Speech act theory considers utterances as actions in and of themselves. Much of modern speech act theory began with the work of Austin (Austin 1962). Austin's theory of speech acts is built upon the notion of performatives. Austin's taxonomy of performatives, for which the act of speaking itself constitutes the action, such as:

1. I now pronounce you man and wife.
2. I hereby declare war on Germany.
3. I apologize for your inconvenience.

As the first two utterances indicate, these performatives are often of a legal or religious nature in which the speaker has the power to change the state of the world by the utterance. For example, the first utterance changes the speaker's relationship to the speaker, the second utterance changes the speaker's relationship to the world, and the third utterance changes the speaker's relationship to the listener.

1. The locutionary act is the actual utterance of a meaningful sentence.
2. The illocutionary act is the sense in which the utterance itself contains force to effect a change in the speaker's relationship to the world, e.g., the promise to perform an action.
3. The perlocutionary act is the sense in which the utterance changes the audience by changing their beliefs or intentions to perform some action.

There have been attempts to associate force with sentence form: some obvious examples include commands, questions, and assertions. For example, "Can you pass the salt?" is usually meant as a request, not a question, while "It's pretty noisy here" might be uttered as a request to turn down the radio volume. (Searle 1976) has also developed a more detailed taxonomy of five classes of locutionary action conveyed by an utterance.

1. Representatives are sentences by which the speaker asserts or commits to the truth of the utterance.
2. Directives are requests to the listener to do something. Questions are directives in that they attempt to evoke a response.
3. Commissives, or promises, commit the speaker to future actions.
4. Expressives create a psychological state rather than causing a physical action. They include apologizing, thanking, etc.
5. Declarations are the performatives mentioned earlier such as declaring war or firing an employee.

The nature of speech acts is a very insightful explanation of the purpose of utterances, and much of our speech can be effectively analyzed according to Searle's taxonomy. But speech act concepts have practical limits. Speech is very social in nature, and many utterances are oriented less towards causing an action than towards establishing or maintaining a social relationship. The Coordinator, a software tool that utilizes speech act theory for tracking work team assignments by requiring users to assign speech act classifications to electronic mail messages, has met with mixed acceptance in actual use (Winograd 1989, Medina-Mora et al. 1992). Although utterances may be successfully classified by speech act theorists, in the actual interplay of office communication correspondents may not wish to be so direct with each other.

Conversational implicature and Speech Acts

Conversational implicature is the principle that an utterance often contains much more meaning than the words themselves indicate directly. Imagine that two people meet on the street and the following conversation ensues.

A: Did you do it?
B: Not again.

This conversation is largely meaningless to us; the conversants make extensive use of shared knowledge and history, and we cannot know to what action they refer. We can, however, make certain inferences about the dialogue: A and B have discussed "it" before, "it" is dominant enough in their relationship not to be ambiguous, and A assumes "it" is unlikely to be ambiguous. Another inference from the conversation is that B has done this "it" in the past and likewise understands that A knows about this prior occurrence (or most recent of several prior occurrences of "it").

To make these inferences with any degree of confidence, we resort to certain implicit assumptions about the use of language. We assume that the utterances have a purpose, e.g., A and B are not total strangers but in fact have a common history. We make assumptions about shared belief, e.g., A does not expect B to know the referent for "it" and not simply teasing B to see whether he or she can guess what is being talked about. In short, we assume an order or regularity to these utterances.

Grice's theory of conversational implicature (Grice 1972) is based on the concept that there is a set of guiding assumptions molding how a conversation is organized. Such guidelines allow us to make the inferences described above, and any deviation from the guidelines must itself have implications beyond the actual words used. Grice's maxims of conversation can be summarized as follows:

- The cooperative principle: Speak to support the accepted purpose or direction of the conversation as if it is at the moment of the utterance. This is the underlying theme of all the maxims.

Higher levels of linguistic knowledge
into turns. The flow of turn-taking provides a collaborative environment for each talker's contributions to the conversation. Conversants maintain a common focus across turns; without this, pronouns could not refer to ideas and objects mentioned in an earlier sentence. Feedback techniques to ensure mutual understanding are discussed in the subsequent section.

Regulation of Conversation

We all know from personal experience that in a conversation the various talkers take turns speaking. After each turn, remarkably little time transpires before the next turn begins. Occasionally turns overlap, as one participant begins before the previous has completely finished, but it is remarkable that conversations can be as dynamic and fast paced as they are without more “stepping on each other’s toes.” Equally remarkable are the conversational processes for selecting a mutually agreeable topic, moving on to new topics, and returning to a previous topic.

Conversation is rich in social conventions that invoke discourse and most utterances occur in a discourse context (Goftman 1981). Turns in conversations are regulated and ordered to allow a chain of utterances to refer to a single topic. Often subsequent utterances can be understood only in the context of the earlier portions of the discourse. The most simple example of this dependency is the adjacency pair (Sacks et al. 1974), in which something is presented or proposed in the first utterance and responded to, accepted, or rejected in the rejoinder. For example:

A: I brought in the mail.
B: Thank you.

A: How much does this cost?
B: Two dollars.

Note that the second utterance in the pair, which brings the pair to some form of closure, has little clarity of its own outside of the adjacency pair.

Where applicable, adjacency pairing simplifies the question of how the listener knows when the talker’s turn is over as the listener must clearly wait for the proposition to be presented in the first member of the pair. Although it may be suggested that all conversations can be reduced to sets of adjacency pairs possibly with inserted sequences of other adjacency pairs between the first and second member of a pair, most conversations are more complex and resist this analysis.

In the absence of simple pairs, it is harder to specify when one turn has ended and another talker may begin a new turn. What constitutes a turn? How does the talker signal this to the listener? Turns are often composed of one or more syntactically or semantically meaningful units. These units may correspond to sentences, but they are equally likely to be smaller phrase-like units; fluent conversation often contains incomplete sentences. One appropriate unit is the breath group, or the string of words between catching one’s breath, which usually expresses one or more coherent thoughts.
For detecting turn boundaries, the problem lies with the "one or more" of the preceding paragraph. If one talks and the second always picks up when the first stopped for breath, turn taking would be more predictable. But the talker may continue to "hold the floor" for multiple utterances or the listener may interrupt when the talker before the talker has even finished or otherwise signal so that the talker modifies the utterance even as it is being produced.

The time between turns is too short (often shorter than pauses within a turn) to believe that the listener simply waits to hear if the talker has more to say. [Duncan 1974, Duncan and 1972] analyzed a number of conversations and suggested (Duncan 1974, Duncan 1972) a number of conversations and suggested that the listener simply waits to hear if the talker has more to say. To believe that the listener simply waits to hear if the talker has more to say, it would be more predictable. But the talker may continue to "hold the floor" for multiple utterances or the listener may interrupt when the talker before the talker has even finished or otherwise signal so that the talker modifies the utterance even as it is being produced.

The completion of each syntactic unit or phrase is a possible end of the turn. If the talker indicates termination by cues such as those just listed, this invites the listener to take a turn. If the current talker desires to continue the turn, i.e., to present a subsequent phrase, the end-of-turn cues can be avoided or more strongly. The opposite behavior can be invoked such as using a rising intonation or beginning a hand gesture.

Conversation does not always break into turns cleanly. Sometimes, either deliberately or accidentally, the listener may interrupt, i.e., begin speaking before the other has finished. Interruption is usually dealt with effectively in conversation, even though party is then the talker tends to repeat or summarize what was said. During overlap, the other party is then the talker tends to repeat or summarize what was said. During overlap, the other party is then the talker tends to repeat or summarize what was said. Past research has not yet resolved who should have a turn; one party may attempt to assert control by emphasizing such as speaking more loudly or with increased pitch range or lengthened syllables. Speech may be used by the listener in a manner that initially seems to be a short turn or an interruption but does not really take a turn away from the talker. Backchannels refer to a number of behaviors whereby the listeners give feedback to the talker [Yngve 1970]. They include paraverbal utterances ("Hmmm..."), completing the other's sentence or offering a paraphrase of it, short interjections ("Of course," "You don't say?"); head nods, and various facial expressions.

Back channels are a cooperative mechanism; listener feedback indicates what is known or accepted so that the talker can continue the exposition with confidence. Back channels make for more productive conversation. For example, in an experiment by [Kraft 1982, Kraft and Lewis 1984], subjects described scenes from a film to a listener who attempted to identify the film. If the listener who was out of sight could not speak back, it took longer for the talker to adequately describe the scene. Even an easeware tab who could never be heard benefited from the listener back channel utterances but not as much as the listener who did. This suggests that some aspects of back channel cooperation produce generally "better" utterances from the talker, while other aspects of performance improvement are specific to the participation of the back channel provider.

**Listen focus**

Back channels are just one aspect of collaborative behavior in conversation. In the course of speaking conversacons change or agree upon the topic of conversation, refer back to previous topics, and reaffirm their basis of mutual belief upon which they can build successful references to world knowledge, either generic or specific and situational. The discussion of turn taking emphasized pairs or short sequences of talk. We now turn our attention to longer conversations with perhaps quite a few turns. At any moment in coherent discourse the conversants usually agree on what is being discussed. From time to time, the topic of conversation changes. The group of sequential utterances that refers to the same topic is a discourse segment. Transitions between discourse segments are often indicated by cue phrases such as "By the way...", "Yes, but...", and "Well...". Throughout a discourse segment, all utterances refer to the same topic or noun phrase; this is the focus or center of the discourse segment.

Identification of the focus of a discourse segment is required to resolve reference, which arises from several sources. Deixis is the reference of certain pronouns, such as "this" and "those" that point at something either physically or conceptually. Anaphora is the reference implied by pronouns such as "he" or "their." The entity referred to by deixis or anaphora is the referent. The referent corresponds to the focus of the discourse segment; changing the referent introduces a new discourse segment.

A discourse segment can be interrupted by the introduction of a new discourse segment, and the original discourse segment can be returned to. For example, consider the discourse fragment.

A: California gets so green in the winter, I love it!
B: Seattle gets a lot of rain in the winter too, but not much sun.
A: It's a nice city, but you should check out Hawaii if you want wonderful winter weather.
B: Last winter we went hikin there.
A: Sometimes it gets a spell of rain in February.
B: But it's not as bad as back there! It's so dreary all winter.
In the first sentence, speaker A references "California." Speaker B then introduces a new reference "Seattle." Speaker A refers back to Seattle at the beginning of the next utterance but then introduces a third focus "Hawaii" using the cue phrase "but." The next two utterances then refer to Hawaii as well. In the last utterance, speaker B jumps back to the focus of "Seattle" without needing to further specify the pronoun. How is this accomplished without further negotiation? 

(Gross and Sidner 1988) proposed a discourse model differentiating the attential structure that specifies the target of reference from the intentional structure that is roughly the pragmatic purpose of the discourse segment. They suggested a stack model for the attential structure. A stack is a data representation in which items are put on ("pushed") and removed ("popped") from the top so that the most recently pushed item is always the one that gets popped. Figure 9.11 shows the stack progressing during the example discourse. In the last snapshot, the top focus "Hawaii" has been popped, leaving "Seattle" exposed as the prime candidate for reference.

This model suggests that once popped, an object cannot be referred to again by a pronoun without being specifically introduced as a new focus so that it appears on the top of the stack again. But this is not entirely true, revealing that the model although powerful is incomplete. Speaker B might say, after a pause and somewhat longer, "It was so nice there, what a great place to walk" referring back to Hawaii. Somehow the conversants would shift back to this focus, aided by the tense shift in B's utterances.

How do we know when a new discourse segment is introduced? In addition to the cue phrases mentioned above, the way reference is used signals a new discourse segment. Gross, Joshi, and Weinstein use the term backward-looking center to refer to the entity in the current utterance that refers to the previous utterance (Gross et al. 1983). They suggested that as long as the center of the current utterance is the same as that of the preceding utterance, a pronoun should be used. If a pronoun is not used, this might suggest that a new discourse segment is being introduced. As the focus of conversation changes, the new topic may be introduced explicitly as the theme of a sentence, or it may be selected by reference from a series of things that have already been talked about (i.e., past backward-looking centers). Which of the possible backward-looking centers is selected depends on their ordering, which is dominated by recency.

**Figure 9.11.** A series of snapshots of a stack model of the shift in focus of the weather discourse. Two new topics are introduced. The last utterance pops the top topic Hawaii to refer back to Seattle.