

## CHAPTER 6

# *Psycholinguistics*

HERBERT H. CLARK AND MIJA M. VAN DER WEGE

Psycholinguistics is the study of the processes by which people use language. In conversation, people engage in actions that range from producing and interpreting speech to steering the course of the conversation—determining what topics are taken up when. In reading, people apply many of the same processes, but by using a skill that has taken years to learn. In writing, authors compose, edit, and rewrite to engineer just the right experience for their readers. When we think of language use, we tend to focus on words, phrases, and sentences, but these are often parts of composite actions that include pointing and other gestures as well.

Psycholinguistics was launched in 1900 with the publication of Wilhelm Wundt's *Die Sprache (Language)* as the first two volumes of his monumental *Völkerpsychologie*. Wundt's enterprise was broad, and it led to such distinguished works as Karl Bühler's *Sprachtheorie (Language Theory)* in 1934. By the middle of the 20th century, psycholinguistics had run into rough weather and, at least in America, had almost disappeared. In the 1960s, it was revived with Noam Chomsky's (1957, 1965) vision of language and linguistics, where it often got narrowed to the study of the "psychological reality of linguistic structures." By its hundredth birthday, psycholinguistics had matured into a field in its own right.

Modern psycholinguistics is diverse in its perspectives, theories, approaches, and goals. At its center is how people process language—from producing speech sounds and understanding words to participating in discourse. But it also includes first and second language acquisition, aphasia, speech disorders, reading, and many other issues. Unlike many areas of psychology, psycholinguistics has borrowed heavily from other disciplines—linguistics, philosophy, computer science, sociology, and anthropology. It has also drawn upon broad evidence—laboratory experiments, field experiments, linguistic intuitions, computer simulations, large corpora of conversations, clinical case studies, and much more. There is no royal road to knowledge in the study of psycholinguistics.

In this chapter, we focus on the core of psycholinguistics—the elements we believe make it a field. Our goal is not to review the field, but to frame it. It is to describe the foundational issues and principles. We begin with communication (why people use language in the first place), then take up speaking and listening, and finally turn to the mental representations necessary for using language.

### COMMUNICATION

To use language—to speak or listen, to read or write—is to take action (Austin, 1962;

Levinson, 1983; Searle, 1969, 1975b; Sacks, Schegloff, & Jefferson, 1974). People choose to speak or not to speak, and they try, or do not try, to attend to, identify, understand, and react to what others say. Psycholinguistics is about the social and cognitive processes by which people carry out these actions.

### Language Settings

Language gets used in a wide range of settings (Clark, 1996). It arises as spoken language in personal and nonpersonal settings (e.g., face-to-face vs. lectures), institutional settings (courts, church, etc.), fictional settings (movies, plays), and private settings (talking to oneself). It comes as written language in just as many settings—personal letters, newspaper stories, institutional letters and labels, fictional novels and comic strips, and private notes to oneself. With the invention of new communication technologies, there seems to be no end to the settings in which people use language.

The processes that people use in these settings range just as widely as the settings themselves. It is self-evident that speaking and listening are different from writing and reading. Speaking requires the execution of *vocal* sounds, words, and phrases in a tight temporal pacing. Writing, in contrast, requires a *manual* skill, learned over years of training, which can be done at any pace and with as much editing and rewriting as needed. Listening requires the *aural* skill of identifying sounds, words, and phrases as they are produced in time. Reading, in contrast, depends on a *visual* skill, also learned over years of training, which can be done at any pace, with as much rereading as needed.

Speaking, listening, reading, and writing themselves change radically with the setting. Take speaking, for example. On television, news anchors read aloud what is already written. In weddings, the bride and groom repeat

what they are told. In plays, actors recite lines already memorized. But in spontaneous conversation, speakers decide what they want to talk about, plan their own words, and produce them. Managing all three processes—especially while under time pressure—is a delicate act of juggling. *Spontaneous speaking* is clearly different from *reading aloud*, *repeating back*, and *reciting*. But how are they alike, and how are they different? And how do listening, writing, and reading change with the setting?

One setting is basic, and that is *face-to-face conversation* (Clark, 1996; Fillmore, 1981). It is the only setting that is universal to all the world's peoples, about a sixth of whom are illiterate. It is the setting in which all of the world's languages evolved before the spread of literacy. It is the only setting that does not require specialized skills such as reading, writing, or oratory. It is the setting in which children acquire the rudiments of their first language; learning from books and television comes later. Other settings can be viewed as secondary to, or derivative from, face-to-face conversation. People understand what they read, for example, largely by treating printed language as if it were a representation of spoken language.

So, psycholinguistics must account first and foremost for face-to-face conversation. It must go beyond reading aloud, repeating back, and reciting, and understanding this speech. It must account for how people plan, speak, listen, and gesture—how they communicate—in the give-and-take of spontaneous dialogue. Eventually, it must account for all language settings, but these accounts differ from setting to setting.

### Language in Joint Activities

People use language *to do things*. In all but one of the settings we have reviewed, people use language to do things *with others*. Using

language is inherently social, and that is nowhere more evident than in face-to-face conversation—the primary setting. But what is dialogue for? To answer this question, we draw on 30 years of close analysis of spontaneous conversation recorded in a variety of settings (e.g., Atkinson & Heritage, 1984; Button & Lee, 1987; Drew & Heritage, 1992; Sacks et al., 1974; Schegloff, Jefferson, & Sacks, 1977).

Joint activities are activities that two or more people can only carry out by coordinating with each other (see Clark, 1996). Such examples include one person helping another person put on his or her coat; four musicians playing a string quartet; people playing a game of football or chess or poker; a person buying goods from a clerk in a store; two lawyers negotiating a contract; and two people gossiping. The participants in each activity, as distinguished from bystanders, assume particular roles (e.g., dealer vs. players in poker) as they presuppose or establish common goals (e.g., completing the poker game) and even pursue their own private agendas. Joint activities have coordinated beginnings, ends, and subsections, and the participants have conventional and nonconventional procedures for achieving this coordination (e.g., dealing cards, saying “I raise you ten,” etc.). Finally, people often engage in more than one joint activity at the same time or intermittently (e.g., gossiping and eating dinner).

Dialogue is a means of coordinating actions in joint activities. Take this brief exchange at a drug store counter between Alan, a customer, and Beth, the server (Merritt, 1976, p. 324):

- (1) Alan Hi. Do you have uh size C flashlight batteries?  
 Beth Yes, sir.  
 Alan I'll have four please.  
 Beth [turns to get]

The basic joint activity is a business transaction, the purchase of batteries. To succeed,

Alan and Beth must coordinate on its *participants, timing, and content*. (1) Who the participants are gets established when Alan addresses Beth with “hi” and she acknowledges (probably by nodding and meeting Alan’s gaze). (2) The time they start is established when Alan says “hi,” and Beth, with her nod and eye-gaze, agrees. (3) The content of their basic activity—its public goal—gets established in two steps as Alan proposes the purchase of four size-C flashlight batteries, and Beth agrees to the proposal by turning to get them. Each piece of the dialogue is designed to coordinate a piece of the basic joint activity.

Simple as this example is, it illustrates several points. First, people distinguish *basic* from *coordinating* activities. If Alan were asked what he did in the drug store, he would answer, “I bought four batteries,” not “I talked to the server” (even though he did). The purchase was primary, and the talk was only secondary—in support of the purchase. Second, people coordinate on basic activities *in increments*. Alan and Beth first establish the participants and starting time (“Hi” plus the nod), then a prerequisite for Alan’s order (“Do you have size C flashlight batteries?” plus “Yes, sir”), and then Alan’s order proper (“I’ll have four please” plus her turning away). Third, the participants’ actions depend turn-by-turn on the actions of the other participants. Beth, for example, could have refused Alan’s “Hi” with “Uh, wait a minute” or “Sorry, I’m busy.” Or she could have said “No, sir” instead of “Yes, sir,” and Alan would have followed up with another direction. These features are characteristic of joint activities.

### Joint Projects

Each increment to a joint activity takes coordination. Alan cannot advance his business with Beth without her agreement, and vice versa, and that normally requires actions from both.

A common way to reach agreement is via *adjacency pairs*, as shown here:

1. Alan Do you have uh size C flashlight batteries?
2. Beth Yes, sir.

An adjacency pair consists of two utterances, by two speakers, in which the first utterance is of a type (e.g., a question) that makes an utterance of a second type (e.g., an answer) conditionally relevant as the next utterance (Schegloff & Sacks, 1973). Once Alan has asked his question, it is conditionally relevant for Beth to answer it. Adjacency pairs must be *spoken* communicative acts, so if Beth had *nodded* instead of saying "yes," that would no longer be an adjacency pair. The following pair of actions would not be an adjacency pair either:

1. Alan I'll have four please.
2. Beth [turns to get]

In this chapter we use the term *projective pair* to cover both spoken and nonspeken pairs of conditionally relevant actions.

A projective pair is really a *minimal joint project* (Clark, 1996). When Alan says, "Do you have uh size C flashlight batteries?" he *proposes*, or projects, a joint action for Beth and him to carry out: She is to tell him whether she has size C flashlight batteries. When Beth says, "Yes, sir," she *takes up* Alan's proposal and tells him what he wants to know, completing her part of the projected joint action. The result is a minimal joint project with two parts:

1. Proposal A proposes a joint project for A and B
2. Uptake B takes up A's proposal

This schema also provides a rationale for Alan and Beth's second pair of actions—"I'll have four please" and "[turns to get]."

Minimal joint projects come in great variety. Here are examples from a single telephone

call from Jane to Kate (see Clark, 1994):

| Joint project     | Example   |
|-------------------|---|
| 1. Summons        | Jane (rings telephone)  |
| 2. Response       | Kate Miss Pink's office   |
| 1. Greetings      | Kate hello  |
| 2. Greetings      | Jane hello  |
| 1. Question       | Kate who is it?   |
| 2. Answer         | Jane oh it's Professor Worth's secretary, from Pan-American College |
| 1. Assertion      | Jane oh it's Professor Worth's secretary, from Pan-American College |
| 2. Assent         | Kate m  |
| 1. Request        | Jane could you give her a message for me                            |
| 2. Promise        | Kate certainly  |
| 1. Promise        | Kate I'll tell her  |
| 2. Acknowledgment | Jane thank you  |
| 1. Thanks         | Kate thank you very much indeed                                     |
| 2. Acknowledgment | Jane right  |
| 1. Good-bye       | Kate bye bye  |
| 2. Good-bye       | Jane bye  |

There are many other types as well.

People can create larger joint projects by combining minimal ones, and there are three main ways of achieving this—chaining, embedding, and pre-sequencing.

1. *Chaining* is illustrated in the telephone call from Jane to Kate in these three turns:

|                        |      |  |
|------------------------|------|--|
| Question 1             | Kate | who is it?   |
| Uptake 1 = Assertion 2 | Jane | oh it's Professor Worth's secretary, from Pan-American College |
| Uptake 2               | Kate | m  |

The first two turns constitute one minimal joint project—a proposal (Kate's question) plus its uptake (Jane's answer). But Jane's answer itself initiates a second joint project. She proposes that Kate assent to her claim of being Professor Worth's secretary, and Kate takes

her up with "m" ("yes"). So Jane's utterance is both the *uptake* in one joint project and the *proposal* of a second, linking the two joint projects together in a chain.

2. *Embedding* is illustrated in this exchange between Susan, a waitress, and Jean, a customer (Merritt, 1976):

Question 1 Susan What'll ya have girls?

Question 2 Jean What's the soup of the day?

Uptake 2 Susan Clam chowder

Uptake 1 Jean I'll have a bowl of clam chowder and a salad with Russian dressing

When Susan asks, "What'll ya have girls?" she projects an answer such as "I'll have a ham sandwich." But Jean doesn't have enough information to answer, so she initiates a second sequence with "What's the soup of the day?" The result is one minimal joint project (question 2 + uptake 2) embedded within another (question 1 + uptake 1). The embedded sequence is called a *side sequence* (Jefferson, 1972) or *insertion sequence* (Schegloff, 1972).

3. *Pre-sequencing* is illustrated in an exchange we have already examined:

|                          |      |   |
|--------------------------|------|---|
| Question 1 = Pre-request | Alan | Do you have uh size C flashlight batteries? |
| Uptake 1                 | Beth | Yes, sir.                                   |
| Request 2                | Alan | I'll have four please.                      |
| Uptake 2                 | Beth | {turns to get}                              |

When Alan asks, "Do you have uh size C flashlight batteries?" he is projecting a local answer of yes or no. At the same time, he is *pre-figuring*, or projecting, a second exchange in which he will request some of those batteries (see Schegloff, 1980). Alan's first utterance is taken to be not only a question, but also a *pre-request*. Indeed, the pre-request may be

taken up with an offer, as here:

|                          |            |                                     |
|--------------------------|------------|-------------------------------------|
| Question 1 = Pre-request | 2 Customer | Do you have the pecan danish today? |
| Uptake 1                 | Server     | Yes we do.                          |
| Uptake 2 = Offer         | 3          | Would you like one of those?        |
| Uptake 3 = Request       | 4 Customer | Yes, please.                        |
| Uptake 4                 | Server     | {turns to get}                      |

The entire sequence may get compressed into two turns, as in this phone call to a liquor store:

|                          |           |   |
|--------------------------|-----------|---|
| Question 1 = Pre-request | 2 Susan   | Do you have a price on a fifth of Jim Beam? |
| Uptake 1                 | Manager   | Yes, I do.                                  |
| Uptake 2 = Offer         | 3 Manager | It's five dollars and fifty-nine cents.     |

It may be compressed even further:

|                          |         |                                      |
|--------------------------|---------|--------------------------------------|
| Question 1 = Pre-request | 2 Susan | Can you tell me what time you close? |
| Uptake 1 = Uptake 2      | Manager | Nine.                                |

Not only are there pre-requests, but pre-questions ("Can I ask you something?"), pre-announcements ("Did you hear what happened?" or "You know what?"), pre-narratives ("Did you hear the joke about the three Irishmen?"), and other pre-sequences.

It takes the strategic use of chaining, embedding, and pre-sequencing to navigate larger joint activities. Pre-sequences, for example, can be used to project subsections of a joint activity—jokes, announcements, request sequences, and more. They can also be used to project entire joint activities. When Jane rings Miss Pink's telephone, she is proposing not just a local acknowledgement, but an entire conversation, and when Kate answers "Miss Pink's office," she takes up both proposals at once. When Alan says "Hi" to Beth, he is proposing not only a greeting, but also a business transaction, and when she takes him up on it, she agrees to both. So, although conversations work turn-by-turn (Sacks et al.,

1974), the participants use these strategies for projecting broader joint activities.

### Speech Acts

People appear to create dialogues one utterance at a time. By tradition, these utterances are called *speech acts*—acts performed in speaking. The philosopher John Austin (1962) introduced this idea and distinguished among several types of speech acts. When Alan says, "I'll have four please," he performs four speech acts (among others):

1. The *phonetic act* of making the sounds in "I'll have four please";
2. The *utterance act* of producing a token of the sentence "I'll have four please";
3. The *illocutionary act* of ordering four batteries from Beth;
4. The *perlocutionary act* of trying to get Beth to sell him four batteries.

The very term "speech act" focuses on speakers and speaking—as if listeners and understanding were incidental. And most of those who followed Austin have focused on illocutionary acts, even though the other levels are also important.

Everyday illocutionary acts can be classified by their public point or purpose. According to one proposal (Bach & Harnish, 1979; Searle, 1969, 1975b), they fall into four major categories:

1. *Assertives*. The point of an assertive is to get addressees to accept or reactivate a

to get Kate to accept her assertion that she is Worth's secretary.

2. *Directives*. The point of a directive is to get addressees to do things. When Alan says, "Do you have size C flashlight batteries," he is trying to get Beth to tell him something. Directives include questions, requests, orders, commands, and even hints.
3. *Commissives*. The point of a commissive is to commit the speaker to a future action. When Jane says, "I'll tell her," she is committing herself to giving Kate's message to Miss Pink. Commissives include promises, offers, and other actions.
4. *Expressives*. The point of an expressive is to express a certain feeling to addressees. When Kate says, "Thank you," she is expressing gratitude to Jane. Expressives also include greetings ("hi"), farewells ("bye"), apologies ("sorry"), and congratulations.

A fifth category, called *declarations*, is a specialized class performed by speakers in their official roles in social institutions. Examples include a judge sentencing a prisoner, a referee saying "foul" in a tennis match, or a poker player saying "I raise you five."

Viewed this way, illocutionary acts are best classified by their role in minimal joint projects. Alan's utterance of "Do you have size C flashlight batteries" is a question, a type of directive, because it projects an answer as uptake. Other illocutionary acts project other types of uptake:

| Type of Act    | A's Proposal                          | B's Projected Uptake        |
|----------------|---------------------------------------|-----------------------------|
| 1. Assertives  | A expresses a belief for B to accept  | B accepts A's belief        |
| 2. Directives  | A directs B to do an act              | B commits to doing that act |
| 3. Commissives | A commits to doing an act for B       | B accepts A's commitment    |
| 4. Expressives | A expresses a feeling for B to accept | B accepts A's feeling       |

certain belief. When Jane says, "Oh it's Professor Worth's secretary," she is trying

That is, speakers use illocutionary acts to perform *perlocutionary acts*, by which they try to

get addressees to take on obligations (as with directives) or to accept the speakers' beliefs, commitments, or feelings (as with assertives, commissives, and expressives). Speakers normally expect addressees to complete the process with their uptake. People in conversation engineer these social exchanges—the acceptance of beliefs, commitments, feelings, and obligations—so as to coordinate their basic joint activities.

If illocutionary acts are partly defined by their role in minimal joint projects, then addressees may help determine how they are to be classified. When a woman named Susan called up restaurants and asked, "Do you accept credit cards?", she got the first answer 40% of the time and the second 14% of the time (Clark, 1979):

- (2) Susan            Do you accept credit cards?  
 Manager A        Yes, we do.
- (3) Susan            Do you accept credit cards?  
 Manager B        We accept MasterCard and Visa.

In 2, managers construed Susan as asking a yes/no question, but in 3, managers construed her as requesting a list of credit cards. In effect, she left the interpretation up to the managers, because she couldn't correct them to the opposite interpretation ("No, I mean. . .") without offending them. What she was taken to mean—the illocutionary act she was construed as performing—was determined not just by *her* words, but by the *manager's* uptake. This conclusion may seem paradoxical (how can what speakers are taken to mean be shaped by their addressees?), but it falls neatly out of the view of language use as joint action.

Traditionally, questions such as "Do you accept credit cards?" and "Can you tell me what time you close?" have been called *indirect requests* (Gordon & Lakoff, 1971; Searle, 1975a). In this view, when Susan asks, "Do you accept credit cards?" (literally, a yes/no question), she is *indirectly requesting* a list of credit cards. The assumption is that Susan has

a specific interpretation in mind, and it is up to the manager to recognize it. The problem with such a view is that it leaves no role for the manager.

Indirect speech acts are better viewed as pre-sequences (Gibbs & Mueller, 1988; Schegloff, 1988). When Susan asks, "Do you accept credit cards?" she is *initiating a negotiation* about what she is to be taken to mean, and it takes the manager to complete the negotiation. The manager can reply, "Yes, we do," as in 2, and let Susan initiate the next step in the negotiation with "Which ones?" The manager can also shortcut the process by *offering* the information he or she believes Susan will ask for, "We accept MasterCard and Visa," as in 3. Finally, the manager can answer her question *and* shortcut the process, as in "Yes, we accept MasterCard and Visa" (which managers did 33% of the time). To succeed, managers must try to infer Susan's larger plans, and they clearly did. One manager replied, "Uh, yes, we accept credit cards. But tonight we are closed." Another replied, "Uh-uh. We're not open anyways." Both inferred that she intended to eat at the restaurant that night.

Some pre-requests are so conventional that they don't seem to allow such a negotiation. It seems impossible to treat "Can you tell me the time?" or "Do you have the time?" merely as yes/no questions. Yet addressees do have options. When Susan asked other businesses, "Can you tell me what time you close?" some managers replied, "Six," but others replied, "Yes, at six," treating the yes/no question explicitly. "Yes, at six" is heard as more polite because it explicitly deals with both the yes/no question and projected request (Clark & Schunk, 1980). Uptake plays a role in even the most conventional pre-sequences.

### Common Ground

Joint activities are carried out against the participants' *common ground*. Common ground

refers to participants' *mutual* knowledge, beliefs, assumptions, and awareness (Clark, 1996; Clark & Marshall, 1981; Lewis, 1969; Stalnaker, 1978). There are two main types of common ground: *communal* common ground and *personal* common ground.

Communal common ground is based on the communities that people belong to. Suppose Kenneth and Jane meet and establish that they both speak English, live in San Francisco, and play classical piano. English speakers, San Franciscans, and classical pianists are three *communities of shared expertise*, and we all belong to many such communities. The expertise of a community may be based on nationality, residence, education, occupation, employment, hobby, language, religion, politics, ethnicity, club, subculture, cohort, or gender. Once Kenneth and Jane establish joint membership in a community, they can take as common ground all the expertise that people in these communities take for granted. As English speakers, they can presuppose basic English vocabulary and grammar. As San Franciscans, they can presuppose the geography, names, and politics of San Francisco. As classical pianists, they can presuppose classical composers, techniques of playing, and musical genres.

Personal common ground is based instead on the personal experiences people have

Joint activities are governed by the participants' common ground. When Alan buys batteries from Beth at the drug store counter, the two of them start with a large body of presuppositions—their *initial common ground*. They presuppose that they are clerk and customer at a drug store counter, that certain practices hold at Philadelphia drug store counters, and that they both speak English. They may be wrong, but that is what they presuppose (see Fussell & Krauss, 1992). As they proceed, they take actions to add to that common ground. They try to update the *current state of their activity*—what they have committed to so far and what is left to do. In their first exchange, Alan and Beth establish *as common ground* that the store sells size C batteries, and in their second, that Alan is committed to buying four. Joint activities would fail without the orderly maintenance of common ground.

### Grounding

Using language is itself a joint activity. When Alan speaks to Beth, the two of them must establish (a) that she is attending to him, (b) that she is identifying his words and gestures, (c) that she understands what he means, and (d) that she is considering taking him up. In general, two people, A and B, have to coordinate their actions at four levels (Clark, 1996; Paek, 2000):

| Level        | A's Action                     | B's Action                        |
|--------------|--------------------------------|-----------------------------------|
| 1. Channel   | A makes sounds, gestures for B | B attends to A's sounds, gestures |
| 2. Signal    | A produces a signal for B      | B identifies A's signal           |
| 3. Intention | A means something for B        | B understands what A means        |
| 4. Project   | A proposes a joint project     | B considers the joint project     |

shared with each other. At the drug store counter, Alan and Beth perceive each other standing there, looking at objects, and hearing the cash register work. They also talk, point, and hand things to each other. Personal common ground is built up from joint perceptual experiences and joint communicative actions.

Indeed, the two of them try to establish, as common ground, the belief that they have succeeded at each of these levels well enough for current purposes, a process called *grounding* (Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). In conversation, people ordinarily try to ground



everything that gets said. They realize, tacitly, that a minor misunderstanding or mishearing now may lead to greater troubles later.

How people ground varies with the level. At the channel level, Alan and Beth may exchange eye gaze as evidence of joint attention (Goodwin, 1981; Kendon, 1967). At the signal and intention levels, Alan looks for positive evidence from Beth, and she tries to provide it. One type of evidence is Beth's uptake, as here:

Alan Do you have uh size C flashlight batteries?  
Beth Yes, sir.

When she replies "Yes, sir," she provides evidence not only that she has attended to Alan's utterance, but that she believes she has identified and understood it. She also shows that she has construed it as a yes/no question. Alan accepts all this evidence by going on to say, "I'll have four please." If she had responded, "My name is Beth," he would have evidence of a *failure* to understand, and he might repeat his question. Other times Beth can assert her understanding with acknowledgements, or *continuers*, like "uh huh," "yes," and "mhm," often called *back-channel responses* (Schegloff, 1982; Yngve, 1970).

Grounding is a two-way process, and addressees often initiate repairs when they fail at the channel, signal, intention, or project level. A common strategy is for addressees to initiate side sequences, as here:

(4) Arthur can I speak to Jim Johnstone please?

Barbara senior?

Arthur yes.

Barbara yes - - -

Although Barbara identified Arthur's utterance, she doesn't understand to which Jim Johnstone he is referring. She implies all this by presupposing success at the channel and

signal levels and asking specifically about Johnstone's identity ("senior?"). Only once Arthur has said "yes" is she willing to go on to her answer. If she hadn't succeeded at the channel level, she might have asked, "What?" and Arthur would have repeated the question.

Grounding is carried out by and for the speaker and addressees, and that does not guarantee success for *overhearers*. In one experiment (Schober & Clark, 1989), two people, whom we will call Ann and Ben, conversed freely as Ann got Ben to arrange 12 Tangram figures (abstract, block-like depictions of people) in a particular order. A third person, whom we will call Oscar, sat nearby but wasn't allowed to speak, and also tried to arrange the 12 figures in that order. That made Oscar an overhearer. The three were separated by barriers, unable to see each other, and all began as strangers. The figures were not easy to describe. In one case, Ann began, "Then number twelve, is (laughs) looks like a, a dancer or something really weird. Um . and, has a square head." Ann and Ben then took several turns to ground that description, often using information that Ben presented (e.g., "and a big fat leg?"). Ben was much more accurate than Oscar in arranging the figures. He made errors 5% of the time, whereas Oscar made errors 22% of the time. Why was Oscar so bad? When Ann and Ben grounded their descriptions, they were opportunistic in using information they happened to share. That often left Oscar in the dark.

In summary, there would be no language as we know it if people didn't engage in joint activities. For two people to play cards, or move a table, or transact business, they need to coordinate their individual actions, and they use dialogue to do that. They use projective pairs to carry out minimal joint projects, which they combine via chaining, embedding, and pre-sequencing, to create larger joint projects. The act of communication is itself a

joint activity, and to coordinate that requires grounding.

## SPEAKING

In face-to-face settings, the current speaker produces words and gestures while the others try to attend to, identify, understand, and consider them. Although the two processes of speaking and listening are not autonomous, researchers have traditionally investigated them separately. Here we consider speaking, or how speakers work their way, as Levelt (1989) put it, "from intention to articulation." The three main steps *en route* are *conceptualizing* what to say, *formulating* how to say it, and *articulating* the result.

In speaking, speakers begin with at least some idea of what they want to do at the moment. When Jane is asked, "Who is it?" she must decide, "Do I want to say who I am, and if I do, how do I want to identify myself?" Speakers normally begin with incomplete plans, and they often change their minds mid-utterance.

Speakers cannot express just anything. What they decide to say (their conceptualizations) must be expressible in the language they are speaking, and they must be able to formulate the right expressions in time. In English, the ideas of motion and manner can be expressed in a single verb, *run* ("go fast"), but in Japanese they must be expressed in two words. Speakers of English and Japanese must conceptualize what they say with these targets in mind and then follow through with the right formulation (Slobin, 1996).

Speakers must also coordinate their actions with their addressees. They often signal delays, describe mistakes, prolong words, and hedge expressions that do not quite fit—all, apparently, to help addressees attend to, identify, and understand their speech and gestures. Speakers devote a part of speaking to managing the process of communicating itself.

## Planning Units

Speakers cannot formulate an expression or gesture without some plan. But where do these plans come from? And what are the plans about? If people use dialogue for advancing basic joint activities, their plans must derive, in part, from these activities.

The major units of planning are easiest to illustrate in narratives. In a study by Chafe (1980), people were shown a short movie, without dialogue, about farm workers picking pears in an orchard and were then asked to describe what happened. The following is an excerpt from one narrative:

- |     |     |                 |   |
|-----|-----|-----------------|---|
| (5) | (a) | (.85) A—nd      | he (.35) sees this three pear<br>(.20) these three baskets<br>of pears, |
|     | (b) | (.15) and then  | sees this man up in the<br>(.50) tree,                                  |
|     | (c) | and             | decides (.45) that he'd like<br>some pears.                             |
|     | (d) | And             | at first looks like he's going<br>to take one or two.                   |
|     | (e) | (.60) Then      | decides that he'd (.15) much<br>rather take a whole<br>basket,          |
|     | (f) | (.55)           | puts the basket on the bike,  |
|     | (g) | (.90) tsk a—nd  | kind of struggles..   |
|     | (h) |                 | cause it's much too big<br>for him.                                     |
|     | (i) | And             | the bike is much too big<br>for him.                                    |
|     | (j) | (1.85) The—n    | he's riding .. across this ..<br>great (.25) expanse,                   |
|     | (k) | (.2) and (1.15) | a girl comes, [continues]   |

Pauses (in seconds) are marked in parentheses ["(.60)"], slight breaks in tempo by double periods (".."), and prolonged words by dashes ("a—nd"). Narratives like this show evidence of three levels of planning.

## Intonation Units

An important level of planning is the intonation unit, represented by each line of the excerpt. As the name suggests, an intonation unit

has a single intonation contour, or melody, with a distinctive ending such as a rising or falling pitch. There is good evidence that these are units of planning. They must be planned as a whole for their intonation contours to come out right. And they often have *entry problems*. Speakers need extra planning time before starting them, and often reformulate them before continuing fluently. In Chafe's (1980) pear stories, 88% of the intonation units were preceded by pauses that averaged 1 second in length. Many also had false starts, prolonged words, and other disfluencies at or near their beginnings.

Intonation units also represent unified *conceptual* plans. They tend to be single, finite clauses, that is, clauses with verbs that have tense, as in lines a, d, h, i, j, and k. When they are not finite clauses, they are usually constituents of a clause, such as the predicate phrases in lines b, c, e, f, and g. In narratives, they are often introduced with *and*, *and then*, *but*, or *so*, signaling a continuation of the story. In the pear stories, intonation units averaged six words long and lasted an average of 2 seconds. They represent what Chafe (1980) called *idea units*—single events or focal points in the larger event being described.

### *Sentential Units*

These consist of one or more intonation units (an average of four in the pear stories) that end with a terminal contour reserved for sentences. In the previous excerpt, intonation units are marked with commas, and sentential units with periods. Lines a to c represent one sentential unit, line d another, lines e to h a third, and so on. Unlike intonation units, sentential units vary enormously in length. Conceptually, they appear to represent a single center of interest in the larger event being described (Chafe, 1979, 1980).

### *Sections*

Sentential units, in turn, are strung together to create sections, which correspond roughly to

written paragraphs (Chafe, 1979; Gee, 1986). Sections are defined in part by their prosody: They tend to begin at a higher pitch and end with a falling-pitch glide. In our excerpt, one section begins at line a, and another at line j. Sections require even more planning than intonation units, for they display more severe entry problems. Line a, for example, begins with longer and more frequent disfluencies, “(.85) A—and (.15) he (.35) sees,” than all the other lines in the section, and so does line j, “(1.85) The—n (.2) he’s.”

Sections represent another level of conceptualization. In narratives, sections have a single topic or theme that reflects a single place, time, and set of characters, and they begin at discontinuities in the event being described. The sentential and intonation units they contain tend to fall into parallel structures. Narrators cannot plan the whole narrative beforehand, so they must keep track of where they are as they create each section, sentential unit, and intonation unit.

Unlike narratives, dialogues are created when, by the participants working together, so many of their plans are local. In Jane's telephone call to Kate, illustrated earlier, Jane asks, “Could you give her a message for me?” and once Kate decides to comply, she plans “Certainly” and produces it. She cannot plan “Certainly” *until* she has understood Jane's request. Yet local plans are part of larger plans. Kate's local plan to comply with Jane's request is part of her larger plan to pass information to Miss Pink. When Alan says “Hi” to Beth at the drug store counter, his local plan is to greet her, but only as the initial move in his larger plan to buy batteries. Most local plans are derived from larger joint activities.

Intonation units, sentential units, and sections are planning units even in dialogues. As noted earlier, people in conversation proceed largely by means of projective pairs. These proposals and uptakes each normally occupy single turns and are often sentential units of one or more intonation units (Ford &

Thompson, 1996). Jane's "Could you give her a message for me?" and Kate's "Certainly" are each single intonation units. Together, a proposal and its uptake constitute a type of section, and these can be combined, through chaining, embedding, and pre-sequencing, to form larger sections. So, although the planning units in dialogues look much like those in narratives, they emerge from the participants acting together.

### Perspective

People taking part in joint activities must coordinate the *content* of their actions—the ideas, beliefs, and assumptions they are presenting. Speakers initiate this process by their choice of words, phrases, clauses, and gestures. Many of these choices have to do with perspective, broadly defined.

Suppose that Burton, who is speaking to Charlotte, wants to describe a scene in which a bartender filled a glass with beer. Here are some of his options:

| Example   | Nominal arguments      |
|---|------------------------|
| (6) a. The bartender filled the glass with beer.    | bartender, glass, beer |
| b. The glass was filled with beer by the bartender. |                        |
| c. The bartender filled the glass.                  | bartender, glass       |
| d. The glass was filled by the bartender.           |                        |
| e. The glass was filled with beer.                  | beer, glass            |
| f. The glass filled with beer.                      |                        |
| g. The beer filled the glass.                       |                        |
| h. The glass was filled.                            | glass                  |
| i. The glass filled.                                |                        |

#### 1. Propositions

Burton must choose the propositions he wishes to express. To form a clause, he must include a verb. As arguments of that verb he can include three arguments (as in lines a and b), two arguments (as in lines c through g), or

just one (as in lines h and i). With two arguments, he can mention the bartender and glass (as in lines c and d), or the beer and glass (as in lines e through g). Even without mentioning the bartender, he can *imply* the presence of an agent (as in lines e and h) or not (as in lines f, g, and i).

The propositions expressed are also determined by *word choice*. Instead of *bartender*, Burton could have used *barman* or *guy behind the bar*. Instead of *beer*, he could have used *brew* or *lager* or *suds*. Instead of *glass*, he could have used *stein* or *schooner*. Or he could have added modifiers, as in *tall glass*, *very dirty glass*, or *glass with a picture of the President on it*. Each choice reflects a different perspective.

Burton must get Charlotte to understand his perspective. Recall the experiment described earlier in which Ann got Ben to arrange 12 Tangram figures in an order (Schober & Clark, 1989). Ann and Ben repeated the task six times with new arrangements of the same figures. The first time through, it took them many words (112 on average) to establish a jointly acceptable perspective, as in this example:

- (7) Ann All right, the next one looks like a person who's ice skating, except they're sticking two arms out in front.  
 Ben Uh huh, okay.  
 Ann Got that one?  
 Ben Yeah.

Here Ann and Ben agreed on the description "person who's ice skating, except they're sticking two arms out in front." (Another pair agreed on the description "person dancing" for the same figure.) By the sixth time through, it took Ann and Ben only 16 words on average, as shown here:

- (8) Ann The ice skater  
 Ben M-hm.

Ann simplified the perspective to "ice skater" based on the perspective she and Ben had

grounded earlier (see also Chantraine & Hupet, 1994; Hupet, Seron, & Chantraine, 1991; Krauss & Weinheimer, 1964, 1966, 1967). When Ann was given a new partner, Carl, she had to return to a fuller perspective and ground it from the beginning (Wilkes-Gibbs & Clark, 1992):

- (9) Ann All right, the second one looks like a person that's ice skating, kind of. They've got a diamond for a head and then they've got two arms sticking out to the right and a leg in back, and a leg-
- Carl To the right or to the left? To the-
- Ann To the left, sorry.
- Carl I got it.

## 2. Subject and Predicate

Even if Burton mentions the bartender and glass, he must decide which to make the subject. In 6c, the bartender is the subject, and what the bartender did is the predicate, but in 6d, the glass is the subject, and what happened to it is the predicate. Many languages also mark a *topic* (what the utterance is about) and *comment* (what is said about the topic), but English does not. Normally, the topic in English is the subject.

## 3. Figure and Ground

Burton can also choose between saying "The glass filled with beer" and "The beer filled the glass" (6f and g). In the first, he views the glass *with respect to* the beer, treating the glass as figure and the beer as ground. In the second, he does the reverse. Using another verb, he has the same choice with (a) "The bartender filled the glass with beer" versus (a') "The bartender poured beer into the glass" (see Talmy, 2000).

## 4. Given and New Information

Burton is exquisitely sensitive to Charlotte's state of mind. As we noted earlier, he keeps track of their current common ground and designs his utterance against it. Most utterances divide into *given information*, which refers to information inferable from current common

ground, and *new information*, which refers to information not yet part of common ground (Clark & Haviland, 1977; Prince, 1981). Consider Burton's two choices:

- (6) j. What the bartender did was bartender, glass  
fill the glass.
- k. The bartender filled a glass. bartender, glass

With 6j, Burton assumes that it is a given—already common ground—that the bartender did *something*, but not *what* it was. He adds the new information that it was "fill the glass." With the accent on *bartender* in 6k, he assumes that it is given that *someone* filled the glass, but not *who* it was. He adds the new information that it was the bartender. Burton's choice of given and new information determines not only the syntax of his utterance, but its intonation.

One choice that depends on common ground is the choice between *definite* and *indefinite descriptions*. Burton would tell Charlotte, "The bartender is filling *the glass* for you" if he thought she could infer the identity of the glass from their common ground (e.g., he had just given her glass to the bartender). But he would reply, "The bartender is filling *a glass* for you," if he thought she could not infer its identity. The general rule is this: Definite descriptions require the referents to be inferable from current common ground; indefinite descriptions do not. Therefore, Burton can say, "I got in my car and grabbed *the steering wheel*," and assume Charlotte will infer "The steering wheel belongs to the car." He can also say, "I walked into the room; *the chandeliers* were burning brightly," and she will infer "There are chandeliers in the room." Inferences like these are called *bridging inferences*, and Burton designs his utterance to make bridging easy, a point we return to later.

Speakers choose perspectives as part of local plans. Suppose Charlotte asks, "What is the bartender filling my glass with?" It would

be natural for Burton to reply, "Beer," "With beer," or perhaps, "He's filling it with beer." It would be odd to reply, "It's being filled with beer," or "Beer is filling it," or even "It is filling up with beer." The natural replies *retain* Charlotte's perspective: the propositions, subject and predicate, figure and ground, and given and new information of her question. The other replies *replace* her perspective. In an experiment by Levelt and Kelter (1982), confederates phoned Dutch merchants and asked the Dutch equivalent of lines 10a, b, c, or d:

- (10) a. What time does your shop close?  
 b. At what time does your shop close?  
 c. What time does your shop close, because I have to come into town especially for this, you see?  
 d. At what time does your shop close, because I have to come into town especially for this, you see?

Although the perspectives in 10a and 10b differ only slightly ("What time" vs. "At what time"), the merchants tended to retain that perspective. They preferred "Five" over "At five" for question 10a, and the reverse for 10b. With the extra clause in 10c and 10d, merchants were more likely to give full answers, such as "We close at five," which are appropriate to either perspective. Retaining a perspective is the easiest, and therefore expected, thing to do.

One reason for a respondent to change perspective is to take issue with the speaker, as in this example:

- (11) Jim    how old, were most of the children, -  
 Kay    well uh only a few of them, were  
          children in fact, . um . I was teaching  
          adults,

In changing perspective, Kay implies disagreement with Jim's presupposition about Kay's students. The general rule is this: To retain a perspective is to presuppose agreement; one way to imply disagreement is to change perspective. All in all, selecting the

appropriate perspective is an important part of planning.

### Functional Processing

It was once believed that speakers produce utterances one word at a time by association—from left to right. One of the revolutions of the 20th century was to overturn that idea. It was replaced by the theory that speakers formulate utterances from the whole to its parts, from the top down. Speakers begin with a message—a selection of propositions under a particular perspective, or enough material for about one clause. They then proceed in three overlapping stages (Bock & Levelt, 1994; Levelt, 1989): (a) functional processing, (b) positional processing, and (c) phonological encoding. Much of the evidence for the top-down view comes from a surprising source: slips of the tongue (Dell, 1986; Fromkin, 1971, 1973; Garrett, 1980).

In functional processing, speakers select the lexical concepts needed for their message and assign them to grammatical functions appropriate to their perspective. Suppose Alan wants to tell Barbara that Ben has been offered a job in engineering. For this message Alan needs six lexical concepts, roughly, "the person speaking," "believe," "male person in focus," "officially propose," "technical profession," and "paying position." These lexical concepts, called *lemmas*, are each associated with a word form, or *lexeme*. The lemma "officially propose," for example, is associated with the lexeme *offer*. Once Alan has formulated each lemma, he must retrieve the corresponding lexemes *I, think, he, offer, engineering, and job*.

Many types of slips of the tongue arise at this stage (Bock & Levelt, 1994; Dell, 1986).

1. *Semantic Substitutions*. One speaker produced "the the *Ca* - . the the *Protestants*, seem just as bad at this." He intended to activate the lemma "Protestant religious group,"

but instead activated "Catholic religious group," a closely related lemma. Slips like this lead to the substitution of one semantically related word for another, not only *Catholic* for *Protestant*, but in other examples such as *high* for *low*, *cherries* for *grapes*, and *Chinese* for *Japanese*.

2. *Blends*. Another speaker referred to a container "that they *swishle* swizzle things around in." Apparently, he activated the lemmas for *swish* and *swizzle* simultaneously (both fit his message) and combined the corresponding two lexemes to form the blend *swishle*. Other attested blends include *momentaneous* from *momentary* and *instantaneous*, *stougher* from *stiffer* and *tougher*, and *hilarics* for *hilarity* and *hysterics*.

3. *Sound-related substitutions*. Another speaker said, "because she'd laughed so much she'd *burnt* a couple *burst* a couple of stitches." She selected the lemma for *burst*, but retrieved the sound-related lexeme *burnt* instead. Other attested examples include *sympathy* for *sympphony*, *bodies* for *bottles*, and *garlic* for *gargle* (Bock & Levelt, 1994).

4. *Tip of the tongue*. Another speaker said, "and can you assess can you . *keva*- what's the word, . *connect* them . . ." Apparently, he had selected the lemma for *connect*, but could not retrieve the lexeme by the time he needed it. Hence the initial attempt "keva-" followed by the comment "what's the word." English has special words for use at such moments, as when another speaker said, "you don't mean the Hussey *thingummy* and *whatsit*."

5. *Collocation substitutions*. As it happened, in our example Alan has trouble retrieving the right lexemes:

- (12) Alan I think he was offered an *engineering degree, engineering - job*, after the first slump.

He is trying to retrieve the stock phrase of *engineering job* for the two lemmas "technical profession" and "paying position," but instead retrieves a stock phrase with the same first word, *engineering degree*. Another example

of collocation substitution is that of *chambermaid* for *chamber music*.

Speakers must assign the lemmas and lexemes they select to syntactic functions (Bock & Levelt, 1994). Their message specifies the perspective for the current clause (which propositions are to be expressed, what is subject, object, and indirect object, what is figure and what is ground, what is given and what is new), and these determine the functional assignments. Alan's message specifies two main propositions: "x *thinks* that y," where y is "z *offers* u to v." He assigns *I* to the role x, *he* to the role v, *engineering job* to the role u, and leaves z unspecified. Also, he assigns *I* to the subject of the main clause, and *he* to the subject of the embedded clause. And, Alan determines that *I* and *he* are given information in focus of attention, therefore making them pronouns and that *engineering job* is new information, therefore making it indefinite. The result is a functional assignment something like this: [I think that [he be offered engineering job]].

Certain slips of the tongue can arise in the process of functional assignment:

6. *Word interchanges*. One speaker wanted to say "writing a letter to my mother," but said "writing a *mother* to my *letter*," exchanging the lexemes *mother* and *letter*. Although the speaker retrieved *mother* and *letter*, he assigned them to the wrong arguments. He must have activated both words at the same time to be able to exchange them. Words may also be anticipated, as in "the *sky* is in the sky" (for "the *sun* is in the sky"), or perseverated, as in "the class will be about discussing the *class*" (for "discussing the *test*"). The word substituted almost always has the same form class as the intended word (such as noun for noun, and verb for verb).

7. *Phrasal interchanges*. One speaker wanted to say "I got into a discussion with this guy," but produced "I got into *this guy* with a *discussion*." He exchanged not just two words, however, but two entire phrases, a *discussion*

and *this guy*. He must have planned these phrases before inserting them into their appropriate slots in the construction of "I got into x with y." Another speaker intending "they must be too tight for you" produced "*you* might be too tight for *them*." He must have switched the lemmas "third person plural" and "first person" and only then selected the lexemes *you* and *them*. Otherwise, he would have produced "*you* might be too tight for *they*."

### Positional Processing

Once speakers have selected the lexemes and their functional assignment, they need to order the lexemes for articulation. The first step is to assemble the lexemes, in their assigned functional roles, into constituents. Alan assembles *I* and *think* into one major constituent, and *he*, *offered*, and *engineering job* into another, and he places them in this order. (If he had reversed the order, he would have said, "He was offered an engineering job, I think.") He then adds the right inflections, making "be + past tense" into *was*, and he spells out the function words, making the indefinite article into *an* to agree with *engineering job*.

Speakers at this stage are sensitive to the weight of each constituent. When they have a choice, they prefer to place heavier constituents later than lighter ones (Arnold, Wasow, Losongco, & Ginstrom, 2000; Behaghel, 1909/1910; Hawkins, 1994; Wasow, 1997). Consider this example, noting the order of constituents in brackets:

- (13) the first European conference on astronomy at  
Leicester, reported [yesterday morning], - [on  
overnight observations of the behaviour of the  
object, - . known as A six uhu two one one  
zero], (1.11a.28)

Ordinarily, the speaker would have said, "A reported [on some observations] [yesterday morning]." But she anticipated that her description of the observations would require a heavy constituent (17 words long), so she

placed the lighter constituent "yesterday morning" (2 words long) first.

Many other types of slips of the tongue occur at this stage (Bock & Levelt, 1994; Dell, 1986).

8. *Morpheme interchanges*. One speaker, intending to say "Singer sewing machine," produced "Singing sewer machine." He kept *sing* and *sew* in the right order, but added the inflections *-er* and *-ing* to the wrong stems. Another example is "he go backs to" for "he goes back to."

9. *Morpheme accommodation*. One speaker, intending to say "Mr. Keene, tracer of lost persons," said "Mr. Keene, loser of traced persons." At the functional level, he exchanged, not the words *tracer* and *lost*, but the verb stems *trace* and *lose*. Then, at the positional level, he added *-er* to *trace* to form *tracer* and made *lose* into a past participle to form *lost*. Speakers also select *a* or *an* to fit the word that follows it, even if that word is itself in error, such as the speaker who misproduced "a meeting arathon" for "an eating marathon."

10. *Mis-derivations*. One speaker produced "these are oral contraception," another "I've just gave given you," and another "he think thinks that Ella's worried." These speakers planned the right words, "contracept + nominal suffix," "give + past participle," and "think + singular," but in deriving the words, added the wrong inflections.

### Phonological Encoding

Once speakers have selected the words, assigned them to functional positions, assembled them in the right order, and filled in the inflections and function words, they are ready to spell out the phonetic segments. They do this, not one intonation unit at a time, but one short constituent at a time. Once again, the evidence comes from slips of the tongue.

11. *Sound interchanges*. These include the anticipation of an upcoming sound, as in



*leading list* for *reading list*, the perseveration of a previous sound, as in *beef needle* for *beef needle*, or *Liverpool lullapie* for *Liverpool lullaby*. The classic "spoonerism" is an exchange of two sounds, as in *lork yibrary* for *York Library*, *speer bill* for *spill beer*, and *flow snurries* for *snow flurries*. Speakers can interchange consonants (e.g., *p* and *b*), vowels (e.g., *ee* and *oo*), consonant clusters (e.g., *fl* and *sn*), and what are called the rimes of two syllables (e.g., *-eer* and *-ill*).

Generally, speakers produce more anticipations (*leading list*) than perseverations (*beef needle*). According to a model developed by Dell, Burger, and Svec (1997), this is because speakers are focused more on the future of their speech planning than on the past. When people have to say tongue twisters such as "chef's sooty shoe soles," people tend to perseverate more often than anticipate words or sounds. After practice, while the overall error rate drops, the errors tend to be anticipations rather than perseverations. People who speak more slowly (e.g., children and people with brain damage) also tend to focus more on the past and produce more perseverations.

Sound interchanges work in remarkably regular ways. The two elements involved almost always come from content words (nouns, verbs, adjectives, adverbs) and not function words (articles, prepositions, etc.). They almost always come from adjacent words (as in *York Library*) or even the same word (as in *aminál* for *animal*). They tend to be similar phonetically and metrically, and in homologous parts of words. The *y* and *l* in *York Library* are similar types of consonants—what are called *liquids*—and both are in the initial position of accented syllables. Therefore, sound exchanges stand in contrast to word exchanges. Word exchanges come from homologous locations in *phrases* and are similar in *meaning* and *function*. Sound exchanges come from homologous locations in *words* and are similar in *sound* and *meter*.

Phonological encoding, therefore, works one short phrase at a time. It assembles these phrases according to their phonetic segments, syllables, and meter, regardless of what they mean. And when it makes errors, it makes them out of the elements in these plans. The final product is a motor program that works the tongue, lips, larynx, jaw, and lungs. There is an analogous process that creates a motor program to work the hands, arms, eyes, face, and torso in gestures. Although less is known about this process, it is linked in both time and content to the functional, positional, and phonological processes for speech. Speakers' gestures are closely tied to the content and timing of the words they use (see the following sections).

### Primary and Collateral Speech

People are not automatons. They are normally aware of what they are doing, able to reflect on what they have just done and are about to do, and if they don't like what they see, they change directions. People are no different when they are speaking. They normally *monitor* what they are about to say and have just said, and what their addressees are doing and saying, and if they don't like what they see, they change directions (Levelt, 1983, 1989). Taking actions based on self-awareness adds a second track to utterances. The distinction is between *primary* and *collateral signals* (Clark, 1996).

Spontaneous speech is replete with actions not found in idealized speech. The following is one example (Svartvik & Quirk, 1980):

- (14) Reynard well, . I mean this . uh Mallet said  
 Mallet was uh said something  
 about uh you know he felt it  
 would be a good thing if uhh . if  
 Oscar went, (1.2.370)

This utterance is full of *supplementary features*—repeats ("if uhh if"), repairs ("Mallet said Mallet was"), fillers ("uh"),

prolonged syllables (“uhh”), and editing expressions (“I mean,” “you know”). These actions each appear to reflect a difficulty in deciding what to say or how to say it. Still, they allow Peter, the addressee, to identify what Reynard really wants to say. Conceptually, Reynard’s utterance divides into two parts. The *primary signals* reflect the official business of the conversation at the moment, namely:

- (14') Reynard well, Mallet said he felt it would be  
a good thing if Oscar went

The *collateral signals* are about the on-going performance itself.

Supplementary features typically divide into two types: problems and solutions. Take Reynard’s “it would be a good thing if uhh . . . if Oscar went.” By the time Reynard reached *thing*, he apparently had a problem—perhaps he didn’t quite know what to say next. Peter, his addressee, may have inferred the problem, but the problem itself remained hidden. All Peter heard was Reynard’s *solution* to the problem. Reynard took four actions: (a) Before suspending his speech, he produced *if* to commit himself to producing an *if*-clause; (b) he produced *uh* to signal that he was delaying the resumption of his speech; (c) he prolonged *uh* to signal that he was continuing an ongoing delay; and (d) upon resuming speech, he repeated *if* to restore continuity to the *if*-clause. These actions are each collateral signals to help Peter deal with the delay with the least effort.

Collateral signals come in many types, which have been discovered in the close examination of spontaneous speech. These include the following:

1. *Editing expressions* such as *I mean, you know, that is, no, and sorry* (Erman, 1987; Levelt, 1983, 1989). Speakers use these to point out expressions they wish to amend and why. In “Mallet was uh said something

about uh *you know* he felt . . .” Reynard points out that he is changing “said something about” to the more accurate “felt.”

2. *Fillers* such as *uh* and *um* (Clark, 1994, 1996; Clark & Fox Tree, 2001). Speakers use these to signal delays in speaking. In “Mallet was uh said something about . . .” Reynard signals a delay with “uh” while he rephrases “was . . .” to “said something about.”
3. *Discourse markers* such as *well, now, oh, like, and so* (Fox Tree & Schrock, 1999; Schiffrin, 1987; Schourup, 1982; Underhill, 1988). Speakers use these to indicate changes in direction and other such things. With “well” in example 14, Reynard indicates that he isn’t giving a direct answer to the question he had been asked.
4. *Back-channel responses* or continuers such as *uh-huh, yeah, and m-hm* (Goodwin, 1986a; Schegloff, 1982; Yngve, 1970). Speakers use these to acknowledge they have heard or understood their partner well enough for current purposes.
5. Certain *gestures*, including certain head nods, eye gaze, smiles, grimaces, and pointing (Bavelas, Chovil, Lawrie, & Wade, 1992; Bavelas & Chovil, 2000; Goodwin, 1981, 1986b; Goodwin & Goodwin, 1986). Speakers use these to acknowledge what is being said and otherwise coordinate with their partners.
6. Certain *strategic silences and overlaps* (Goodwin, 1981; Schegloff, 1987). Speakers use these to indicate such things as reluctance or demands to speak.
7. *Nonreduced vowels* (such as “thee” instead of “thuh” for the word *the*) and prolonged syllables (Fox Tree & Clark, 1997). Speakers use these to indicate they are suspending speech or adding a delay because of some problem in production. When one speaker said, “when you come to look at thee . thuh literature,” he signaled that he

was having problems deciding on *literature*, which he immediately amended, "I mean you know the actual statements."

8. *Preliminary commitments* (Clark & Wasow, 1998). Speakers often produce a word or phrase on its own to commit themselves to speaking before they are able to proceed fluently. When Reynard says "if uh . if" he produces the first *if* to commit himself to the upcoming *if*-clause that he cannot yet produce.

In summary, speaking has many origins and constraints. People speak primarily to advance their joint activities—from business exchanges to telling stories. They form plans at many levels—from sections, sentences, and intonation units down to words, suffixes, and phonetic segments. At the same time, people monitor what they and their interlocutors are doing and saying. They create not only primary signals for their official business, but collateral signals to deal with the on-going performance itself.

## LISTENING

For every action in speaking, there must be a corresponding action in listening. (Compare Newton's third law of motion: "For every action, there is an equal and opposite reaction.") Just as speaking divides into four levels, listening does also by:

1. Attending to the speakers' vocalizations and gestures (channel level);
2. Identifying the speakers' signals (signal level);
3. Understanding what the speakers mean by those signals (intention level); and
4. Considering the joint projects proposed (project level).

Listening has been investigated mostly at the signal and intention levels in artificial set-

tings. Still, these investigations have established many of the processes by which listening takes place.

Listeners begin with the raw material they hear and see—the speaker's vocalizations and gestures. They recognize that speakers produced these in attempts to advance the current joint activity—whether it was diplomacy or gossip, a business transaction, or a card game. So, listeners recognize that these signals must satisfy two constraints: (a) They must be consistent with the raw material heard and seen; and (b) they must contribute to the speakers' moves in their current joint activity. Listening works both from perception up and from purpose down. Early on, most investigations were on the processes that work from the bottom up, but more and more have revealed processes that work from the top down.

### Identifying Words

Speech doesn't come parsed into words, phrases, clauses, and sentences. Most intonation units, the main units of speech identifiable from prosody, are uninterrupted streams of speech sounds. "I'll have four please" might come off "Illhavefourplease," with no noticeable gaps. Worse yet, the pronunciation of many words and phrases further obscures their boundaries. "In boats" is regularly pronounced "im.boats" (the period marks a syllable boundary), "an egg" as "a.negg," "to eat" as "to.weat," and "the apple" as "the.yapple." When speech is informal and quick, "Why don't you eat?" may sound like "Wain.cheat?" Listeners must have a remarkable ability to discover order in apparent disorder.

Evidence suggests that listeners identify words one speech segment at a time. In one study (Marslen-Wilson & Tyler, 1980), people were asked to listen to speech and, when they heard a specific target word, to press a button as quickly as possible. Some people, for example, listened for the target word *lead* in

"The church was broken into last night. Some thieves stole most of the lead off the roof." Listeners identified the target words, which averaged 420 ms long, a mean of 50 ms before the ends of the words. If it is assumed that pressing the button takes about 200 ms, then listeners identified the word about 250 ms before the end of a word—less than half way through. How is that possible?

According to a model by Marslen-Wilson and colleagues (e.g., Marslen-Wilson, 1987), listeners begin with the first sound of a word and then use the succeeding sounds to narrow down the possibilities until they arrive at a unique word. Take *trespass*. After the first sound "t-," listeners activate in memory the entire cohort of English words that begin with "t-." With over 1,000 such words, each one gets only a small activation. After "tr-," listeners reduce that cohort to words that begin with "tr-," which may run into the hundreds. By "tresp-," listeners have reduced the cohort to a unique word, *trespass*, which gets all the activation. So the "p" in "trespæs" is called the *uniqueness point*. It has been shown that the earlier the uniqueness point, the earlier listeners can identify the word.

Are all of these preliminary words activated in memory? The evidence suggests that they are. Consider *captain*. Before the "e" in "kæpten," listeners should activate not only *captain*, but also *captive* (and other words). As a result, they should be primed to identify words related to both *captain* and *captive*, for example *ship* and *guard*. But just after the "e" (the uniqueness point for *captain*), listeners should activate only *captain*, which primes *ship*, but not *guard*. Indeed, this is precisely what Zwitserlood (1989) found. Research shows that listeners activate entire cohorts of words, which they reduce to unique words when they get enough evidence (cf. Elman, 1989; McClelland & Elman, 1986). They can narrow down the options even faster by taking note of the potential referents. In

one experiment, listeners sat at a table with candy and other objects on it and were told to "Pick up the candy." Their eyes darted toward the candy even before the end of the word *candy*. When there was both candy and a candle on the table, they took longer because the candle delayed the uniqueness point for identifying the word *candy* (Dahan, Swingley, Tanenhaus, & Magnuson, 2000; Tanenhaus & Spivey-Knowlton, 1996).

If "I'll have four please" is pronounced "I'llhavefourplease," how do listeners know when to start a new word? They don't. In a study by Shillcock (1990), when listeners heard, "He carefully placed the *bone* on the table," they were primed by *bone* to identify the word *rib*. This is not surprising. But when other listeners heard, "He carefully placed the *trombone* on the table," they were just as primed by *trombone* for the word *rib*. Listeners apparently hear *bone* in *trombone*—at least briefly. In working bottom up, listeners initially activate a wide range of extraneous words.

Identifying a word requires not only its phonological shape, or lexeme—such as "trespæs" or "kæpten"—but its intended sense, or lemma. Most words are ambiguous, so listeners must select from a range of lemmas. An example of this is the word *bug* in "He found several bugs in the corner of his room." Without knowing more about what the speaker is trying to say, *bug* could equally mean "insect" or "hidden microphone." As Swinney (1979) showed (see also Tanenhaus, Leiman, & Seidenberg, 1979), people who listened to this sentence were primed by *bug* to identify words related to *both* of these meanings—say, *ant* for "insect" and *spy* for "hidden microphone" (compared to neutral *sew*). Surprisingly, however, other listeners were just as primed by *bug* to identify *both ant* and *spy* (compared to *sew*) in an utterance that had no ambiguity at all: "He found several spiders, roaches, and other bugs in the corner of

his room." Listeners were primed for both *ant* and *spy* immediately after *bug*, but just a few syllables later they were primed for *ant*, but not for *spy*. These listeners had quickly *deactivated* the unintended lemma "hidden microphone." From this and many other investigations, it appears that listeners activate all common senses of a word and then deactivate those that don't fit.

Listeners, then, have a dual problem: how to identify the lexemes within the continuous stream of speech, and how to settle on the intended lemmas. What makes it such a problem is that listeners activate too many lexemes and too many lemmas. They need powerful top-down methods for settling on the right ones.

### Sentence Structures

Sentences have an orderliness that listeners can count on as they try to hear words as parts of larger structures. Although languages of the world differ, they tend to conform to a small number of principles about sentence structures. It would be odd if listeners did not exploit these principles, and they do. The next section describes four such principles.

1. *Grouping*. As Behaghel noted over a century ago, "What belongs together mentally is placed together syntactically" (see Venneman, 1973, 1975). Another way to phrase this claim is that words that jointly refer to the same object, event, or process tend to be placed in a single constituent. In English, "I'll have a bowl of clam chowder and a salad with Russian dressing" divides into constituents as follows, where each constituent is enclosed in a pair of square brackets:

[I'll [have [[a [bowl [of [clam chowder]]]]] and  
[a [salad [with [Russian dressing]]]]]]]

Mentally, *clam* and *chowder* go together (both refer to the soup) and, indeed, they form a constituent, a noun phrase. Likewise, *of* and *clam*

*chowder* go together, and they form a prepositional phrase. English relies heavily on grouping for denoting the relations among words, so listeners should try hard to identify constituents.

2. *Ordering*. "Relations among propositions tend to be marked by word order" (see Greenberg, 1963). In English, typical sentences are subject + verb + object (as in "I'll have a bowl of clam chowder"), but in Japanese, they are subject + object + verb. In addition, "word pairs that are alike in function tend to have the same internal ordering" (see also Lehmann, 1972, 1973). In English, modifiers tend to come before nouns, as in *Russian dressing*, *clam chowder*, *that dog*, and *two hamburgers*. If the modifiers are complex, they tend to come after nouns, as in *bowl of clam chowder*, *dog that I saw*, and *hamburgers good enough to eat*. In French and Spanish, even simple modifiers tend to follow nouns. Therefore, it is wise for listeners to attend to word order in order to identify subjects, verbs, objects, and modifiers.

3. *Case-marking*. "Words marked for case denote distinct roles." In English, *I* and *we* are used for denoting subjects, whereas *me* and *us* are used for objects of verbs, objects of prepositions, and other functions. Nominative, accusative, and possessive pronouns have distinct functions. In German, articles, adjectives, and some nouns also use case-marking. *The man* is translated as *der Mann*, *dem Mann*, and *den Mann* for the nominative, dative, and accusative cases depending on whether *the man* is, for example, the subject, indirect object, or direct object of the verb. It is information that German listeners rely on.

4. *Agreement*. "Words that agree (in number, gender, etc.) tend to refer to the same object, event, or process." In French, the three words in *le soleil rond* ("the round sun") are each masculine, and those in *la lune ronde* ("the round moon") are feminine. French listeners can count on agreement to help them

identify *le*, *soleil*, and *rond* as referring to the same object. English makes almost no use of agreement.

Grouping and ordering are exploited by English listeners, as simple examples demonstrate. Consider "John said he will come yesterday." One reason this sentence sounds strange is that listeners try to make a constituent out of *he will come yesterday*, and that makes no sense. Listeners have trouble seeing that *yesterday* goes with *John said* because that would create a discontinuous constituent.

Or consider: (1) "John figured out that Susan wanted to take the train to New York" versus (2) "John figured that Susan wanted to take the train to New York out." In sentence 1 it is easy to create the verb *figure out* because *figure* and *out* form a constituent. In sentence 2 it is difficult to see the verb as *figure out* because the verb is discontinuous and because *the train to New York out* forms an interpretable constituent.

Finally, consider "The man pitched the ball threw the ball." As we go along, we form a subject-verb-object constituent of *the man pitched the ball*, but then we are left with the fragment *threw the ball*. The sentence seems to make no sense. Change it to *The man thrown the ball pitched the ball*, and the problem disappears. *Thrown* cannot be the main verb, so we realize that *the man thrown the ball* is a noun phrase and *pitched the ball* is the main verb and object. It is easy to see why parsing "The horse raced past the barn fell" is so difficult (Bever, 1970).

Languages differ in how they mark syntactic relations. English makes heavy use of grouping and ordering, whereas German makes greater use of case-marking and less of ordering. Walpiri, a language of Australia, makes heavy use of case-marking and almost none of ordering. Parsing strategies should reflect these differences, and evidence suggests they do. In this volume, Rayner and Clifton review the processes for parsing and

comprehending utterances, so this chapter addresses only the basic issues.

### Comprehension Processes

How do English listeners identify syntactic relations? According to some proposals, listeners work largely or solely bottom up. Suppose people read "The reporter saw her friend . . ." one word at a time. If they realize that *saw* most often takes concrete objects, they should infer that "reporter saw friend" is subject + verb + object. So when the sentence goes on, ". . . was not succeeding," they should be startled at *was* and recover only after a delay, as was the case in an experiment by Holmes, Stowe, and Cupples (1989). Listeners were *not* startled, however, when the sentence began "The reporter saw *that* her friend . . ." Nor were they startled at *would* in "The candidate doubted his sincerity would be appreciated." For this utterance, they apparently assumed that the direct object of *doubt* is most often a full clause, such as "that his sincerity would be appreciated." Evidence like this suggests that listeners know about the constructions that words are most likely to occur in, and they use that knowledge in parsing utterances into constituents (MacDonald, Pearlmutter, & Seidenberg, 1994).

Listeners also work top down. Consider "The burglar blew open the safe with the new lock." The phrase *the safe with the new lock* makes sense if there are two safes, one with a new lock and one without. It makes less sense if there is only one safe. In one part of an experiment by Altmann and Steedman (1988), people read one of the two versions of this passage:

A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was a safe with a new lock and a safe with an old lock [or: *a strong-box with an old lock*]. The burglar blew open the safe with the new lock.

People read the phrase *with the new lock* faster in the version with two safes than in the version with a safe and a strongbox. But when *with the new lock* was replaced by *with some dynamite*, they were faster with one safe than with two. Readers used their knowledge of the situation already described (one vs. two safes) to help them identify which relation was being introduced by *with* (a modifier for *safe* or a modifier for *blow up*).

Listeners can also exploit their knowledge of the scene around them. In an experiment by Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy (1995), people sitting at a table that had apples, towels, and boxes on it were instructed to (among other things) "Put the apple on the towel in the box." As a sentence, "Put the apple on the towel in the box" is ambiguous. Is an apple to go on a towel in a box, or is an apple on a towel to go in a box? Without context, people tend to choose the first grouping: Put [the apple] [on the towel in the box]. Indeed, when there was one apple on one towel, and a second towel, listeners were confused. Their eyes darted first to the apple on the towel, then to the second towel, and only after a delay did they put the apple in the box. But when there were *two* apples on the table, one on a towel and one not on a towel, they had no trouble at all. Their eyes immediately settled on the apple on the towel, and they put it into the box. They used their knowledge of visual layout to help them parse the utterance as intended: Put [the apple on the towel] [in the box].

Top down processes appear to be pervasive, but no one knows how pervasive. For years there was a sign in a London hospital that read, "No head injury is too trivial to ignore," but no one had noticed that it made no sense (Wason & Reich, 1979). It was taken to mean that "You should never ignore a head injury, no matter how trivial" even though it literally means "There is no head injury that is so trivial, so small, that it shouldn't be ignored." Examples like this are common

(Erikson & Mattson, 1981; Fillenbaum, 1971, 1974; Reder & Cleeremans, 1990). They suggest that people do only a partial analysis of many constructions, cutting the process short by introducing plausible interpretations. How complex must sentences be for people to take these shortcuts? Probably, no construction is too trivial to ignore.

### Implicatures

Speakers ordinarily mean much more than they say. When Jane places a phone call to Miss Pink's office and asks the secretary, "Is Miss Pink in?" she appears to be asking, literally, whether or not Miss Pink is in. But she expects the secretary, Kate, to recognize that the question is a pre-request. In terminology introduced by Grice (1975, 1978, 1991), what Jane *says* (in Grice's special sense) is a question to be answered yes or no. But *by* saying that, she also *implicates* that she wants to talk to Miss Pink. Indeed, Kate first answers the question and then deals with the implicature, "Well, she's in, but she's engaged at the moment." In Grice's view, speakers intend their addressees to *work out* these implicatures as part of what is meant.

Listeners must therefore infer what speakers are implicating. Traditionally, these inferences have been divided into *backward* and *forward* inferences (Clark, 1977a,b; Clark & Haviland, 1977; see Garrod & Sanford, 1994; Sanford & Garrod, 1994; Singer, 1994; van den Broek, 1994). Although the two types of inferences have been investigated mostly in reading artificial narratives, the findings probably extend to listening as well.

### Backward Inferences

In Grice's scheme (see also Sperber & Wilson, 1986), speakers are expected to follow the maxim: "Be relevant." They are assumed to make their current contribution relevant to the on-going joint activity. Working out how it is relevant leads to implicatures, as

in the following two artificial fragments of a discourse:

- (14) I just bought a shirt and tie at Macy's. *The shirt* was on sale.  
 (15) I just bought a shirt at Macy's. *The price* was just right.

In each sequence, addressees must determine how the second sentence is relevant to the first, and draw the inferences needed to establish that relevance. Recall that definite references (such as *the shirt* and *the price*) require their referents to be inferable from current common ground. So in 14, addressees infer that *the shirt* in the second sentence refers to the shirt mentioned in the first. The inference is trivial, but essential to establishing what the speaker means. In 15, the inference is more complex. Addressees infer that the shirt was bought for a price, which is the referent for *the price* in the second sentence.

Inferences needed for establishing relevance or coherence have been called *bridging inferences* (Clark & Haviland, 1977) and *accommodation* (Lewis, 1979). Bridging inferences take many forms, as the following sequences illustrate (see also Clark, 1977a; Mann & Thompson, 1986; Prince, 1981; Singer & Halldorson, 1996; Sperber & Wilson, 1986):

- (16) I went for a walk this afternoon. The park was beautiful.  
 [Bridge: One place where I walked was a park, the referent of *the park*.]  
 (17) Duncan has a black eye. It was Bob who hit him.  
 [Bridge: Duncan has a black eye because someone hit him.]  
 (18) Margaret went horseback riding last week. She was sore for three days.  
 [Bridge: Margaret was sore in the way riders get sore because of the ride.]  
 (19) They're having a party again next door. I couldn't find a parking place.  
 [Bridge: I believe they're having the party because I couldn't find parking.]

As these examples show, bridging inferences are part and parcel of what people under-

stand. Still, there is no unified account of how they are created (Garrod & Sanford, 1994; Sanford & Garrod, 1994; Singer, 1994; van den Broek, 1994). Sometimes, they take measurable time to create; other times they do not. Most bridging inferences show up in tests of memory of a passage, but some do not. Two points seem clear: Addressees base their bridging inferences on the current joint activity or situation—what they are doing with their partners at the moment; and the bridging inferences they draw are the simplest inferences needed to establish the speaker's utterance as the relevant next move in that activity.

### Forward Inferences

In Grice's scheme, speakers are also expected to adhere to two other maxims: (a) "Make your contribution as informative as is required (for the purposes of the exchange)" but "no more informative than is required," and (b) "Be brief (avoid unnecessary prolixity)." What follows is a breathtaking variety of implicatures. To give an idea of their range, we present three heuristics that follow from the maxims, as characterized by Levinson (2000):

**Heuristic 1.** "What isn't said, isn't." When Ann is asked "How many children do you have?" and she answers, "I have two sons," she implicates that these are all of her children. If she had had others, she would have mentioned them.

**Heuristic 2.** "What is simply described is stereotypically exemplified." When Charles says, "The accountant dried her hands," listeners take him as implicating that she dried her hands in the ordinary way and not, say, on her dress. What is ordinary, or stereotypical, depends on the situation. At the dinner table, the accountant might be expected to dry her hands on a napkin, but in the washroom, to dry them on a towel.



Implicatures based on this heuristic are often called *elaborative inferences* (Garrod & Sanford, 1994; Sanford & Garrod, 1994; Singer, 1994; van den Broek, 1994), and they have been widely studied in comprehension and memory. In one experiment, people who heard "The man dropped the delicate glass pitcher on the floor" often misrecognized it later as "The man *broke* the delicate glass pitcher on the floor" (Johnson, Bransford, & Solomon, 1973). In another study, people who had just read "Steve threw a delicate porcelain vase against the wall" were able to name (read aloud) the word *break* faster than people who had just read "Steve went out and purchased a delicate porcelain vase" (Murray, Klin, & Myers, 1993). On the other hand, when people were presented with the sentence "The director and the cameraman were ready to start shooting when suddenly the actress fell from the 14th floor," they were primed for the word *dead* only after a delay (McKoon & Ratcliff, 1992).

Elaborative inferences often anticipate backward inferences that will be needed later, as in the following sequences:

- (20) Keith took his car to London. *The car kept overheating.*  
 [Bridge: The car mentioned in the first sentence is the referent of *the car*.]
- (21) Keith drove to London. *The car kept overheating.*  
 [Bridge: Keith drove a car, the referent of *the car*.]

In one study (Garrod & Sanford, 1982), it took no longer to read "The car kept overheating" in sentence 21 than in 20. Apparently, when participants read "Keith drove to London," they inferred the stereotypical vehicle, a car; therefore it was as easy to draw the bridging inference as when the car was mentioned explicitly. But what if *drove* is replaced by *went*?

- (22) Keith went to London. *The car kept overheating.*  
 [Bridge: Keith drove a car, the referent of *the car*.]

Depending on the participants' common ground, the stereotypical means of transportation could be a car, bus, train, or airplane, and the backward inference would be more work. Elaborative inferences like this are essential to narratives.

**Heuristic 3.** "What's said in an abnormal way, isn't normal." When Michael says, "Susan stopped the car," he implicates that she stopped it in the stereotypical way—by using the foot brake. But when he says, "Susan caused the car to stop," he selects the wording *cause to stop* over the expected *stop*. By doing so, he implicates that Susan's method was *not* normal; for example, she may have used the emergency brake.

Many of the implicatures created by heuristic 3 have been investigated as instances of *indirection*. Here, again, is a pre-request:

- (23) Susan (on telephone) Can you tell me what time you close?  
 Store manager Yes, we close at nine.

Instead of asking, "What time do you close?" Susan went out of her way to create a pre-request. Why go to all that work? The answer, according to many, is to be polite (Brown & Levinson, 1987; Clark & Schunk, 1980; Goffman, 1967; Lakoff, 1973). It is polite (a) to offer the addressee a way out of the request, and (b) to add to the addressee's self-regard. Both help the speaker and addressee maintain face. So speakers set up pre-requests, like Susan's, to deal with the greatest obstacles to compliance, and addressees infer this (Francik & Clark, 1985; Gibbs, 1986). In this example, Susan pretends that the greatest obstacle is the manager's ability to tell her the closing time. It would be odd to pretend that it was his happenstance knowledge: "Do you happen to know what time you close?" Addressees often infer the point of pre-sequences without extra time or apparent effort (Gibbs, 1979, 1983; Gibbs & Mueller, 1988).

In summary, listeners seem to work bottom up. They identify speech sounds and use them to identify words, then phrases, and then entire intonation units. They use the successive segments of a sound stream to narrow down on the intended word, activating the lemmas, or word senses, of all the potential words at any moment. But listeners also work top down. They are normally engaged in a joint activity (e.g., listening to a narrative, answering a question, talking about a scene) and that allows them to narrow in on intended words more quickly, eliminate inappropriate lemmas, and parse utterances into their appropriate parts. Also, they have procedures for inferring what speakers mean. Some of these lead to bridging inferences that establish reference and coherence with what has come before. Others lead to elaborative inferences and inferences about indirection.

## MEANING AND SIGNALS

Speakers mean things by what they say. When Alan asks Beth (at the drug store counter), "Do you uh have size C flashlight batteries," he means that she is to say whether the store has size C flashlight batteries available for sale to him at that moment. This is what is called *speakers' meaning* (Grice, 1957, 1968, 1991). Speakers' meaning is a type of intention (an intention that speakers intend their addressees to recognize), and it arises from what the speakers are trying to accomplish in the current joint activity. Speakers get their addressees to recognize these intentions by speaking, winking, gazing, nodding, smiling, pointing, and making other gestures. These actions are *signals*, or actions by which one person means something for others.

### Methods of Signaling

The meaning of a signal is very different from a speaker's meaning. It is not an intention, but

a specification of the relation between the signal and the world. The word *battery*, for example, can mean "an artillery emplacement," "an array of objects," or "a device for producing direct current." These are its *type meanings*. Alan used it at the drug store counter with the *token meaning* "a device for producing direct current." Winking, gazing, nodding, smiling, pointing, and others signals have meanings too. What are their meanings, and how do they acquire those meanings?

The late-19th-century philosopher Charles Sanders Peirce offered one influential answer in his theory of *signs* (Buchler, 1940). According to Peirce, signs represent "objects" (physical things, actions, events, properties) under certain interpretations. A portrait of Napoleon is a sign that represents a particular man under the interpretation "Napoleon Bonaparte." Signs, in turn, come in three types:

1. *Icons*. Icons represent their objects by means of a *perceptual resemblance* to the objects. Napoleon's portrait represents Napoleon by its perceptual resemblance to Napoleon.
2. *Indexes*. Indexes represent their objects by means of a *physical or causal connection* to those objects. A road sign represents a village by pointing to the village.
3. *Symbols*. Symbols represent their objects by means of *rules*. Both *dog* and *chien* signify domesticated canines, one by a rule of English and the other by a rule of French.

What, then, about signals? Speakers make signals by using, creating, or forming signs for their addressees. In spontaneous speech, speakers have three basic *methods* of signaling:

1. *Demonstrating*. Demonstrating is signaling by means of icons. When Alan shapes his hand like a telephone and places it to his ear (forming an icon), he is *demonstrating* the act of telephoning.

2. *Indicating*. Indicating is signaling by means of indexes. When Alan points at a car (forming an index), he is *indicating* the car.
3. *Describing-as*. Describing-as is signaling by means of symbols. When Alan gives a "thumbs-up" (a symbol) or says "Great" (another symbol) to a tennis serve, he is *describing* the serve *as* excellent.

Most signals are *composite signals*, which are fusions of two or more of these methods. When Alan points at a car and says, "that car," he is referring to the car with a single signal, but the signal is a composite of indicating the car and describing it as a car.

Most work has focused on symbols because those are what researchers generally think of as "language." Traditional linguistics includes the study of phonetics, morphology, syntax, semantics, and pragmatics, all of which are primarily symbols. But face-to-face conversation relies on symbols, indexes, and icons in both linguistic and nonlinguistic methods.

### Describing

The prototypical symbols are words and the sentences created from them. Whenever people select words and create sentences, they are using symbols, by *describing* something *as* something. How do these symbols work?

For the past thirty years, most accounts of language use have assumed that people possess *mental lexicons*, or dictionaries in the head. A mental lexicon is an organized list of dictionary entries, called *lexical entries*, to which people refer when producing and comprehending utterances (see Dell & O'Seaghdha, 1991; Levelt, 1989; Levelt et al., 1991; Levelt, Roelofs, & Meyer, 1999). As we discuss earlier in the chapter, each lexical entry has two parts: (a) the phonological form of the word, its lexeme; and (b) the mean-

ing of the word, its lemma. The lexical entry links the two parts. The lexical entry for *dog* is a pairing of lexeme and lemma: [/*dɔg*/, "domesticated canine"]

The notion of mental lexicon raises a number of issues for psycholinguistics. We consider four of them: (1) conventions, (2) lexical items, (3) communal lexicons, and (4) symbolic gestures.

### 1. Conventions

Researchers ordinarily assume that language is conventional—in particular, that words are conventional. But what is a convention? The answer is often treated as self-evident, but it is not. The issue is central to the notion of mental lexicon.

The modern analysis of conventions comes from David Lewis (1969). As Lewis argues, people, such as Alan and Beth, have to coordinate with each other to reach a common goal. They face a *coordination problem* in reaching that goal. Suppose they want to greet each other. Should they hug, kiss, shake hands, or what? The first time they meet, they may solve the problem by agreeing to shake hands. Agreeing to shake hands is a *coordination device*—a solution to their coordination problem. If they meet regularly, they have a *recurrent* coordination problem for which they need a general solution. They may come to *mutually expect* to shake hands, and shaking hands becomes a convention. For Lewis (though the wording is ours):

A convention is:

- (a) a regularity in behavior
- (b) that is in common ground in a given community
- (c) as a coordination device
- (d) that is partly arbitrary
- (e) for a recurrent coordination problem.

Shaking hands is (a) a regularity in behavior. It is (b) common ground for Alan and Beth (c) as

a coordination device (e) for the recurrent coordination problem of greeting each other. It is (d) partly arbitrary because, with a different history, they might expect to hug instead.

Most conventions evolve slowly and are learned as part of one's culture, but, in the right circumstances, they can also develop quickly. In a study by Garrod and Doherty (1994; also Garrod & Anderson, 1987), pairs of people sat at separate computer terminals and tried to negotiate their way through mazes on their screens. Although they had the same underlying maze (an incomplete matrix), they were shown different elements of it. To succeed, they had to exchange information, which made them coordinate on how they talked about locations. One pair might refer to a location as "four lines down and two boxes over," (using *lines* and *boxes*), but another might say, "row four column two" (using *rows* and *columns*). In one condition, pairs of people played with each other multiple times. In another, people were grouped into an informal community, and each played as many times as in the first condition, but once with every other member of the community. The isolated pairs developed local agreements for referring to location, but each pair tended to develop a different one. In contrast, the pairs in the community began with different local agreements, but soon converged on the same solution—typically rows and columns. The convention evolved as a solution to the recurrent, community-wide coordination problem.

Conventions, Lewis (1969) argued, are the basis for natural languages. In talking, Alan and Beth have the recurrent coordination problem of how Alan is to get Beth to see that he is denoting a domesticated canine. They recognize that they are both members of the community of English speakers in which it is common ground that *dog* can be used to denote such a beast. They can solve their coordination problem by Alan using *dog* and Beth interpreting him as denoting a domesticated

canine. The word *dog* is, therefore, conventional. This solution is partly arbitrary, because if English history had been different, we might be using *hound*, *chien*, or *perro* instead. The mental lexicon is a *system* of such conventions organized into lexical items and communal lexicons.

## 2. Lexical Items

In common parlance, most words have more than one sense. The word *ear* has at least three:

- Sense 1.** The visible organ of hearing, as in "floppy ears";
- Sense 2.** The sense of hearing, as in "good ear for jazz";
- Sense 3.** The spoke from which corn grows, as in "three ears of corn."

But how many "words" do these represent? Let us consider three models. In Model A, there are three distinct words (*ear*) that just happen to sound the same. This model treats senses 1 and 2 as unrelated, and that seems wrong. In Model B, there is just one word *ear*, which has three senses. This model also seems wrong because it misses the fact that sense 3 is conceptually unrelated to senses 1 and 2. In Model C, there are two words or *lexical items*, one for senses 1 and 2, and a second for sense 3:

- ear*<sub>1</sub>: [/ɪr/, "the visible organ of hearing"]  
[ɪr/, "the sense of hearing"]
- ear*<sub>2</sub>: [/ɪr/, "the spike on which corn grows"]

In this view, a lexical item is a collection of related lexical entries. Indeed, most dictionaries of English divide *ear* into just these parts.

Model C reflects a difference between *polysemy* and *homonymy*. *Ear*<sub>1</sub> is polysemous because it has more than one related lexical entry. But *ear* is also a homonym because it has two unrelated sets of lexical entries, represented by *ear*<sub>1</sub> and *ear*<sub>2</sub>. It is often easy to

identify homonyms by examining other languages. French, for example, has different words for *ear*<sub>1</sub> and *ear*<sub>2</sub>—*oreille* and *épi*—but like English, *oreille* has the two senses of English *ear*<sub>1</sub>.

How do we decide whether or not “visible organ of hearing” and “sense of hearing” belong to distinct lexical entries of *ear*<sub>1</sub>? The answer isn’t simple. *Line*, for example, has five apparently distinct senses:

- Sense 1.** A physical mark, as in “Two parallel *lines* never meet”;
- Sense 2.** A demarcation, as in “His car was checked at the state *line*”;
- Sense 3.** A continuous arrangement, as in “We stood in *line* for the tickets”;
- Sense 4.** A continuous sequence of words, as in “The actress learned her *lines*”;
- Sense 5.** A sequence of constructs, as in “What *line* of work are you in?”

In a study by Caramazza and Grober (1976), people judged sense 1 to be the most central sense of *line* and sense 5 the least central. From these and other judgments, Caramazza and Grober argued that *line* has a core meaning, “an extension,” and the five senses are derived from it. But do these five senses represent *five* distinct lexical entries, each with a different lemma? Or is there just one lexical entry with the lemma “an extension”?

The issue is one of *sense selection* versus *sense creation* (Clark & Clark, 1977; Clark, 1983; Clark & Gerrig, 1983; Rapp & Gerrig, 1999). People invent new senses every day, as these attested examples show: “The initiative is aimed at preventing the *New Yorking* of the San Francisco skyline”; “The photographer asked him to do a *Napoleon* for the camera”; and “We’re looking for a *size 10* with a steam iron” (a female roommate who wears size 10 and owns a steam iron). The words *New York*, *Napoleon*, and *size 10* do not come with the needed lexical entries. The novel senses had

to be *created* from the known lexical entries for *New York*, *Napoleon*, and *size 10*.

In the right circumstances, it takes listeners no longer to interpret novel words than conventional words. In one study (Gerrig 1989b; see also Gerrig & Bortfeld, 1999; Gerrig & Gibbs, 1988), readers were given brief stories that ended with a noun compound like *snow-ball* or *fire-ball*. Readers were faster to read and understand *snow-ball*, a compound they could access quickly, when the story led up to its conventional meaning (“ball made of snow”) than when it led up to a novel meaning (“dance in honor of a big snowstorm”). For a compound like *fire-ball*, whose conventional meaning could not be accessed as quickly, readers were just as fast in reading and understanding it with the novel meaning (“dance in honor of a famous fire”) as with the conventional one (“ball made of fire”). People appear able to access conventional meanings at the same time as they create novel meanings, and the novel meanings sometimes arrive before the conventional ones.

The line between conventional and innovative is difficult to draw. At the conventional end, we have “*to fly* to Amsterdam,” and at the innovative end, “*to KLM* to Amsterdam.” But look at these examples of the word *newspaper*: “The newspaper is on the table” (the physical newspaper); “The newspaper says it’s going to rain today” (an article in today’s edition of a newspaper); “I used to work for the newspaper” (the publishing company); “The newspaper called me today for an interview” (someone who works for the publishing company); and “I stopped by the newspaper for my interview” (the office of the newspaper company). The list begins at the conventional end, but is the last sense of *newspaper* conventional, or do we create it on the spot (as we do for *to KLM*)?

Lexical entries are therefore organized into lexical items (like *ear*<sub>1</sub> and *ear*<sub>2</sub>), which further organize themselves—if *line* is any

indication. But when people use a word, they often treat one of its conventional lexical entries as a starting point for creating a novel sense for that occasion—a nonce sense. It is the only way to interpret *New Yorking*, *a Napoleon*, *size 10*, *KLM*, *newspaper*, and many other such expressions.

### 3. Communal Lexicons

In Lewis's scheme, a convention holds only for a *particular community of people*. Most accounts assume a single community for the entire English lexicon—the community of English speakers. That cannot be right. For example, the words *sclerotic* and *myocardial* are in common ground for medical doctors, like the words *mortmain* and *nonfeasance* for lawyers. They are common ground only within these communities of shared expertise—medicine and law. If so, lexical entries are organized, not into a single monolithic lexicon, but into many *communal lexicons* (Clark, 1998). The largest lexicons reflect shared expertise in a language like English or Japanese. The smallest reflect esoteric types of expertise like contract law, lacrosse, or Palo Alto.

Almost every community has evolved a lexicon for its shared expertise, and Lewis' account of conventions makes it easy to see why. Conventions arise as solutions to *recurrent* coordination problems. Most of us have little need (especially a recurrent need) to refer to the notion of "tissue death." But doctors do, so they have evolved the term *infarct*. As a community, they find it a *useful* term. Most of us, even after being introduced to the term, do not have the expertise or background to use it. Doctors, as a community, do, so they find it a *usable* term. For a word to arise in a community, it must be both *useful* and *usable*. It is these twin requirements that lead to the size and number of communal lexicons.

Communal lexicons are essential to speaking and listening. For two people to talk, they must use the same vocabulary, and to find one,

they need to establish joint membership in a community and use its lexicon. When Alan, an American, steps off the plane in Tokyo, he might approach Yuko, a stranger, and ask, "Do you speak English?" If she says, "Yes," the two of them can assume joint membership in the community of English speakers. Still, he cannot go on to "My heart has an infarct" without establishing that both are English-speaking doctors. When people first meet, they generally spend time establishing common ground for further conversation. That includes joint membership in communities of shared expertise.

### 4. Symbolic Gestures

Words and constructions are not the only symbols of language use. There is also a class of gestures called *emblems* (Ekman & Friesen, 1969; McNeill, 1992). For North Americans, these include: thumbs-up, thumbs-down, greeting wave, farewell wave, thumb and index finger in circle ("excellent"), winks, index finger to protruding lips ("be quiet"), crossed fingers, and shoulder shrugs (see Johnson, Ekman, & Friesen, 1975). The two most common are head-nods ("yes") and head-shakes ("no"). Most emblems are not used as constituents of spoken utterances, but on their own. Most correspond to one-word interjections such as *yes*, *no*, *okay*, *hello*, *goodbye*, *excellent*, or *quiet*, or to simple sentences such as "I'm kidding" or "I don't know."

Emblems have conventional meanings and are, in certain other respects, like words. The same gesture (e.g., crossed-fingers) means radically different things from one community to the next (Morris, Collett, Marsh, & O'Shaughnessy, 1979). Many emblems belong to highly specialized communities. In baseball, an umpire sticking his right thumb behind his head means "You're out." So emblems must have lexical-like entries that link form and lemma, such as [head-nod, "yes"] and [wink, "I'm kidding"], and that belong to communal lexicons. Sign languages such as

American Sign Language are complete languages built on emblem-like gestures.

The process of *describing*, therefore, works with symbols, or signs associated with objects by rule. The most basic symbols, words and emblems, have conventional lexical entries, such as [/dɒg/, domesticated canine] and [wink, "I'm kidding"]. These are organized into lexical items, which are organized into communal lexicons. In speaking and listening, people must do more than match the correct lemma with the correct lexeme. They must establish and use joint communal lexicons. Often, too, they must create or interpret novel words, deriving or inferring nonce meanings from conventional meanings.

### Indicating

*Indicating* is a method of signaling by which people create indexes for the objects to which they want to refer. The prototype is pointing with the finger (*index*, in Latin, means "finger"), which is often called a *deictic gesture*. In a bookstore, Alan points at a copy of Melville's *Moby Dick* and asks Beth, "Have you ever read *that*?" His pointing (a) specifies a location, and (b) gets Beth to attend to a thing at that location. There is an *intrinsic connection* between his gesture, the aiming of his finger, and the thing itself. In Peirce's scheme for indexes, indicating requires an additional step—an interpretation. Every indication refers to a thing *under a particular description*. For Alan's referent, the description is "something Beth may have read."

Indicating seems so simple that there is nothing to explain. But appearances belie reality. The following sections describe several complications.

### Directing Attention

Pointing with the finger is a technique we refer to in this chapter as *directing-to*. When Alan points at a copy of *Moby Dick*, he is *directing* Beth's attention *to* the book by getting her to

follow the bearing of his finger. To indicate, speakers can use any device that directs their addressee's attention to the referent.

1. *Parts of the body*. Speakers can point with the finger (Alan's "Have you ever read *that*?"), sweep over an area with the arm ("All *this* is yours"), nod at a thing ("She was standing over *there*"), touch or tap on a thing with the hand or foot ("*This* is the book [or rug] I want"), turn the head or torso toward a person ("Let *us* talk"), and gaze at a person ("I want *you* and *you* and *you* to come with me"). In some societies, speakers conventionally point with pursed lips or a protruding upper lip.
2. *Voice*. Speakers can indicate their locations by the source of their voices ("I'm *over here*"), and their identities by the sounds of their voices (on the telephone: "It's *me*"). Speakers can indicate points in time by the timing of their vocalizations (race official: "Ready . . . set . . . go"). Most interpretations of *I*, *here*, and *now*, the so-called essential indexicals (Perry, 1979), rely on this form of indicating.
3. *Conspicuous events*. When Alan and Beth hear an unexpected noise, Alan can ask, "What is *that*?" Or, when Beth says, "I'd like a bowl of *vichyssoise*," he can ask, "How do you spell *that*?" with confidence that Beth will see that he is referring to the most conspicuous unspellable word in her utterance.
4. *Appendages*. People can point with wooden or laser pointers, using them as extensions of their arms—as appendages. They can also direct attention by ringing a doorbell, or by telephoning or paging someone.

Most forms of directing-to are parts of composite signals. When Alan says, "It's *me*" on the telephone, he refers to himself with his voice, which is an index to himself, plus the conventional word *me*, a symbol that refers to the person indexed by the voice. Alan's "me"

is a composite indication-plus-description, as are the other examples of directing-to.

### *Placing-for*

Another technique for indicating things is by *placing* them *for* others (Clark, in press). When Alan places money on the ticket counter of a cinema for the ticket-seller to take, he is indicating the money as "payment for a ticket he is buying." And when the ticket-seller places the ticket on the counter for him, she is indicating the ticket as "the ticket he is now buying." People can also indicate themselves by placing themselves for others. When a waiter places a bowl of *vichyssoise* in front of Beth, he is indicating it as "what she ordered." But when he places himself next to Beth, he is indicating himself as "a waiter waiting for her order."

### *Interpreting Indications*

Interpreting even the simplest indication is complex. When Alan points at the copy of *Moby Dick* and asks, "Have you ever read that?" he is drawing Beth's attention to a *perceptually conspicuous site*. Yet he is referring not to the site itself, or to the physical book, but to *any* printed edition of *Moby Dick*. His reference takes a chain of indexes: (a) his finger is an index to the site of the book; (b) the site is an index to that copy of *Moby Dick*; and (c) that copy is an index to any edition of *Moby Dick*.

A major challenge is to say how Alan designs his composite signal—"that" plus his pointing—and how Beth creates the right chain of indexes. With the same gesture, he could have referred to the physical book ("That is torn"), the intangible story of *Moby Dick* ("That is such an exciting novel"), Herman Melville ("He was born in 1819 in New York City"), or even the publisher ("They publish such great novels"). Each requires a different chain of indexes.

In face-to-face conversation, indicating is everywhere, as people point, place things,

orient and place themselves, or simply speak up (for indicating *I*, *here*, and *now*). To indicate and to understand indicating, people must consult mental representations of the space around them, objects in that space, and things physically or causally connected with those objects.

### *Demonstrating*

In demonstrating an object (a thing, event, state, or property), people create an icon that resembles it perceptually. A demonstration is really a *selective depiction* of the object (Clark, 1996; Clark & Gerrig, 1990), and most are created by two *depictive techniques*.

1. *Modeling*. Alan can denote a telephone by forming his right hand into the shape of a telephone (making a fist with thumb and pinkie extended). This could also be called *sculpting*. Another form of modeling is *enacting*, as when Alan denotes a person jumping by playing the role himself and jumping.
2. *Sketching*. Alan denotes a round plate by drawing a circle in the air with his finger. This form of sketching could be called *tracing*. Another form is *delimiting*, as when Alan denotes the length of a fish by placing the flat palms of his hands the right distance apart.

These techniques are often used in combination. Alan can denote a person telephoning by forming a telephone with his hand (sculpting) and placing it to his ear (enacting). Speakers can use these techniques with their hands, bodies, faces, and voices.

Demonstrations with the hands or arms are called *iconic gestures*, or *illustrators* (Ekman & Friesen, 1969; Goodwin, 1981; Kendon, 1980; McNeill, 1992; Schegloff, 1984). They have three main stages: (1) *preparation*; (2) *stroke*, the meaningful portion of the gesture; and (3) *recovery*. They can be timed very precisely with speech, and are generally



associated with *intonation units*, the apex of the stroke coinciding with an accented syllable. Iconic gestures tend to begin about 1.0 seconds before the words they go with and to last about 1.5 seconds beyond them (Butterworth & Beattie, 1978; Morrel-Samuels & Krauss, 1992). People can also demonstrate with their faces, as they model sympathetic grimaces, disappointed faces, or thinking faces (Bavelas, Black, Lemery, & Mullett, 1986; Goodwin & Goodwin, 1986).

Most demonstrations with the voice come in the form of *quotations* (Clark & Gerrig, 1990). In the following example, Kate is telling friends about being in the hospital on an intravenous (I-V) system (Polanyi, 1989):

(24) I went out of my mind and I just screamed. I said, "Take that out! That's not for me!" ... And I shook this I-V and I said, "I'm on an I-V, I can't eat. Take it out of here!"

In her two quotations Kate does more than say the words. She enacts an angry person by shouting the words and pretending to shake an I-V. Therefore, although some quotations enact merely what someone said, many enact what someone did. Most quotations are not verbatim, nor are they intended to be (Tannen, 1989; Wade & Clark, 1993). They can be created even for speechless entities, as the following example shows (Clark & Gerrig, 1990):

(25) The problem is this [the speaker holds up ring finger] will say, "I'm gonna curl," and then this guy [the pinkie] will say, "Yeah, I'm gonna curl too!" But then it goes "Aaaaigh!"

Some quotations are all gesture, as in "The kid went [rude gesture] and ran away," where the gesture is a type of quotation.

Most demonstrations are performed during speech, yet they can bear several relations to the speech:

1. *Embedded parts*. Quotations, whether speech or gestures, are embedded parts of utterances or the discourse. In sentence 25,

the quotation "I'm gonna curl" is the direct object of the verb *say*, and the sentence would not be complete without it.

2. *Composite parts*. Many demonstrations are parts of composite signals. For example, when Alan points at a book and says, "Have you ever read that?" he creates a composite signal—a description (*that*) plus indication (pointing). The same is true for many iconic gestures. When a woman, Fran, was telling a story about the film *Some Like it Hot*, she extended her arms overhead while saying, "and the girl jumps up" (Kendon, 1980). She created a composite of a demonstration (the gesture) plus a description ("jumps up").

3. *Independent signals*. Some demonstrations are independent of the utterance or discourse being produced. They are neither embedded nor composite parts.

4. *Self-talk*. Some demonstrations are not performed for the addressees, but for the speakers themselves. When solving a problem by themselves, people sometimes gesture to help them think about objects, events, and relations in the problem.

There has been much debate about the communicative role of demonstrations. In one view (Krauss, Morrel-Samuels, & Colasante, 1991; Rauscher, Krauss, & Chen, 1996; Rimé & Schiaratura, 1991), most iconic gestures are self-talk. Speakers are more likely to gesture when they have difficulty retrieving a word (Morrel-Samuels & Krauss, 1992), and they are more likely to be hindered in describing scenes when their hands are immobilized (Bilous, 1992; Krauss, 1998; Rimé, Schiaratura, Hupet, & Ghysselelinckx, 1984). This evidence favors a self-directed role for iconic gestures.

Still, most iconic gestures probably *are* communicative (Kendon, 1987, 1994). All quotations, whether speech, gestures, or a combination, are embedded parts of utterances, so they are communicative (see also

Kita, 1997). Most iconic gestures carry information not carried in the associated words, and listeners register this information as part of what is communicated (Engle, 1998, 2000; McNeill, 1992). Speakers use few iconic gestures when their addressees cannot see them, treating most of the gestures as being for their addressees (Cohen & Harrison, 1973). Speakers use as many iconic gestures in retelling a story as they do in telling it for the first time. They do so even though they no longer have trouble retrieving words (Beattie & Coughlan, 1998; see also 1999; Beattie & Shovelton, 1999). Additionally, narrators tell better stories when their addressees can react with iconic gestures, which both parties treat as part of their communication (Bavelas, Coates, & Johnson, 2000).

In summary, signals mean what they do by a combination of three methods: describing-as, indicating, and demonstrating. Describing-as is an immense *memory retrieval process*. Speakers and listeners have up to 100,000 lexical entries in their mental lexicons, which they consult about five times a second in the course of a normal utterance. Indicating, in contrast, is a process of *spatial cognition*. For each indication, speakers and addressees must consult representations of the space around them, locate objects in it, and find connections among them. Demonstrating, finally, is a process of *depicting and imagining appearances*. With each demonstration, speakers and their addressees must call on their knowledge of what things look or sound like and imagine a thing from the features of the demonstration.

Speaking and listening, therefore, cannot be reduced to words, phrases, and sentences. A close look at any face-to-face conversation reveals it to be an intricate mix of describing-as, indicating, and demonstrating, not only with language (e.g., words, speech timing, and quotations), but also with gestures (e.g., emblems, pointing or placing, and iconic gestures). People talk with their entire body.

## REPRESENTATIONS OF DISCOURSE

People carrying out basic joint activities need to represent those activities and update their representations as they go along. When Alan buys batteries from Beth at the drug store counter, the two of them start with their initial common ground and update the current state of their activity as they proceed. These two representations, the initial common ground and the current state of the activity, have been investigated under such headings as situational models, mental models, scripts, and schemas. The trouble is that most investigations have focused on narratives, especially written ones, so the picture is incomplete at best.

### Visual and Spatial Representations

When people communicate, they not only describe, but also indicate and demonstrate. As we just noted, indicating requires speakers and listeners to represent the surrounding space and the objects in it, and demonstrating requires them to represent what things look, sound, and feel like. The very act of communicating demands that people create and update the visual and spatial representations of what they are discussing. A large body of evidence shows that they do.

### Spatial Relations

Bransford, Barclay, and Franks (1972) produced a classic demonstration of spatial relations. As part of their experiment, people read either sentence 26a, b, c, or d, and were asked to remember it.

- (26) a. Three turtles rested on a floating log and a fish swam beneath it.  
 b. Three turtles rested on a floating log and a fish swam beneath them.  
 c. Three turtles rested beside a floating log and a fish swam beneath it.  
 d. Three turtles rested beside a floating log and a fish swam beneath them.

The scenes described in 26a and 26b are alike spatially, for if a fish swam beneath the log ("it"), it also swam beneath the turtles ("them"). However, the scenes described in 26c and 26d are not alike spatially, for if a fish swam beneath the log, it did not necessarily swim beneath the turtles. Later, in a multiple-choice test with the four sentences in random order, people who had seen 26a often chose 26b by mistake, but those who had seen 26c rarely chose 26d by mistake. Readers must have represented not the sentence *per se*, but a visual or spatial representation of the scene described.

People consult visual and spatial representations to interpret even single words, such as *approach* in these three descriptions:

- (27) a. I am standing on the porch of a farmhouse looking across the yard at a picket fence. A tractor [or: mouse] is just *approaching* it.  
 b. I am standing across the street from a post office with a mailbox in front of it. A man crossing the street is just *approaching* the post office [or: mailbox].  
 c. I am standing at the entrance to an exhibition hall looking at a slab of marble. A man is just *approaching* it with a camera [or: chisel].

In an experiment by Morrow and Clark (1988), people were given one of the two alternatives of these and other descriptions and were asked to estimate the distance of, say, the tractor, or mouse, from the picket fence. The following table gives the average estimates of those distances:

- (27') a. tractor to fence, 39 feet; mouse to fence, 2 feet  
 b. man to post office, 28 feet; man to mailbox, 13 feet  
 c. man with camera to marble slab, 18 feet; man with chisel to marble slab, 5 feet

Apparently people arrive at a denotation for *approach* by considering how near an object must be to a landmark in order to be in "interaction with it" for its assumed purpose. These judgments depend on the size of the referent object (as in 27a), the size of the landmark (27b), and the purpose of the person approaching (27c).

These findings shouldn't be surprising, and they are just a sample of a large literature on such effects. They remind us that listeners need visual and spatial imagination for even the simplest descriptions. They need to imagine the appearance or arrangement of turtles, logs, tractors, mice, and fences to come to the right interpretations.

### *Point of View*

Most stories are told from a narrator's or protagonist's point of view. In Mark Twain's *Tom Sawyer*, Tom Sawyer is the protagonist, and a separate third-person narrator tracks his location as he moves from place to place. In Mark Twain's *Huckleberry Finn*, Huck Finn is both protagonist and first-person narrator, so when *he* moves from place to place, he describes what he sees from his own perspective. To represent point of view, readers must represent Tom's and Huck's immediate surroundings and their location in it. We surely do not represent these surroundings as fully or vividly as we do our own, but we need at least some representation of that space.

Tracking a first-person narrator requires following a *deictic center*—the *I*, *here*, and *now* of the narrator's point of view. This is especially important for interpreting *deictic expressions*. These are expressions whose interpretation depends on the speaker's or addressee's current point of view. Examples include: *come and go*, *this* and *that*, *here* and *there*, *this side* and *the other side*, *in front of* and *behind*, and *left of* and *right of* (see Bühler, 1982; Duchan, Bruder, & Hewitt, 1995; Fillmore, 1975; Levinson, 1996). In Hemmingway's *The Killers*, the narrator opens his story this way:

- (28) The door to Henry's lunchroom opened and two men came in.

As Fillmore (1981) noted, the narrator must be inside the lunchroom, because he describes the door as opening by unseen forces and the

men as "coming" in, not "going" in. The deictic center is inside the room. Point of view is essential to many of the narrator's choices, and imagining the scene from the narrator's or protagonist's vantage point is crucial to getting that point of view right.

Abrupt changes in point of view require abrupt changes in the imagined spatial representation, and these are sometimes difficult to perform. In a demonstration by Black, Turner, and Bower (1979), people read simple descriptions such as the two following examples:

- (29) Bill was sitting in the living room reading the paper, when John came [or: went] into the living room.
- (30) Alan hated to lose at tennis. Alan played a game of tennis with Liz. After winning, she came [or: went] up and shook his hand.

We, as readers, can think of point of view in sentences 29 and 30 by setting up a camera to view the scenes. For the first clause in 29, we would set it up in the living room and leave it there when John "comes" in. This is not the case when John "goes" in, for the camera would need to start out of the living room and then follow John into the room. In sentence 30, the camera would be near Alan for the first two sentences, so it would not need to be moved when Liz "comes" up to him. It *would* need to be moved when she "goes" up to him, following Liz when she moved. Changing point of view (as with "went" in 29 and 30) should be disruptive to understanding, as the study showed. Participants took longer to read the passages with the changed points of view, and were also more likely to recall them incorrectly (see also Bruder, 1995).

People keep track of changing points of view even without deictic expressions. In an experiment by Morrow (1985), people memorized the layout of a small model house and then read brief narratives about it, one sentence at a time. One narrative ended in

these two sentences about Kathy's movements, which were followed by a question:

- (31) She walked from the study into the bedroom. She didn't find the glasses in the room. Which room is referred to?

For different people, the first sentence had different prepositions (*from* vs. *through* vs. *past the study*, and *into* vs. *to the bedroom*) and different verb modalities (*walked* vs. *was walking*). All these differences influenced which room people inferred to be the referent of *the room* in the second sentence. The following are the results of two variants (in percent of choices by the participants):

- (32) She walked *from* the study *into* the bedroom  
The room referred to: the bedroom, 77%;  
the study, 21%; other rooms, 2%
- (33) She walked *past* the study *to* the bedroom  
The room referred to: the bedroom 21%;  
the study 73%; other rooms, 6%

In sentence 32, most people took Kathy to be in the bedroom, but in 33, most took her to be near the study. To figure out where Kathy was in 32 or 33, people had to consult their representation of the model house and, against that, interpret the combination of *walked*, *from* or *past the study*, and *into* or *to the bedroom*. There are no clear answers as to how they did that.

People also track larger features of the spatial surroundings (Bower & Morrow, 1990; Morrow, Bower, & Greenspan, 1989; Morrow, Greenspan, & Bower, 1987). In a study by Glenberg, Meyer, and Lindem (1987), people were given paragraphs to read, one sentence at a time. Some read one of the two versions of 34:

- (34) Warren spent the afternoon shopping at the store. He picked up [or: set down] his bag and went over to look at some scarves. He had been shopping all day. He thought it was getting too heavy to carry.

The pronoun *it* in the last sentence refers to the bag mentioned in the second sentence. When

the verb in the second sentence is *picked up*, Warren keeps the bag with him as he looks at the scarves, but when the verb is *set down*, he leaves the bag behind. In this study, the bag's location was important to the interpretation of the pronoun. People read the final sentence a full 0.6 seconds faster when the verb was *picked up* than when it was *set down*. The assumption is that they could readily locate the referent for *it* when the bag was still with Warren, but they could not locate the referent as readily when Warren did not have the bag. Participants must have been consulting a spatial representation in determining the referent.

Deploying spatial representations in discourse is, therefore, complicated. To make these judgments, people need to represent the protagonist's surroundings and keep track of where the protagonist is. To do that, they must consult their common ground with the writer, including their practical knowledge of houses, department stores, acts of walking, and other common items and events. They must combine this with information from the descriptions, such as the verb (*walked*), the prepositional phrases (*from the study* and *into the bedroom*), and other items (*the bag*). The issue is how people combine such disparate sources of information to arrive at their understanding (see Glenberg, Kruley, & Langston, 1994).

### *Mental Maps*

It is often assumed that people consult mental maps of their homes, neighborhoods, and cities as they travel through them. If so, do they create and consult the maps in using language? In a classic study by Linde and Labov (1975), people were asked, "Could you tell me the layout of your apartment?" Almost all responded by taking the questioner on an imaginary tour, as in this example:

- (35) You walk in the front door.  
 There was a narrow hallway.  
 To the left, the first door you came to was a tiny bedroom.

Then there was a kitchen,  
 and then bathroom,  
 and then the main room was in the back, living  
 room, I guess.

They would begin at the front door and describe a tour that passed through each room precisely once. Apparently, they imagined someone ("you") taking the tour, for they described landmarks in relation to the tourist's instantaneous positions with such deictic terms as *to the left* and *straight ahead* (see Ehrich & Koster, 1983; Levelt, 1982; Shanon, 1984; Ullmer-Ehrich, 1987). Descriptions like these are *route descriptions*. Only a few people gave *survey descriptions*, in which they described the scene from a bird's eye's view—as a mental map. In these cases, they located landmarks with absolute terms such as *to the north of* and *next to*. Apparently, it was more natural to describe apartments with route than with survey descriptions.

How do listeners *understand* route and survey descriptions? In studies by Taylor and Tversky (1992, 1996; see also Perrig & Kintsch, 1985), people read either a route or survey description of a small town and were then asked inferential questions from either a route perspective or a survey perspective. People were just as fast at answering questions from either perspective regardless of whether they had read the route or survey description. Apparently, they created representations of the town that were independent of the type of description they read. These representations must be more than simple maps in the head, for they allow people to jump back and forth between route and survey perspectives almost at will.

### **Schemas and Mental Models**

People appear to have special cognitive tools for narrating stories or conversing about everyday activities. These tools include schemas and scripts, mental models, and mental simulations.

### Schemas

In the early 1900's, psychologists developed the notion of *schema* to account for how people understand and remember stories. A schema is a set of cultural preconceptions about causal or other types of relationships—part of *communal* common ground. In the classic experiments by Bartlett (1932), people were told a Native American folk story, "The War of the Ghosts," which included many elements unfamiliar to Western norms. In retelling that story, people often distorted it to fit their cultural expectations, such as changing "hunting seals" into "fishing," a more likely pastime according to their schema.

Schemas of a different type were proposed for the structure of stories themselves. According to one account (Rumelhart, 1975), stories consist of a setting followed by an episode; an episode consists of an event plus a reaction to it; a reaction consists of an internal response plus an external response; and so on. Listeners are assumed to parse stories into these functional sections much as they parse sentences into constituents. A rather different account (Labov, 1972) is that narratives of personal experience have six parts:

1. An *abstract*, briefly summarizing the story;
2. An *orientation*, a stage setting about the who, when, what, and where of the story;
3. A *complicating action*;
4. An *evaluation* of these actions;
5. The *result or resolution* of the complicating action; and
6. A *coda*, a signal of completion.

Narrators and their audiences are assumed to refer to such schemas in producing and understanding stories.

A third class of schemas, known as *scripts*, was proposed as representations for events (Schank & Abelson, 1977). The idea is that scripts guide people's expectations about the

presence and order of everyday events. When we go to a restaurant, our "restaurant script" informs us that we need to order from a menu, wait for our food, and pay at the end. When we hear a description about going to a restaurant, we appeal to the same script. Even if not explicitly told, we assume that the protagonist ordered food and paid the bill in the proper order (Bower, Black, & Turner, 1979). If we are told that the events occurred in an unusual order, such as the protagonist paid before ordering his or her food, we may recall the events in their usual order because that fits our "restaurant script." Scripts are part of communal common ground, so they vary with the community. The restaurant script in North America is strikingly different from those in Greece and Japan.

Schemas were designed, then, to explain how people could have a mental representation of a narrative that is more detailed than the original. People take the limited input and, by applying schemas, elaborate on it in appropriate ways.

### Mental Models

Whereas schemas represent cultural preconceptions, mental models are mental constructions in which people represent specific objects, events, and relationships in utterances or narratives (Johnson-Laird, 1983; Garnham & Oakhill, 1996). They are mental instantiations of the world being described. People create mental models based on the discourse, the situation, and the purposes they have to serve. So, people trying to understand "Three turtles rested on a floating log and a fish swam beneath it" create mental models of ponds, logs, fish, and turtles so that they can estimate where they are in relation to each other. People trying to interpret *approach* in "The tractor approached the fence" create mental models of the scene described in order to judge where tractor and fence must be. Mental models begin, in effect, with the generic information represented in schemas (in *communal*

common ground), and add visual and spatial relationships to represent instantiations of a scene or event (in *personal* common ground).

Mental models can also represent dynamic events. If a person is asked how many windows there are in his or her house, that person is likely to imagine him- or herself walking around the house counting the windows—a dynamic process (Shepard & Cooper, 1982). According to Hegarty (1992; Hegarty, Just, & Morrison, 1988), people understand diagrams of pulleys in much the same way—through dynamic mental models (see also Gentner & Stevens, 1983). These seem eminently suited for representing the dynamic course of events people consult in telling and understanding narratives.

### *Mental Simulations*

Mental simulations were proposed by Kahneman and Tversky (1982) as a type of dynamic mental model in which people can modify the initial settings of the model and compare the outcomes. People might simulate a process for many purposes: (a) to predict its outcome, (b) to assess its probability, (c) to assess counterfactual alternatives (“if only . . .”), and (d) to project the effects of causality. When people simulate alternative endings to a story, for example, they tend to make “downhill changes” to scenarios, that is they remove unusual or unexpected aspects of the situation. They rarely make “uphill changes,” or changes that introduce unusual aspects, and never make “horizontal” changes, or changes that alter arbitrary aspects. Mental simulations represent the process of *imagining* working through an event.

Mental simulations are well suited for imaginary experiences (see Davies & Stone, 1995), and these include *emotional* experiences. When people go back over fatal accidents of loved ones, they often experience guilt, anger, or regret as they mentally simulate alternatives for those accidents—as they think “If only she hadn’t driven down that

street,” or “What if he had left two minutes earlier?” (Kahneman & Tversky, 1982). Mental simulations require the active participation of the participants, and they introduce a boundary between reality and the simulation (taking the system “off-line” and feeding it counterfactual inputs). However, many of the specifics of mental simulation have yet to be tested experimentally.

### **Fictional Worlds**

The evidence in this chapter has focused on people’s reliance upon visual, spatial, schematic, and scriptal representations for the actual world—apartment layouts, visits to restaurants, personal experiences. However, people need something more when the situations are fictional, and that “something” is called *joint pretense*. People engage in a simple pretense whenever they act *as if* they were doing something they are not actually, really, or seriously doing at that moment (Goffman, 1974). An example is lying. Liars act *as if* they were actually, really, or seriously claiming what they appear to be claiming. Fiction, however, requires a *joint* pretense, when two people coordinate on the pretense, mutually aware that they are doing so.

The prototype of joint pretense is the game of make-believe (Clark, 1996; Walton, 1978, 1983, 1990). Suppose that Ned and Kenneth, both 5 years old, are pretending to be lion and lion-tamer. To succeed, they must coordinate their imaginings. They must simulate the way a lion and lion-tamer would behave toward each other. They must both imagine the back yard as a circus ring, the back porch as a lion cage, and much, much more. In playing their game, they are simultaneously engaged in two *layers of joint action*.

**Layer 1:** Ned and Kenneth are playing a game of make-believe, jointly pretending to be taking the actions at layer 2.

**Layer 2:** Ned and Kenneth are a lion and lion-tamer performing in a circus.

The domain of layer 1 is the real or actual world. The domain of layer 2 is a fictional world. Both are part of Ned and Kenneth's current common ground.

All fiction requires two or more layers of joint action (Bruce, 1981; Clark, 1996; Currie, 1990; Walton, 1978, 1983, 1990). Take the first lines of a joke told by Sam to Reynard (Svartvik & Quirk, 1980):

- (35) let me tell you a story, - - -  
 a girl went into a chemist's shop, and asked for,  
 contraceptive tablets, - -  
 so he said "well I've got . all kinds, and . all  
 prices, what do you want,"  
 she said "well what have you got,"

There are three layers to this example. In layer 1, the real or actual world, Sam is announcing the story to Reynard, "Let me tell you a story." In layer 2, a fictional world, a reporter is telling a friend about a conversation between a chemist (a pharmacist) and a young woman. The quotation in line 3 shows the third layer, the world of the chemist speaking to the young women.

To participate in this joke, Sam and Reynard must engage in joint pretense. When Sam produces "A girl went into a chemist's shop and asked for contraceptive tablets," he is asking Reynard to join with him in pretending that an actual reporter is telling an actual friend about a young woman going into a chemist's shop. Crucially, the deictic center changes with each layer. In layer 1, *I* and *you* are Sam and Reynard; in layer 2, *I* and *you* are the reporter and his friend; and in layer 3, *I* and *you* are the chemist and the young woman. Reynard cannot interpret *me* and *you* in line 1, *went* in line 2, or *I* and *you* in line 3 without keeping track of these layers. The same goes for many other features as well.

Joint pretense is valuable because it allows participants to have vivid experiences in

the safety of imagination. Ned and Kenneth, the two 5-year-olds, play their game because it is exciting to imagine living in the circus and to simulate experiences they could not have in the actual world. Reynard's understanding of the joke becomes more exciting and more vivid, when he can imagine an actual chemist saying, "Well I've got all kinds, and all prices, what do you want?" Novels, jokes, and short stories are a mixture of telling and showing—of *diagesis*, or description, and *mimesis*, or demonstration. As novelist David Lodge (1990) noted, "[The] alternation of authorial description and characters' verbal interaction remains the woof and warp of literary narration to this day."

### *Imaginal Props*

Novels, jokes, and short stories aren't the only venues for fictional language. There are also stage plays, radio plays, operas, operettas, puppet shows, films, television situation comedies, soap operas, film cartoons, comic books, songs, and pantomimes. Many narratives have appeared in several media. Jane Austen's *Emma* comes as a novel, audio recording, and film, and it could probably be produced as a radio play, comic book, stage play, and opera. These forms are not all alike. They range in how they engage our imagination—and in how effectively they do that.

Imaginal props are one device for engaging the imagination. Imaginal props are devices that support the *imagining* of a situation. They are engineered to get the audience engrossed in a fictional world (Clark & Van Der Wege, 2001), such as the following examples demonstrate:

1. *Quotation.* In sentence 35, Sam quotes the chemist as saying, "Well, I've got all kinds and all prices." If he delivers the line well, Sam can help Reynard imagine, or experience, not only the chemist's point of view,



but also his accent and sympathy. Even the barest quotations add vividness.

2. *Iconic and deictic gestures.* In spontaneous stories, speakers often use iconic and deictic gestures to depict and point to things in the fictional world (Haviland, 1996; McNeill, 1992). In an example discussed earlier (Kendon, 1980), Fran tells an anecdote from the film *Some Like it Hot*. At one point she says, "They wheel a big table in [sweeping her arm to depict the motion], with a big with a big [1.08 sec] cake on it [tracing a horizontal circle to depict its shape], and the girl [raises arm to depict jumping up], jumps up." The gestures clarify what she is saying and make the fictional scene that much more vivid.
3. *Enactments.* When a stage actor is Hamlet, or a film actress is Emma, they do more than recite their lines. They enact their characters. When we read *Emma*, we work hard to imagine what Emma looks like—her hair, clothing, and mannerisms. Without a background in 19th-century English style, we may get it wrong. In seeing the film *Emma*, we are *shown* what she looks like, including her hair, clothing, mannerisms, and what she does. All we must imagine is that this particular actress (say, Gwyneth Paltrow) is in fact Emma.
4. *Staging.* Stage plays, films, operas, and comic books rely on staging. The production crew engineers the scenery, scene changes, timing, close-ups, and other features to help engross the audience in the right fictional world. Nothing kills imagination like bad production.
5. *Sound effects.* It may seem that the greater the verisimilitude of the imaginal prop, the better the aid to imagination, but that isn't always true. In Verdi's opera, Aida's singing is hardly realistic speech, yet it helps us create her happy or melancholy

moods in fictional Egypt. The same goes for background music in films. Excitement, suspense, sadness, and other moods would be harder to create without it.

Imaginal props are tricks of the fictional trade. In the hands of the best writers, storytellers, film directors, and actors, they help engross us in the fictional world. The issue is how.

### *Experiential Representations*

When we get engrossed in a story, we often experience emotions (see Gross & Levenson, 1995). Consider what Walton (1978) calls *quasi-fear*. When we see a horror film, we are afraid of what the monster will do to the heroine. Our hearts beat faster, our muscles tighten, and our knuckles turn white as the monster approaches her. But do we warn her as we would if all this were happening in front of us? This is what makes it quasi-fear and not real fear. Next, consider what Gerrig (1989a, 1993; see also Gerrig & Prentice, 1991; Prentice & Gerrig, 1999; Prentice, Gerrig, & Bailis, 1997) called *anomalous suspense*. Ordinarily, suspense is a state in which we "lack knowledge about some sufficiently important target outcome (p. 79)." Yet, as Gerrig demonstrated in a series of experiments, when we read suspense stories, we often feel suspense even when we know how they turn out. Like Walton's quasi-fear, we compartmentalize our experience as part of the story world and separate from the actual world.

Most narratives are engineered to elicit emotion. Novels are classified into genres largely by the emotions they evoke. Mysteries lead to suspense and fear; adventures to excitement, fear, and elation; horror stories to horror, loathing, and fear; light romances to sexual excitement; heavier romances to erotic arousal; satires to amusement; and so on. Films are classified in much the same way.

We imagine story worlds as if we were now experiencing them before our very eyes. At the same time, we recognize that we are still in the actual world.

How, then, do people represent fiction? A complete answer must account for at least four phenomena (Clark & Van Der Wege, 2001):

1. *Experience.* People experience selective features of the narrative world as if they were actual, current experiences. These include visual appearances, spatial relations, points of view, movement and processes, voices, and emotions.
2. *Imaginal props.* People's imaginings appear to be aided by well-engineered imaginal props, such as direct quotation, gestures, stage sets, sound effects, and background music.
3. *Participation.* Speakers and writers design what they say to encourage certain forms of imagination, but listeners and readers must cooperate with them to succeed.
4. *Compartmentalization.* In participating in narratives, people distinguish their experiences in the story world from their experiences in the real world.

It isn't enough to posit visual or spatial representations, schemas, scripts, mental models, and even mental simulations. It takes layering and joint pretense to account for these four phenomena. However, more investigation is needed to determine how they work.

## SUMMARY

In the century since Wundt, psycholinguistics has come a long way. First, there have been breakthroughs in research methods. In Wundt's time, one could investigate written language, or slips of the tongue, as Freud did, but not much more. With advances in technology, investigators now exploit audio- and

video recordings, both for laboratory experiments and for the analysis of spontaneous conversation. They can measure reaction times, analyze and synthesize speech sounds, and track brain activity. Second, there have been breakthroughs in theory. Since Wundt's time, linguistic theory has become a highly sophisticated, if controversial, area of study. Researchers have also developed major theories for speaking, parsing, speech perception, language acquisition, reading, and brain activation, to name just a few areas. All of these theories stand on a foundation of strong empirical results.

Still, the essence of language use is found in face-to-face talk. It is here that speaking and listening arise in their natural, universal states. It is here that researchers can study why speakers say the things they say and how listeners interpret these things—ultimately, as a way of coordinating joint activities. It is here that researchers can study people's use of common ground—in everything from identifying words like *candy* to drawing elaborative inferences. It is here that researchers can study how speakers combine description, indication, and demonstration to say what they say. The problem is that too little is known about spontaneous language and how it differs from reciting, reading aloud, listening to idealized speech, and other such forms. Understanding language in its natural habitat is a major challenge for the second century of psycholinguistics.

## REFERENCES

- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30(3), 191–238.
- Arnold, J., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of complexity and information structure on constituent ordering. *Language*, 76, 28–55.

- Atkinson, J. M., & Heritage, J. (Eds.). (1984). *Structures of social action: Studies in conversation analysis*. Cambridge: Cambridge University Press.
- Austin, J. L. (1962). *How to do things with words*. Oxford: Oxford University Press.
- Bach, K., & Harnish, R. M. (1979). *Linguistic communication and speech acts*. Cambridge: MIT Press.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Bavelas, J. B., Black, A., Lemery, C. R., & Mullett, J. (1986). "I show you how you feel": Motor mimicry as a communicative act. *Journal of Personality and Social Psychology*, 50, 322-329.
- Bavelas, J. B., & Chovil, N. (2000). Visible acts of meaning. An integrated message model of language in face-to-face dialogue. *Journal of Language and Social Psychology*, 19, 163-194.
- Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, 15, 469-489.
- Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality & Social Psychology*, 79(6).
- Beattie, G., & Coughlan, J. (1998). Do iconic gestures have a functional role in lexical access? An experimental study of the effects of repeating a verbal message on gesture production. *Semiotica*, 119(3-4), 221-249.
- Beattie, G., & Coughlan, J. (1999). An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. *British Journal of Psychology*, 90(1), 35-56.
- Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language & Social Psychology*, 18(4), 438-462.
- Behaghel, O. (1909/1910). Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen*, 25, 110-142.
- Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279-352). New York: Wiley.
- Bilous, F. R. (1992). *The role of gestures in speech production: Gestures enhance lexical access*. Unpublished Ph.D. dissertation, Columbia University.
- Black, J. B., Turner, T. J., & Bower, G. H. (1979). Point of view in narrative comprehension. *Journal of Verbal Learning and Verbal Behavior*, 18, 187-198.
- Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945-984). San Diego: Academic Press.
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, 11, 177-220.
- Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science*, 247(4938), 44-48.
- Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive vs. interpretive approach. *Cognitive Psychology*, 3, 193-209.
- Brown, P., & Levinson, S. C. (1987). *Politeness*. Cambridge: Cambridge University Press.
- Bruce, B. (1981). A social interaction model of reading. *Discourse Processes*, 4, 273-311.
- Bruder, G. A. (1995). Psychological evidence that linguistic devices are used by readers to understand spatial deixis in narrative text. In J. F. Duchan, G. A. Bruder, & L. E. Hewitt (Eds.), *Deixis in narrative: A cognitive science perspective* (pp. 243-260). Hillsdale NJ: Erlbaum.
- Buchler, J. (Ed.). (1940). *Philosophical writings of Peirce*. London: Routledge and Kegan Paul.
- Bühler, K. (1934). *Sprachtheorie: die Darstellungsfunktion der Sprache*. Jena: Fischer.
- Bühler, K. (1982). The deictic field of language and deictic words. In R. J. Jarvella & W. Klein (Eds.), *Speech, place, and action* (pp. 9-30). Chichester, England: Wiley.
- Butterworth, B., & Beattie, G. (1978). Gestures and silence as indicators of planning in speech. In

- R. Campbell & P. T. Smith (Eds.), *Recent advances in the psychology of language: Formal and experimental approaches*. NATO Conference Series III:4B (pp. 347–360). New York: Plenum.
- Button, G., & Lee, J. R. (Eds.). (1987). *Talk and social organisation*. Philadelphia: Multilingual Matters.
- Caramazza, A., & Grober, E. (1976). Polysemy and the structure of the subjective lexicon. In C. Rameh (Ed.), *Georgetown University round table on language and linguistics* (pp. 181–206). Washington, DC: Georgetown University Press.
- Chafe, W. (1979). The flow of thought and the flow of language. In T. Givon (Ed.), *Syntax and semantics 12: Discourse and syntax* (pp. 159–181). New York: Academic Press.
- Chafe, W. (1980). The deployment of consciousness in the production of a narrative. In W. Chafe (Ed.), *The pear stories* (pp. 9–55). Norwood, NJ: Ablex.
- Chantraine, Y., & Hupet, M. (1994). Efficiency of the addressee's contribution to the establishment of references: Comparing monologue with dialogues. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, 13(6), 777–796.
- Chomsky, N. (1957). *Syntactic structures*. 's-Gravenhage: Mouton.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, 55, 430–477.
- Clark, H. H. (1977a). Bridging. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: Readings in cognitive science* (pp. 411–420). Cambridge: Cambridge University Press.
- Clark, H. H. (1977b). Inferences in comprehension. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, NJ: Erlbaum.
- Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive psychology*, 11, 430–477.
- Clark, H. H. (1983). Making sense of nonce sense. In G. B. Flores d'Arcais & R. Jarvella (Eds.), *The process of language understanding* (pp. 297–331). New York: Wiley.
- Clark, H. H. (1994). Discourse in production. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 985–1021). San Diego: Academic Press.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H. (1998). Communal lexicons. In K. Malmkjær & J. Williams (Eds.), *Context in language learning and language understanding* (pp. 63–87). Cambridge: Cambridge University Press.
- Clark, H. H. (in press). Pointing and placing. In S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet*. Hillsdale, NJ: Erlbaum.
- Clark, H. H., & Brennan, S. A. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspective on socially shared cognition* (pp. 127–149). Washington, DC: APA.
- Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.
- Clark, H. H., & Fox Tree, J. E. (2001). Using uh and um in spontaneous speaking, Stanford University. Manuscript.
- Clark, H. H., & Gerrig, R. J. (1983). Understanding old words with new meanings. *Journal of Verbal Learning & Verbal Behavior*, 22(5), 591–608.
- Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, 66, 764–805.
- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Hillsdale, NJ: Erlbaum.
- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge: Cambridge University Press.
- Clark, H. H., & Schaefer, E. R. (1989). Contributing to discourse. *Cognitive Science*, 13, 259–294.
- Clark, H. H., & Schunk, D. H. (1980). Polite responses to polite requests. *Cognition*, 8, 111–143.
- Clark, H. H., & Van Der Wege, M. (2001). Imagination in discourse. In D. Schiffrin &

- D. Tannen (Eds.), *Handbook of discourse*. Oxford: Blackwell.
- Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37(3), 201-242.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.
- Cohen, A. A., & Harrison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology*, 28, 276-279.
- Currie, G. (1990). *The nature of fiction*. Cambridge: Cambridge University Press.
- Dahan, D., Swingle, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, 42(4), 465-480.
- Davies, M., & Stone, T. (Eds.). (1995). *Mental simulation*. Oxford: Blackwell.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283-321.
- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, 104(1), 123-147.
- Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review*, 98(4), 604-614.
- Drew, P., & Heritage, J. (Eds.). (1992). *Talk at work: Interaction in institutional settings*. Cambridge: Cambridge University Press.
- Duchan, J. F., Bruder, G. A., & Hewitt, L. E. (1995). *Deixis in narrative: A cognitive science perspective*. Hillsdale, NJ: Erlbaum.
- Ehrich, V., & Koster, C. (1983). Discourse organization and sentence form: The structure of room descriptions in Dutch. *Discourse Processes*, 6(2), 169-195.
- Ekman, P., & Friesen, W. (1969). The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica*, 1, 49-98.
- Elman, J. L. (1989). Connectionist approaches to acoustic/phonetic processing. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 227-260). Cambridge: MIT Press.
- Engle, R. A. (1998). Not channels but composite signals: Speech, gesture, diagrams, and object demonstrations are integrated in multimodal explanations. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Engle, R. I. (2000). *Toward a theory of multi-modal communication: Combining speech, gestures, diagrams, and demonstrations in instructional explanations*. Unpublished doctoral dissertation, Stanford University, CA.
- Erikson, T. A., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20, 540-552.
- Erman, B. (1987). *Pragmatic expressions in English: A study of you know, you see, and I mean in face-to-face conversation*. Stockholm: Almqvist & Wiksell International.
- Fillenbaum, S. (1971). On coping with ordered and unordered conjunctive sentences. *Journal of Experimental Psychology*, 87, 93-98.
- Fillenbaum, S. (1974). Pragmatic normalization: Further results for some conjunctive and disjunctive sentences. *Journal of Experimental Psychology*, 102, 574-578.
- Fillmore, C. (1975). *Santa Cruz lectures on deixis*. Bloomington, IN: Indiana University Linguistics Club.
- Fillmore, C. (1981). Pragmatics and the description of discourse. In P. Cole (Ed.), *Radical pragmatics* (pp. 143-166). New York: Academic Press.
- Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 134-184). Cambridge: Cambridge University Press.
- Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, 62(2), 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(2), 280-295.
- Francik, E. P., & Clark, H. H. (1985). How to make requests that overcome obstacles to compliance. *Journal of Memory and Language*, 24, 560-568.
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27-52.
- Fromkin, V. A. (Ed.). (1973). *Speech errors as linguistic evidence*. The Hague, Netherlands: Mouton.
- Fussell, S. R., & Krauss, R. M. (1992). Co-ordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality & Social Psychology*, 62(3), 378-391.
- Garnham, A., & Oakhill, J. V. (1996). The mental models theory of language comprehension. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 313-339). Hillsdale, NJ: Erlbaum.
- Garrett, M. F. (1980). Syntactic processes in sentence production. In B. Butterworth (Ed.), *Speech production* (pp. 170-220). New York: Academic Press.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2), 181-218.
- Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3), 181-215.
- Garrod, S. C., & Sanford, A. J. (1982). Bridging inferences in the extended domain of reference. In A. Baddeley & J. Long (Eds.), *Attention and performance IX* (pp. 331-346). Hillsdale, NJ: Erlbaum.
- Garrod, S. C., & Sanford, A. J. (1994). Resolving sentences in a discourse context: How discourse representation affects language understanding. In M. A. Gernsbacher & et al. (Eds.), *Handbook of psycholinguistics* (pp. 675-698). San Diego: Academic Press.
- Gee, J. P. (1986). Units in the production of narrative discourse. *Discourse Processes*, 9, 391-422.
- Gentner, D., & Stevens, A. (Eds.). (1983). *Mental models*. Hillsdale, NJ: Erlbaum.
- Gerrig, R. J. (1989a). Suspense in the absence of uncertainty. *Journal of Memory and Language*, 28(6), 633-648.
- Gerrig, R. J. (1989b). The time course of sense creation. *Memory & Cognition*, 17(2), 194-207.
- Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. New Haven: Yale University Press.
- Gerrig, R. J., & Bortfeld, H. (1999). Sense creation in and out of discourse contexts. *Journal of Memory and Language*, 41(4), 457-468.
- Gerrig, R. J., & Gibbs, R. W. (1988). Beyond the lexicon: Creativity in language production. *Metaphor & Symbolic Activity*, 3(1), 1-19.
- Gerrig, R. J., & Prentice, D. A. (1991). The representation of fictional information. *Psychological Science*, 2(5), 336-340.
- Gibbs, R. W. (1979). Contextual effects in understanding indirect requests. *Discourse Processes*, 2, 1-10.
- Gibbs, R. W. (1983). Do people always process the literal meanings of indirect requests? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 524-533.
- Gibbs, R. W. (1986). What makes some indirect speech acts conventional? *Journal of Memory and Language*, 25(2), 181-196.
- Gibbs, R. W., & Mueller, R. A. (1988). Conversational sequences and preference for indirect speech acts. *Discourse Processes*, 11(1), 101-116.
- Glenberg, A. M., Kruley, P., & Langston, W. E. (1994). Analogical processes in comprehension: Simulation of a mental model. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 609-640). San Diego: Academic Press.
- Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, 26(1), 69-83.
- Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. Garden City, NY: Anchor Books.
- Goffman, E. (1974). *Frame analysis*. New York: Harper and Row.

- Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.
- Goodwin, C. (1986a). Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies*, 9, 205-217.
- Goodwin, C. (1986b). Gestures as a resource for the organization of mutual orientation. *Semiotica*, 62, 29-49.
- Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62, 51-75.
- Gordon, D., & Lakoff, G. (1971). Conversational postulates, *Papers from the seventh regional meeting of the Chicago Linguistic Society* (pp. 63-84). Chicago: Chicago Linguistic Society.
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of language* (pp. 58-90). Cambridge: MIT Press.
- Grice, H. P. (1957). Meaning. *Philosophical Review*, 66, 377-388.
- Grice, H. P. (1968). Utterer's meaning, sentence-meaning, and word-meaning. *Foundations of Language*, 4, 225-242.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3: Speech acts* (pp. 113-128). New York: Seminar Press.
- Grice, H. P. (1978). Some further notes on logic and conversation. In P. Cole (Ed.), *Syntax and semantics 9: Pragmatics* (pp. 113-127). New York: Academic Press.
- Grice, H. P. (1991). *In the way of words*. Cambridge: Harvard University Press.
- Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition & Emotion*, 9(1), 87-108.
- Haviland, J. B. (1996). Projections, transpositions, and relativity. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 271-323). Cambridge: Cambridge University Press.
- Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 1084-1102.
- Hegarty, M., Just, M. A., & Morrison, I. R. (1988). Mental models of mechanical systems: Individual differences in qualitative and quantitative reasoning. *Cognitive Psychology*, 20(2), 191-236.
- Holmes, V. M., Stowe, L., & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language*, 28(6), 668-689.
- Hupet, M., Seron, X., & Chantraine, Y. (1991). The effects of the codability and discriminability of the referents on the collaborative referring procedure. *British Journal of Psychology*, 82(4), 449-462.
- Jefferson, G. (1972). Side sequences. In D. Sudnow (Ed.), *Studies in social interaction* (pp. 294-338). New York: Free Press.
- Johnson, H. G., Ekman, P., & Friesen, W. V. (1975). Communicative body movements: American emblems. *Semiotica*, 15(4), 335-353.
- Johnson, M. K., Bransford, J. D., & Solomon, S. K. (1973). Memory for tacit implications of sentences. *Journal of Experimental Psychology*, 98(1), 203-205.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Harvard University Press.
- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In P. Slovic, D. Kahneman, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201-208). Cambridge: Cambridge University Press.
- Kendon, A. (1967). Some functions of gaze direction in two-person conversation. *Acta Psychologica*, 16, 22-63.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *Relationship of verbal and nonverbal communication* (pp. 207-227). Amsterdam: Mouton de Gruyter.

- Kendon, A. (1987). On gesture: Its complementary relationship with speech. In A. W. Siegman, & S. Feldstein (Eds.), *Nonverbal behavior and communication* (2nd ed., pp. 65–97). Hillsdale, NJ: Erlbaum.
- Kendon, A. (1994). Do gestures communicate? A review. Special Issue: Gesture and understanding in social interaction. *Research on Language and Social Interaction*, 27(3), 175–200.
- Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics*, 35(2[348]), 379–415.
- Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, 7(2), 54–60.
- Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, 61(5), 743–754.
- Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1(5), 113–114.
- Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality & Social Psychology*, 4(3), 343–346.
- Krauss, R. M., & Weinheimer, S. (1967). Effect of referent similarity and communication mode on verbal encoding. *Journal of Verbal Learning and Verbal Behavior*, 6(3), 359–363.
- Labov, W. (1972). The transformation of experience in narrative syntax. In W. Labov (Ed.), *Language in the inner city: Studies in the Black English vernacular* (pp. 354–396). Philadelphia: University of Pennsylvania Press.
- Lakoff, R. (1973). The logic of politeness; or, minding your p's and q's. *Papers from the ninth regional meeting of the Chicago Linguistics Society* (pp. 292–305). Chicago: Chicago Linguistics Society.
- Lehmann, W. P. (1972). On converging theories in linguistics. *Language*, 48, 266–275.
- Lehmann, W. P. (1973). A structural principle of language and its implications. *Language*, 49, 47–66.
- Levelt, W. J. M. (1982). Cognitive styles in the use of spatial direction terms. In R. J. Jarvella & W. Klein (Eds.), *Speech, place, and action: Studies in deixis and related topics* (pp. 251–268). Chichester, England: Wiley.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.
- Levelt, W. J. M. (1989). *Speaking*. Cambridge: MIT Press.
- Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, 14(1), 78–106.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, 22(1), 1–75.
- Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechman, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, 98(1), 122–142.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, S. C. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 108–169). Cambridge: MIT Press.
- Levinson, S. C. (2000). *Presumptive meanings*. Cambridge: MIT Press.
- Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge: Harvard University Press.
- Lewis, D. K. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, 8, 339–359.
- Linde, C., & Labov, W. (1975). Spatial networks as a site for the study of language and thought. *Language*, 51(4), 924–939.
- Lodge, D. (1990). Narration with words. In H. Barlow, C. Blakemore, & M. Weston-Smith (Eds.), *Images and understanding* (pp. 141–153). Cambridge: Cambridge University Press.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703.



- Mann, W. C., & Thompson, S. A. (1986). Relational propositions in discourse. *Discourse Processes*, 9, 57-90.
- Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8(1), 1-71.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. Special Issue: Spoken word recognition. *Cognition*, 25(1-2), 71-102.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1-86.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99(3), 440-466.
- McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.
- Merritt, M. (1976). On questions following questions (in service encounters). *Language in Society*, 5, 315-357.
- Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(3), 615-622.
- Morris, D., Collett, P., Marsh, P., & O'Shaughnessy, M. (1979). *Gestures: Their origins and distribution*. New York: Stein & Day.
- Morrow, D. G. (1985). Prepositions and verb aspect in narrative understanding. *Journal of Memory and Language*, 24, 390-404.
- Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language*, 28(3), 292-312.
- Morrow, D. G., & Clark, H. H. (1988). Interpreting words in spatial descriptions. *Language and Cognitive Processes*, 3, 275-291.
- Morrow, D. G., Greenspan, S. E., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language*, 26, 165-187.
- Murray, J. D., Klin, C. M., & Myers, J. L. (1993). Forward inferences in narrative text. *Journal of Memory and Language*, 32(4), 464-473.
- Paek, T. S. Y. (2000). *Formal and computational methods for achieving mutual understanding in conversations between humans and computers*. Unpublished doctoral dissertation, Stanford University, CA.
- Perrig, W., & Kintsch, W. (1985). Propositional and situational representations of text. *Journal of Memory and Language*, 24(5), 503-518.
- Perry, J. (1979). The problem of the essential indexical. *Nous*, 13(1), 3-21.
- Polanyi, L. (1989). *Telling the American story*. Cambridge: MIT Press.
- Prentice, D. A., & Gerrig, R. J. (1999). Exploring the boundary between fiction and reality. In S. Chaiken, Y. Trope & et al. (Eds.), *Dual-process theories in social psychology* (pp. 529-546). New York: Guilford Press.
- Prentice, D. A., Gerrig, R. J., & Bailis, D. S. (1997). What readers bring to the processing of fictional texts. *Psychonomic Bulletin & Review*, 4(3), 416-420.
- Prince, E. F. (1981). Towards a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223-256). New York: Academic Press.
- Rapp, D. N., & Gerrig, R. J. (1999). Eponymous verb phrases and ambiguity resolution. *Memory & Cognition*, 27(4), 612-618.
- Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7(4), 226-231.
- Reder, L. M., & Cleeremans, A. (1990). The role of partial matches in comprehension: The Moses illusion revisited. In A. C. Graesser & G. H. Bower (Eds.), *The psychology of learning and motivation: Inferences and text comprehension* (pp. 233-258). San Diego: Academic Press.
- Rimé, B., & Schiaratura, L. (1991). Gesture and speech. In R. S. Feldman, & B. Rimé (Eds.), *Studies in emotion & social interaction* (pp. 239-281). New York: Cambridge University Press.
- Rimé, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, 8, 311-325.

- Rumelhart, D. E. (1975). Notes on schemas for stories. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 211–236). New York: Academic Press.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50, 696–735.
- Sanford, A. J., & Garrod, S. C. (1994). Selective processing in text understanding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 699–719). San Diego: Academic Press.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.
- Schegloff, E. A. (1972). Notes on a conversational practice: Formulating place. In D. Sudnow (Ed.), *Studies in social interaction* (pp. 75–119). New York: Free Press.
- Schegloff, E. A. (1980). Preliminaries to preliminaries: "Can I ask you a question?" *Sociological Inquiry*, 50, 104–152.
- Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk. Georgetown University Roundtable on Languages and Linguistics, 1981* (pp. 71–93). Washington, DC: Georgetown University Press.
- Schegloff, E. A. (1984). On some gestures' relation to talk. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 262–296). Cambridge: Cambridge University Press.
- Schegloff, E. A. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organization. In G. Button & J. R. E. Lee (Eds.), *Talk and social organization* (pp. 70–85). Clevedon, UK: Multilingual Matters.
- Schegloff, E. A. (1988). Presequences and indirection: Applying speech act theory to ordinary conversation. *Journal of Pragmatics*, 12(1), 55–62.
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361–382.
- Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica*, 8, 289–327.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21(2), 211–232.
- Schourup, L. C. (1982). *Common discourse particles in English conversation*. New York: Garland.
- Searle, J. R. (1969). *Speech Acts*. Cambridge: Cambridge University Press.
- Searle, J. R. (1975a). Indirect speech acts. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 59–82). New York: Seminar Press.
- Searle, J. R. (1975b). A taxonomy of illocutionary acts. In K. Gunderson (Ed.), *Minnesota studies in the philosophy of language* (pp. 334–369). Minneapolis: University of Minnesota Press.
- Shanon, B. (1984). Room descriptions. *Discourse Processes*, 7, 225–255.
- Shepard, R. N., & Cooper, L. A. (Eds.). (1982). *Mental images and their transformations*. Cambridge: MIT Press.
- Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24–49). Cambridge: MIT Press.
- Singer, M. (1994). Discourse inference processes. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 479–515). San Diego: Academic Press.
- Singer, M., & Halldorson, M. (1996). Constructing and validating motive bridging inferences. *Cognitive Psychology*, 30(1), 1–38.
- Slobin, D. I. (1996). From "Thought and Language" to "Thinking for Speaking." In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge: Cambridge University Press.
- Sperber, D., & Wilson, D. (1986). *Relevance*. Cambridge: Harvard University Press.

- Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics 9: Pragmatics* (pp. 315-332). New York: Academic Press.
- Svartvik, J., & Quirk, R. (Eds.). (1980). *A corpus of English conversation*. Lund, Sweden: Gleerup.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-660.
- Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge: MIT Press.
- Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, 18(4), 427-440.
- Tanenhaus, M. K., & Spivey-Knowlton, M. J. (1996). Eye-tracking. *Language & Cognitive Processes*, 11(6), 583-588.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.
- Tannen, D. (1989). *Talking voices: Repetition, dialogue and imagery in conversational discourse*. Cambridge: Cambridge University Press.
- Taylor, H. A., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, 31(2), 261-292.
- Taylor, H. A., & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language*, 35(3), 371-391.
- Ullmer-Ehrich, V. (1982). The structure of living space descriptions. In R. J. Jarvella & W. Klein (Eds.), *Speech, place, and action: Studies in deixis and related topics* (pp. 219-249). Chichester, England: Wiley.
- Underhill, R. (1988). Like is, like, focus. *American Speech*, 63(3), 234-246.
- Van den Broek, P. (1994). Comprehension and memory of narrative texts: Inferences and coherence. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 539-588). San Diego: Academic Press.
- Venneman, T. (1973). Explanation in syntax. In J. Kimball (Ed.), *Syntax and semantics* (Vol. 2, pp. 1-50). New York: Seminar Press.
- Venneman, T. (1975). An explanation of drift. In C. N. Li (Ed.), *Word order and word order change* (pp. 269-305). Austin: University of Texas Press.
- Wade, E., & Clark, H. H. (1993). Reproduction and demonstration in quotations. *Journal of Memory and Language*, 32(6), 805-819.
- Walton, K. L. (1978). Fearing fictions. *Journal of Philosophy*, 75, 5-27.
- Walton, K. L. (1983). Fiction, fiction-making, and styles of fictionality. *Philosophy and Literature*, 8, 78-88.
- Walton, K. L. (1990). *Mimesis as make-believe: On the foundations of the representational arts*. Cambridge: Harvard University Press.
- Wason, P. C., & Reich, S. S. (1979). A verbal illusion. *Quarterly Journal of Experimental Psychology*, 31, 591-597.
- Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change*, 9, 81-105.
- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31(2), 183-194.
- Wundt, W. M. (1900). *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte*. Leipzig: Englemann.
- Yngve, V. H. (April 1970). *On getting a word in edgewise*. Paper presented at the sixth regional meeting of the Chicago Linguistic Society, Chicago, IL.
- Zwitzerlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32(1), 25-64.