Speech Acts

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 started by considering **performative** utterances, for which the act of speaking itself constitutes the action, such as:

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

As the first two utterances indicate, these performative utterances are often of a legal or religious nature in which the talker has been granted a higher authority to speak for a society rather than merely speaking as an individual. Austin extended the notion of performative utterances with the concept that many utterances manifest a **force,** or ability to change the state of the world by their very existence, due to their impacts on the beliefs or intentions of the conversants. He identified three senses in which action is performed through an utterance.

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 state among conversants, e.g., the utterance may act as a promise, offer, decree, statement of commitment, or admission of guilt.
3. The **perlocutionary act** is the manner in which the utterance changes the audience by changing their beliefs or causing them to perform some action.

There have been some attempts to associate force with sentence form; some obvious examples include commanding, using the imperative form, and questioning via the interrogative form. But the mapping from force to form becomes problematic for **indirect speech acts** in which the force depends on speaking conventions. "Can you pass the salt?" is usually meant as a request not a question about the other's ability to pick up the salt shaker. In a similar vein, "It's pretty noisy in here" might be uttered as a request to turn down the radio volume.

[Searle 1976] offered a more detailed taxonomy of five classes of illocutionary 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 talker to future actions. Threats are also examples of commissives.
4. **Expressives** create a psychological state rather than causing a physical action. They include welcoming, apologizing, thanking, etc.
5. **Declarations** are the performative statements mentioned earlier such as declaring war or firing an employee.

The notion 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 product 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 1988, 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 some 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., that A does expect B to know the referent for "it" and is 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 1975] 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 it is at the moment of the utterance. This is the underlying theme of all the maxims.

- **The maxim of quality:** Speak the truth; do not speak what you know to be false nor that for which you do not have adequate evidence of truth.
- **The maxim of quantity:** Be direct. Say as much as is required at the current point of the interchange, but do not say more than is required.
- **The maxim of relevance:** Make your utterances relevant and to the point.
- **The maxim of manner:** Be brief and orderly; avoid ambiguity and obscurity.

The cooperative principal defines conversation as an organized, purposeful, and efficient series of spoken exchanges, and the remaining maxims are specific techniques to support this principle. In this framework, conversants work together in a conversation, their utterances are mutually relevant, and they do not speak nonsense or seek to confuse each other.

In reality, language often is not nearly as orderly as these underlying principles seem to imply. However, in many cases when Grice's principles appear to be violated, they unify otherwise disjoint or seemingly purposeless utterances. Consider the following exchanges.

A: Did you feed the cat?
B: There's a pile of feathers at the doorstep.

B appears to violate the maxim of relevance; A wishes to know whether B has fed the cat, yet B talks about a pile of feathers. By assuming that B's reply must be relevant, we can infer that B suspects that the cat has caught a bird, and perhaps that this should substitute for the cat's dinner or that the cat should not be fed as punishment.

DISCOURSE

Discourse refers to multiple utterances over time often by different talkers. Discourse is a broad term covering several distinct areas of language understanding. The utterances in a discourse are not disjointed but rather related or connected. Discourse deals with understanding the purpose of these multiple utterances, including issues such as plans that extend across several utterances, references in one utterance to objects or concepts specified in a prior utterance, and the use of multiple utterances to comprise a single speech act. Discourse issues in conversations include the use of language to regulate the conversation's flow, the temporal distribution of contributions among its participants, and the coordination of the effective exchange of information among the participants such that all arrive at the same understanding of what was said.

Discourse issues have formed a theme of conversational interaction and feedback throughout this book. This section emphasizes how conversation is broken

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 invite discourse and most utterances occur in a discourse context [Goffman 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 talker spoke and the second always picked 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 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 1972] analyzed a number of conversations and suggested the following indicators in addition to the completion of a syntactic unit of subject and predicate by which the talker can signal the end of a turn.

- **Intonation:** A level or falling pitch at the end of a sentence indicates termination of the thought being expressed.
- **Syllable lengthening:** The final syllable, or more correctly the final stressed syllable at the end of a turn, is longer than it would be otherwise. Duncan refers to this as "drawl."
- **Gesture:** Termination of a hand gesture while speaking acts as a cue that the accompanying turn is drawing to a close.
- **Key phrases:** Certain phrases such as "you know . . ." at the end of a syntactic unit are often spoken at turn termination.
- **Visual attention:** Talkers often avert their gaze from the listener during an utterance so looking back to the listener could cue the end of a turn.

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 accidently, the listener may **interrupt,** i.e., begin speaking before the other has finished. Interruption is usually dealt with effectively in conversation: Often one party will back off and quickly cease speaking, in which case whichever party is then the talker tends to repeat or summarize what was said during the period of overlap to insure that it was heard correctly. During overlap, the conversants have not yet resolved who should have a turn; one party may attempt to assert control by emphasis 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. **Back channels** refer to a number of behaviors whereby the listeners give feedback to the talker [Yngve 1970]. They include paraverbal utterances ("Hmmm," "Uh-huh"), 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 [Kraut *et al.* 1982, Kraut 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 eavesdropper who could never be heard benefited from the listener's back channel utterances but not as much as the listener 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.

Discourse Focus

Back channels are just one aspect of collaborative behavior in conversation. In the course of speaking conversants 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 hiking 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?

[Grosz and Sidner 1986] proposed a discourse model differentiating the **attentional 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 attentional 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 course of 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 longingly, "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. Grosz, Joshi, and Weinstein use the term **backward-looking center** to refer to the entity in the current utterance that refers to the previous utterance [Grosz *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.

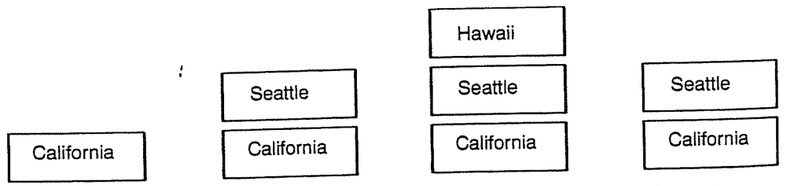

Figure 9.11. A series of snapshots of a stack model of the shift in focus of the weather discourse. Two new topics are introduced. The last utterance pops the top topic Hawaii to refer back to Seattle.

Intonation is a strong cue to shifts in focus. New information, i.e., potential new centers, are usually stressed more than old information. Intonational cues can also be used to override the default ordering of centers.

1: John hit Bill and then he called the police.
2: John hit Bill and then HE called the police.

In sentence one John calls the police; in sentence two Bill does. Ordinarily we would expect that "he" refers back to John, but if this word is emphasized as in sentence two it indicates to choose an alternate center, i.e., Bill.

Focus management requires the agreement of all participants in a conversation—if one party introduces a new topic without the appropriate cues, then other conversants do not receive the intended message. Conversation is a cooperative process. Another aspect of collaborative conversational behavior relates to synchronizing mutual beliefs. Clark *et al.* identify a number of conversational moves to make sure that both conversants agree to the identity of the current focus [Clark and Marshall 1981, Clark and Schaefer 1989, Clark and Wilkes-Gibbs 1986]. Clark and Brennan [Clark and Brennan 1990] introduce the concept of **grounding** to describe the process whereby we ensure that the listener understands our utterances as we intend them to be understood and that we agree that they stand for the same entities in the world.

Although much of conversation is purposeful and can be categorized as an attempt at communicating a concept or performing an action, conversation also serves a social nature. Part of the feedback provided by back channels, for example, informs the talker "I am paying attention to you. I am listening. I value your contribution." Sometimes talk exists as much to fill a communication space between people who feel awkward in silence as it does to change the other conversant's opinion or affect changes in the physical world [Goffman 1981].

CASE STUDIES

This section presents case studies of several projects that attempted to maintain interactive conversations utilizing the aspects of higher-level linguistic knowledge described in this chapter. Although only fragmentary use was made of the formalisms just described, these case studies offer some evidence of the potential of the topics in this chapter for enabling more sophisticated conversational interaction. Syntactic and semantic knowledge can be used to detect speech recognition errors and to determine the meaning of an utterance. Pragmatics relates an utterance to the larger world situation, and discourse structure helps cue appropriate responses.

Grunt

Grunt was an experiment that explored the utility of a discourse model in isolation from other linguistic knowledge [Schmandt 1988]. Grunt attempted to main-