Watts 1989: 218). Incidentally and by the way, which mark temporary digression, can be exploited to permanently steer conversations away from topics the speaker doesn't want to discuss (Bublitz 1988). Oh can be used to show information is new when it isn't, or withheld to show it's not new when it is (Fox Tree & Schrock 1999; Heritage 1984). The deceptive use of discourse markers is only possible because speakers can reliably count on their being interpreted in particular ways.

Another way of approaching the worrisome multifunctionality issue is by arguing that a particular function of a discourse marker arises out of a combination of an underlying function and a particular pronunciation, position in turn, syntactic placement, or pragmatic environment (Erman 1987; Östman 1981; Schiffrin 1987; Stenström 1990).

A third way of exploring multifunctionality is by directly measuring what's understood. Although evidence using experimental approaches is sparse in comparison to corpora analyses, it is a useful direction for future studies. Evidence from reading experiments has supported the

claims that discourse markers are used in predictable ways and are beneficial to discourse processing. When connectives such as because and although linked two sentences, the second sentence was read more quickly and integrated better, as measured by accuracy on comprehension questions, than when the connectives were absent (Millis & Just 1994). When then, and, so, because, and but were used to link sentences, the relationships between sentences were interpreted differently from when the connectives were absent; for example, readers were more likely to agree that two sentences were temporally related when a marker was present than when it was absent (Segal et al. 1991). Listeners may use markers like readers do. But they also may not. As discussed above, the problems listeners face are different from the problems readers face, so markers may function differently in the two modalities. This is on top of a perhaps more basic problem in extrapolating reading findings to listening: different discourse markers are used in writing and speech (Flowerdew & Tauroza 1995; Stenström 1990).

Direct tests of how spoken discourse markers affect listeners' interpretations have supported the idea that discourse markers are beneficial. A videotaped lecture containing markers was understood better than the same lecture with the markers edited out, as measured by the amount of lecture material recalled and the accuracy of responses to a test on the material (Flowerdew & Tauroza 1995). As another example, overhearers can complete a referential card task better when there are discourse markers than when there aren't (Fox Tree 1999). But in this study, it is unclear whether it is the number of markers or other co-occurring factors such as the number of differing perspectives that is driving the effect (Fox Tree 1999). Finally, other experimental tests using word monitoring tasks (Fox Tree & Schrock 1999) and semantic verification tasks (Fox Tree & Schrock 1999; Gernsbacher & Jescheniak 1995) have also demonstrated the beneficial contribution of discourse markers and other spontaneous speech

signals to language comprehension.

In word monitoring, people listen for the occurrence of a particular word in an utterance and press a button if they hear the word. The speed at which they press the button is related to the comprehensibility of the utterance up to that point (Fox Tree 1995; Marslen-Wilson & Tyler 1980). In a study on the use of the discourse marker oh, researchers found that listeners were faster at detecting a word in an utterance after they had heard an oh than they were at detecting the same word when the oh had been digitally removed, demonstrating the on-line beneficial effects of oh to language comprehension (Fox Tree & Schrock 1999). Oh can signal addressees to halt the building of bridging inferences or to expect an updating of earlier information, and addressees benefit from this signal.

In semantic verification, people see a word appear on a computer screen as they are listening to an utterance, and they press a button corresponding to whether or not the word they saw had been said in the utterance they heard (Fox Tree & Schrock 1999; Gernsbacher & Jescheniak 1995). The discourse marker oh was found to have a predictable effect with this task as well; people were faster at verifying that a word had been said when the visual target was presented after an oh than when it was presented at the same point in the utterance but with the oh excised, once again demonstrating the usefulness of oh in on-line language comprehension (Fox Tree & Schrock 1999). Stressing a word and using this cataphorically, as in “So a man walks into a bar with this parrot on his shoulder,” was also found to improve the incorporation of upcoming information into the discourse model (Gernsbacher & Jescheniak 1995: 26). For example, listeners were faster at confirming that the word ashtray had been said when the utterance was “she just had to buy this ashtray” than when it was “she just had to buy an ashtray” (Gernsbacher & Jescheniak 1995: 44).

Other evidence that discourse markers are important to interpretation comes from observing second language learners. Omitting discourse markers can cause the speech of nonnative speakers to sound odd (Stubbs 1983), as can using them incorrectly. After living in the U.S. for 17 months, some people learning English still had not mastered the pragmatic use of and as a temporal marker instead of as a logical conjunction (Bouton 1994). When given the information “Sandy went to Philadelphia and stole a car,” these speakers had difficulty determining if one act preceded another or if the two were simultaneous (Bouton 1994: 162); native English speakers agreed that going to Philadelphia came first. Interpreting discourse markers correctly is a skill that has to be learned.

The spontaneous speech phenomena we have just discussed, including discourse markers, stock phrases, and cataphoric this, are all used regularly by conversational participants. They play a role in coordinating discourse, achieving grounding, and expressing attitudes among conversational participants. They constrain interpretations between utterances and the types of bridging inferences that are built. In their absence, people make a greater variety of interpretations about the relationship between two ideas (Segal et al. 1991). They help coordinate spontaneous talk by creating connections between the sometimes disparate utterance productions that can result from the lack of sufficient planning time in speaking. And when all else fails, they can be used to show that what’s being said is not exactly what’s meant, or what would have been said had there been enough time to plan. Without them, information would be lost; listeners would not have forewarning about an impending need to adjust grounding assumptions, they would not have signals that they should hold a certain concept in mind for future reference, and they would not have information about how to organise or evaluate speakers’ talk. Discourse markers not only have clear functions in spontaneous speech, which can be used to promote

understanding or exploited for deceptive ends, but they can also be required.

Discourse markers are not unwanted interruptions but devices that are purposefully used to overcome grounding problems that arise with the communicative medium of spontaneous speech. They are important precisely in those communicative situations where they occur, where language is not written down or prepared in advance but is produced on the fly in spontaneous discourse. We now turn to another byproduct of speaking on the fly, disfluencies.

#### <a>Coordinating repairs

In order to maintain turns in a conversation and speak in a timely manner while ensuring moment-by-moment interlocutor comprehension, speakers often find themselves either speaking before they're ready or needing to adjust what they've said on the fly based on addressees' feedback. These problems arise because of the absence in spontaneous speech of one of the grounding factors described earlier, the lack of private reviseability. How do listeners follow speech with errors, restarts, revisions, or the sudden insertion of out-of-order information? Several devices exist to help interlocutors maximise understanding in the face of speech disfluencies and errors.

The terms disfluency and error cover a wide range of problems in spontaneous speech. These include 1.) leaving long silent gaps between words, 2.) slips of the tongue, such as saying knoor dob instead of door knob, 3.) mispronunciation, such as pronouncing Yosemite as /yo-seh-mayt/ instead of /yo-seh-mi-tee/, 4.) repaired speech, as in the house- the red house, and 5.) pragmatic errors, such as a clerk's answering a customer's question What time is it? with We close at seven when the question was asked with the intention of finding out the actual time and not as an indirect way to find out what time the shop closed. These categories are not mutually exclusive: slips of the tongue can be repaired and a pragmatic error can contain long silent gaps.

A problem in speech can be either repaired or unrepaired. Repaired and unrepaired disfluencies and errors pose different problems for interlocutors. Unrepaired errors either result in communicative breakdown, which we are not concerned with here in our investigation of how interlocutors successfully navigate problematic speech, or they result in no threat to communicative success, such as with some misplaced word accents (Cutler 1983). Disfluencies and errors are repaired when interlocutors feel it is important to do so for communicative success. The repair poses problems for both speakers and addressees. Speakers need to make repairs that clearly indicate what is to be removed from the discourse record and what is to be retained. Addressees need to follow these repairs so that they can piece together what the speakers' intended utterance is.

Repairs can be categorised into two main types: self-repairs and other-repairs. In self-repairs, speakers notice and correct problems without recourse to other conversational participants. With other-repairs, speakers and interlocutors work together to resolve problems, as in the following example (adapted from Svartvik & Quirk 1980: 1.13.613 -1.13.614):

(18) A: this is one that's Milligan tsh uh . \*fire bone china\*

B: \*Milligan china\*

While Speaker A searched for the way to express the type of dishware, Speaker B jumped in to complete the partial noun phrase that began with Milligan by saying Milligan china. Both interlocutors work to resolve the word finding problem. We'll talk about self-repairs first, and then other-repairs.

Self-repairs can be broken down into three main types. One is repetition, as in "well I'll I'll get them through . quickly" (Svartvik & Quirk 1980: 1.1.220). Repetitions can be of words or phrases, as in "I mean it isn't just . it isn't just this morning" (Svartvik & Quirk 1980: 3.3.814).

A second type of self-repair is a restart, where speakers start to say something, but then restart their utterances, as in “but it’s far more than . well it lasts quite a time” (Svartvik & Quirk 1980: 1.4.937). A third is midway between a repetition and a restart, where some aspects of the information are repeated, and some are not, as in “they had no . riuh they had never they had no plans whatever” (Svartvik & Quirk 1980: 2.3.120). These will be called replacements.

Self-repairs are common in spontaneous speech because speakers are constantly monitoring what they say. This means checking both that they are making sense and that what they are saying is appropriate to the situation. If there is a problem, they need to fix it as they speak. In other language production situations, such as writing email or letters, people don’t make as many errors because they have time to review their words before addressees receive the information. Although revising talk publicly is a necessary component of speaking spontaneously, the revision process has also developed so that it is less burdensome to listeners.

One way that speakers help listeners recoup after a stumble is by starting repairs with words that are systematically related to the speech they replace. This relatedness can be exploited in determining exactly what the intended utterance was. In interpreting repairs, listeners can use either the word-identity convention or the category-identity convention (Levelt 1989: 493):

(19) The Word-Identity Convention

If the first word of the repair is identical to some word w of the original utterance, the repair is to be interpreted as a continuation of the original utterance from w on. (If there is more than one such word in the original utterance, take the last one.)

(20) The Category-Identity Convention

If the syntactic category of the first word of the repair is identical to the syntactic

category of some word w of the original utterance, the repair is to be interpreted as a continuation from w on, with the first repair word replacing w. (If there is more than one such word in the original utterance, take the last one.)

Listeners can use one or both of these conventions to accurately determine where to attach the repair syntactically. The word-identity convention allows for a replacement with a lead-in word, such as the word to in “Right to yellow, uh to white,” the category-identity for replacements without lead-ins, as in “From the green disc to up to a pink . . ., orange disk” (Levelt 1983: 90; italics removed). In a corpus of route descriptions, there were few violations of these conventions; by the first word of the repair, listeners had enough information to connect repairs to earlier aborted speech (Levelt 1989: 495).

By systematically relating the first word of the repair to an earlier word in the utterance, speakers can indicate to listeners where to begin a repair. They can also signal how far back in the speech stream the correction lies by the editing expressions they use between the reparandum and the repair. Editing expressions are words or phrases like um, I mean, sorry, and “oh, that’s impossible; I will start again, ok?” (Levelt 1989: 482). Reparandums are the stretches of the speech that are to be replaced, and repairs are the stretches that do the replacing. Saying uh or um before the repair signals that the reparandum is likely to be only a short distance back, around 1.7 syllables; saying sorry or that is signals that the reparandum is likely to be further back, around 4.3 syllables (Levelt 1989: 484).

Editing expressions might also be used to distinguish between two categories of repairs: error repairs, where a wrong word or phrase is used, and appropriateness repairs, where an inappropriate or imprecise word or phrase is used (Levelt 1983 1989). In the route description corpus, 62% of error repairs had editing expressions, compared to 28% of appropriateness



repairs (Levelt 1989: 483). The choice of editing device might further distinguish the kind of repair taking place. In Dutch, uh or no were used for error repairs, and so was used for appropriateness repairs (Levelt 1989: 483). This systematicity allows listeners to predict whether a replacement or an adjustment will follow after an editing expression is heard.

Non-experimental observations of the use of um and uh support the hypothesis that these words are informative. With an um or an uh, speakers can indicate “not just the unavailability of a word, but a *relevant* unavailability - one that impedes the ongoing development of the talk in progress” (Goodwin & Goodwin 1986: 55-56). That is, speakers are indicating not just that a word is not immediately accessible, but that finding the word is imperative for communicative success. When they can’t find a word but also don’t need to find the word, speakers don’t say um; instead they do something else, like saying something or other in place of an inaccessible last name in the example “what the hell was her name. Karen. Right. Karen. her name was Karen something or other” (adapted from Goodwin & Goodwin 1986: 54). These observations support the claim that ums and uhs are not said randomly but instead supply information.

Experimental observation also supports the hypothesis that at least uhs are informative. People are faster to identify words after uh than they are to identify the same words when the uh is digitally excised, both in English and in Dutch (Fox Tree 1997). People are also faster to identify which object is being described when the description is in its original form, uh plus repair, than when uh and repair have been edited out (Brennan & Schober 1997).

So both corpora analyses and experimentation support the idea that editing expressions are useful in listeners’ recovering from repairs. But editing expressions are infrequent with one kind of repair, restarts. Only 6% of restarts contained the editing expressions um or uh (Levelt 1989: 494).

When speakers completely restart what they are saying, the repaired part shares little or nothing in common with the reparandums, so it's possible that editing expressions' potential benefit in signalling how far back the correction lies or what kind of correction it is is just not useful enough.

Without the benefit of self-repair conventions or editing expressions, how do listeners recover from restarts? The answer is they don't recover as easily from restarts as from other repairs. A restart that requires figuring out what information to excise from the discourse record and what to retain will slow listeners down in comprehending spontaneous speech, but a different kind of repair, a repetition, will not (Fox Tree 1995). It takes longer to monitor for the word looks in and the next figure, that has- it looks a little like a uh like a hammer than it does to monitor for it in and the next figure, it looks a little like a uh like a hammer, the identical sentence with the false start this has digitally excised. In contrast, monitoring for the word shield in and the inner part that looks like a uh like the shape of a of a shield or a weapon is not slowed down by the presence of the repetition of a (Fox Tree 1995). Fortunately, there is reason to believe that if a speaker is to make a repair, restarts are a last resort. In a sample of 607 repaired noun phrases beginning with the in the Svartvik and Quirk corpus (1980), about 36% contained repetitions, but only 11% contained restarts (Fox Tree & Clark 1997).

Restarts cause noticeable trouble and are unlikely to be forewarned by editing expressions. But there are alternative forewarning tools speakers may use. One of these is pronouncing words preceding trouble in marked ways. For example, in noun phrases the determiners the and a may be pronounced in elongated form as thee and ay instead of thuh and uh (Fox Tree & Clark 1997). In one analysis, when thee was used instead of the unmarked pronunciation thuh, 81% of the time the determiner was immediately followed by a suspension

of speech, such as a pause, um or uh, or a repair. For comparison, only 7% of thuhs were immediately followed by suspensions (Fox Tree & Clark 1997). Thee also signaled to listeners that the suspension was likely to be relatively severe. For example, looking only at the noun phrases with suspensions, 76% of restarts were after thee (24% after thuh) compared to 42% of replacements (58% after thuh; Fox Tree & Clark 1997). So although there may be fewer editing expressions before restarts, there are still other signals of impending trouble, such as marked pronunciations of the words preceding the restarts. Forewarning can mitigate the effects of speaker revisions by alerting listeners that they should pay more attention to the upcoming speech because it is likely to require a repair.

In addition to forewarning repairs and making repairs in particular ways, speakers may also aid listeners in recovering from speech suspensions by continuing to speak after suspensions in ways that preserve continuity (Clark & Wasow 1998). For example, speakers may choose to repeat what they said before the suspension after the suspension, as in “I uh I wouldn’t be surprised at that,” instead of merely continuing after the suspension, as in “I uh wouldn’t be surprised at that” (Clark & Wasow 1998: 236). In this view, repetitions are not problems as much as they are solutions. Disrupting the continuity of phonological phrases can lead to slower recognition of words in speech (Fox Tree 1995); preserving continuity may do the opposite and aid processing.

Evidence of listeners’ active work in creating successful repairs is demonstrated by the existence of other-repairs, which contrast with the repairs within a speaker’s own speech discussed up to now. If listeners were not actively engaged in the communicative process, other-repairs would not exist. One common kind of other-repair is pragmatic repair. Unlike disfluencies, which can often be described at the word level or at most the sentence level,

pragmatic errors are errors that span several sentences as in the following constructed example where the boldface utterance is disjointed from the rest of the passage (Kreuz & Roberts 1993: 244):

(21) A: Would you believe I got *another* speeding ticket today?

B: Not again! What happened this time?

A: Same thing as always. They have this speed trap I have to drive through to get to work, and I was late again, so . . .

B: You should find some other way to get to work. Why don't you try using that exit at Washington Street?

**A: Don't the police have anything better to do?**

B: Well, what's wrong with the Washington exit?

A: They've got it blocked off for the next few months. I think they're paving it.

B: Oh. That doesn't leave you too many options, does it?

In this example, the line "Don't the police have anything better to do?" introduces the pragmatic error, but the whole discourse is necessary for the line to be interpreted as an error.

Pragmatic repairs have different effects on overhearers from lexical or phonological repairs. Spoken passages containing pragmatic errors were compared to the same passages with lexical or phonological errors substituted for the pragmatic errors; in the above example, "Don't the police have anything better to do?" was replaced with "They've got Jefferson, I mean, Washington blocked off for the next few months," with the subsequent dialogue altered accordingly (adapted from Kreuz & Roberts 1993: 244). Both kinds of repairs caused the personality of the speaker of the problematic sentence to be rated more negatively than in passages without repairs. But in contrast to lexical or phonological errors, with pragmatic errors,

the addressees were also rated more negatively than without the error. Though the utterer of a lexical or phonological error seem to be solely responsible for those mistakes, both interlocutors seem to be responsible for pragmatic errors. Overhearers treat the resolution of pragmatic errors as a mutual task where both interlocutors can be held accountable. Addressees have a recognised responsibility in avoiding and resolving communicative breakdowns.

The idea of achieving repair by mutual work helps explain why not all errors are corrected. When a correction is not necessary to achieve mutual understanding, errors can be left unrepaired. When a word is pronounced with a mistaken stress that preserves vowel quality, such as “You think it’s sarCASm, but it’s not,” speakers don’t correct the word; but when the mispronunciation leads to a distortion of the word that might threaten comprehension, it is corrected (Cutler 1983: 85-86). Likewise, when sentence accent is anomalous but doesn’t lead to an interpretation different from its intended interpretation it isn’t corrected, as in “The only trouble WITH it -- (pause) -- is the hood is too small;” but when the accent does affect the meaning, it is corrected, as in “and what I’M saying -- what I’m SAYing is” (Cutler 1983: 86-87). The driving force behind whether or not talk will get corrected or not is whether or not the correction is necessary for understanding.

When people need to adjust what they’re saying, they don’t just stop what they’re saying and say what they really mean. Instead, they adjust their speech in ways that help listeners successfully correct their discourse model to accommodate the repair. By (1) preserving continuity, (2) using editing expressions and elongation to forewarn the presence and type of upcoming repairs, and (3) making repairs that follow the conventions about relating the first word of the repair to the reparandum, speakers can help listeners identify errors and make corrections smoothly. It is in speakers’ best interest to make repairs that help listeners to follow

along, and in listeners' best interest to use all available information to process speech.

#### <a>Conclusion

In this chapter, I discussed the use of timing, discourse markers, editing expressions, and other devices in resolving the problems that arise in the communicative medium of spontaneous speech. The main obstacles discussed were the effervescence of speech, which leads to an inability to review the discourse record, and the cotemporality of the interlocutors, which limits private reviseability and creates a pressure to speak before being fully prepared, resulting in speech errors and disfluencies.

Speaking on the fly -- in unplanned, unrehearsed conversation -- leads to the production and use of specialised words, phrases, and pronunciations in order to achieve the grounding criterion, coordinate turns, organise ideas, correct errors, and warn listeners of upcoming speech production trouble. The nonessential words and phrases of spontaneous speech are not so nonessential after all. Instead, these hallmarks of conversations need to be attended to on a moment-by-moment basis as discourse is building up. They cannot be ignored even for one moment, or else one person risks being misinterpreted by a conversational partner.

## <a>References

Adams, M. R. (1982). Fluency, nonfluency, and stuttering in children. Journal of Fluency Disorders, *7*, 171-185.

Aijmer, K. (1988). "Now may we have a word on this": The use of "now" as a discourse particle. In M. Kytö, O. Ihalainen, & M. Rissanen (Eds.), Corpus Linguistics, Hard and Soft: Proceedings of the 8th International Conference on English Language Research on Computerized Corpora. Amsterdam: Rodopi.

Argyle, M., Lalljee, M., & Cook, M. (1968). The effects of visibility on interaction in a dyad. Human Relations, *21*, 3-17.

Atkinson, M. (1984). Our Master's Voices: The Language and Body Language of Politics. New York: Methuen.

Bernstein, B. (1962). Social class, linguistic codes and grammatical elements. Language and Speech, *5*, 221-240.

Bestgen, Y., & Costermans, J. (1994). Time, space, and action: Exploring the narrative structure and its linguistic marking. Discourse Processes, *17*, 421-446.

Bouton, L. F. (1994). Conversational implicature in a second language: Learned slowly when not deliberately taught. Journal of Pragmatics, *22*, 157-167.

Brennan, S. E., & Schober, M. F. (1997). When do speech disfluencies *help* comprehension? Paper presented at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, PA.

Brennan, S. E., & Williams, W. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. Journal of Memory and Language, *34*, 383-398.

Brody, J. (1987). Particles borrowed from Spanish as discourse markers in Mayan languages. Anthropological Linguistics, 29(4), 507-521.

Bublitz, W. (1988). Supportive fellow-speakers and cooperative conversations: Discourse topics and topical actions, participant roles, and 'recipient action' in a particular type of everyday conversation. Philadelphia: John Benjamins Publishing Company.

Clark, H. H. (1983). Language use and language users. In G. Lindzey & E. Aronson (Eds.), Handbook of social psychology, (3 ed., ). Reading, MA: Addison-Wesley.

Clark, H. H. (1996). Using Language. New York: Cambridge University Press.

Clark, H. H. & Brennan, S. E. (1991). Grounding in communication. In Resnick, L. B., Levine, J. M., and Teasley, S. D. (Eds.), Perspectives on Socially Shared Cognition. Washington, D. C.: American Psychological Association.

Clark, H. H., & Haviland, S. E. (1975). Comprehension and the given-new contract. In R. O. Freedle (Ed.), Discourse production and comprehension. Norwood, NJ: Ablex.

Clark, H. H. & Schaefer, E. F. (1987). Collaborating on contributions to conversations. Language and Cognitive Processes, 2 (1), 259-294.

Clark, H. H. & Schaefer, E. F. (1989). Contributing to discourse. Cognitive Science, 13, 259-294.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. Cognitive Psychology, 37, 201-242.

Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. Cognition, 22, 1-39.

Cutler, A. (1983). Speakers' conceptions of the function of prosody. In A. Cutler & D. R. Ladd (Eds.), Prosody: Models and Measurements. New York: Springer-Verlag.



Duncan, S., Jr. (1972). Some signals and rules for taking speaking turns in conversation. Journal of Personality and Social Psychology, 23(2), 283-292.

Duncan, S., Jr., & Niederehe, G. (1974). On signalling that it's your turn to speak. Journal of Experimental Social Psychology, 10, 234-247.

Edwards, J. A. (1993). Survey of electronic corpora and related resources for language researchers. In J. A. Edwards & M. D. Lampert (Eds.), Talking Data: Transcription and Coding in Discourse Research. Hillsdale, NJ: Lawrence Erlbaum Associates.

Erman, B. (1987). Pragmatic Expressions in English: A Study of You Know, You See, and I Mean in Face-to-Face Conversation. Stockholm, Sweden: Almqvist & Wiksell International.

Erman, B., & Kotsinas, U.-B. (1993). Pragmaticalization: The case of *ba'* and *you know*. In J. Falk, K. Jonasson, G. Melchers, & B. Nilsson (Eds.), Stockholm Studies in Modern Philology (Vol. 10, pp. 76-93). Stockholm: Almqvist & Wiksell International.

Flowerdew, J., & Tauroza, S. (1995). The effect of discourse markers on second language lecture comprehension. Studies in Second Language Acquisition, 17, 455-458.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. Journal of Memory and Language, 34, 709-738.

Fox Tree, J. E. (1997). Listeners' uses of *ums* and *uhs* in on-line speech processing. Poster presented at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, PA.

Fox Tree, J. E. (1999). Listening in on monologues and dialogues. Discourse Processes, 27, 35-53.

Fox Tree, J. E. & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. Cognition, 62, 151-167.

Fox Tree, J. E. & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. Journal of Memory and Language, 40, 280-295.

Gernsbacher, M. A., & Jescheniak, J. D. (1995). Cataphoric devices in spoken discourse. Cognitive Psychology, 29, 24-58.

Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. Semiotica, 62-1/2, 51-75.

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds.), Structures of Social Action: Studies in Conversation Analysis. Cambridge: Cambridge University Press.

Holmes, J. (1986). Functions of *you know* in womens' and men's speech. Language in Society, 15, 1-22.

Isaacs, E. A. & Clark, H. H. (1987). References in conversation between experts and novices. Journal of Experimental Psychology: General, 116 (1), 26-37.

James, D. (1972). Some aspects of the syntax and semantics of interjections. Papers from the Eighth Regional Meeting of the Chicago Linguistics Society.

Jefferson, G. (1973). A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. Semiotica, 9, 47-96.

Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), Conversation: An Interdisciplinary Perspective. Philadelphia: Multilingual Matters, Ltd.

Johnson, W. (1961). Measurements of oral reading and speaking rate and disfluency of adult male and female stutterers and nonstutterers. Journal of Speech and Hearing Disorders, 7, 1-20.

Jucker, A. H. (1993). The discourse marker 'well': A relevance theoretical account. Journal of Pragmatics, 19, 435-452.

Kotthoff, H. (1993). Disagreement and concession in disputes: On the context sensitivity of preference structures, 193-216.

Kreuz, R. J., & Roberts, R. M. (1993). When collaboration fails: Consequences of pragmatic errors in conversation. Journal of Pragmatics, 19, 239-252.

Kuiper, K. (1996). Smooth Talkers: The Linguistic Performance of Auctioneers and Sportscasters. Mahwah, NJ: Lawrence Erlbaum Associates.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. Cognition, 14, 41-104.

Levelt, W. J. M. (1989). Speaking: From Intention to Articulation. Cambridge, MA: MIT Press.

Levin, H., Silverman, I., & Ford, B. L. (1967). Hesitations in children's speech during explanation and description. Journal of Verbal Learning and Verbal Behavior, 6, 560-564.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. Cognition, 8, 1-71.

Maschler, Y. (1994). Metalanguaging and discourse markers in bilingual conversation. Language in Society, 23, 325-366.

McLaughlin, M. L., & Cody, M. J. (1982). Awkward silences: Behavioral antecedents and consequences of the conversational lapse. Human Communication Research, 8(4), 299-316.

Millis, K. M., & Just, M. A. (1994). The influence of connectives on sentence comprehension. Journal of Memory and Language, 33, 128-147.

Östman, J.-O. (1981). You Know: A Discourse Functional Approach. Amsterdam: John Benjamins.

Redeker, G. (1991). Linguistic markers of discourse structure. Linguistics, 29, 1139-1172.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. Language, 50, 696-735.

Schegloff, E. A. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organisation. In G. Button & J. R.E. Lee (Eds.), Talk and Social Organisation (Vol. 1, pp. 70-85). Clevedon, England: Multilingual Matters, Ltd.

Schiffirin, D. (1982). Discourse Markers: Semantic Resources for the Construction of Conversation. Unpublished PhD Dissertation, University of Pennsylvania.

Schiffirin, D. (1987). Discourse Markers. New York: Cambridge University Press.

Schourup, L. C. (1985). Common Discourse Particles in English Conversation. New York: Garland.

Segal, E. M., Duchan, J. F., & Scott, P. J. (1991). The role of interclausal connectives in narrative structuring: Evidence from adults' interpretations of simple stories. Discourse Processes, 14, 27-54.

Siegmán, A. W., & Reynolds, M. A. (1983). Speaking without seeing, or the effect of interviewer absence on interviewee disclosure time. Journal of Psycholinguistic Research, 12(6), 595-602.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. Journal of Memory and Language, 32, 25-38.

Stenström, A.-B. (1990). Lexical items peculiar to spoken discourse In J. Svartvik (Ed.), The London-Lund Corpus of Spoken English: Description and Research. Lund, Sweden: Lund University Press.

Stubbe, M., & Holmes, J. (1995). *You know, eh* and other 'exasperating expressions': An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. Language and Communication, 15(1), 63-88.

Stubbs, M. (1983). Discourse Analysis: The Sociolinguistic Analysis of Natural Language. Oxford: Basil Blackwell.

Svartvik, J., & Quirk, R. (Eds.). (1980). A Corpus of English Conversation. Lund: CWK Gleerup.

Takahara, P. O. (1998). Pragmatic functions of the English discourse marker *anyway* and its corresponding contrastive Japanese discourse marker. In A. H. Jucker & Y. Ziv (Eds.), Discourse Markers: Descriptions and Theory. Amsterdam: John Benjamins.

Watts, R. J. (1989). Taking the pitcher to the 'well': Native speakers' perception of their use of discourse markers in conversation. Journal of Pragmatics, 13, 203-237.