CHAPTER 5

Discourse and Dialogue

SUSAN E. BRENNAN, ANNA K. KUHLEN, AND JEANNE CHAROY

INTRODUCTION

Discourse is language used in social context—typically, utterances or sentences, connected in paragraphs or stories, whether expressed as spoken monologues or written texts. *Dialogue* is discourse that unfolds in a coordinated fashion between two or more people as they interact (whether in spoken conversation, over the telephone, or conducted via e-mail or another of social media's many textual formats). Both discourse and dialogue, whether the medium is text or speech, are produced with addressees in mind (explicitly or implicitly):

> Addressivity, the quality of turning to someone, is a constitutive feature of the utterance; without it the utterance does not and cannot exist. (Bakhtin, 1986, p. 99)

Language scientists focus at many different grains of analysis such as sound, word, and sentence in order to achieve clarity and control in their experiments. However, it is worth keeping in mind that "in the wild," language use and processing occur within discourse and dialogue contexts that shape how these smaller units of language scale up:

> Language is a temporal phenomenon, a process that flows through time. That is partly because time is an essential ingredient of sound, but more importantly it is because thoughts flow through time as well, and language is first and foremost a way of organizing and communicating this flow of thoughts. It is futile to limit our attention to isolated sentences. The shape a sentence takes can never be appreciated without recognizing it as a small, transient slice extracted from the flow of language and thought, when it has not simply been invented to prove some point. (Chafe, 2002, p. 256)

An utterance plucked out of context is ambiguous, whereas within its natural dialogue context, this is much less often the case. Both discourse and dialogue recruit the planning, creation, integration, interpretation, and grounding of linguistic elements whose meanings depend on extra-linguistic context and knowledge (for discussion, see Graesser, Gernsbacher, & Goldman, 2003). Relevant context and knowledge may include any and all aspects of the situation at hand, including the current goals of the participants (be they speakers, addressees, or bystanders—or in the case of books and other text formats, writers, readers, or even characters); the identities of the participants and the presumed common ground that exists between them; the genre,

This material is based on work done while SB was serving at the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Fourth Edition, edited by John T. Wixted.

DOI: 10.1002/9781119170174.epcn305

or conventions, associated with the situation or tasks at hand; the medium within which discourse is conducted; and in the case of conversational exchanges in spoken dialogue, nonverbal aspects such as gesture, prosody, and multimodal information about a partner's attention, intention, emotion, and actions.

These definitions cover quite a bit of human social and cognitive activity. This chapter aims to describe the nature of discourse and dialogue by highlighting their important features and discussing some of the most enduring and influential models and results. We will do this primarily from an experimental psycholinguistic perspective, but with an eye to social-interactional, computational, and neuroscience approaches. *Discourse and dialogue are inherently multidisciplinary topics*; for a deep understanding, it is necessary to consider approaches from multiple fields of investigation. Sociolinguistic or ethnomethodological approaches such as those taken by *conversation analysts* (e.g., Goodwin, 1979, 1981; Jefferson, 1973; Sacks, Schegloff, & Jefferson, 1974) can serve as useful starting points, as they yield a wealth of detailed and descriptive data about talk-in-interaction, or language use as it arises in (and is inseparable from) social context. This approach is data driven (rather than hypothesis driven) and the findings often resist generalizing. *Discourse analysts* tend to look for and count details of particular interest within a particular genre of text or spoken dialogue, in order to identify distributions of forms and sometimes to relate them to functions in a more generalizable way. *Psycholinguists* attempt to uncover the mechanics of language processing, sometimes by conducting experiments on one subject at a time in laboratory settings that strip away social context in the interest of gaining experimental control. These experimental data are used to test principled models and generalize about linguistic processes and representations. Some psycholinguists retain social context in their studies of spoken dialogue, staging tasks with two or more people communicating, in order to test theories of language use and processing in parallel with interpersonal coordination as the participants do a meaningful task together (also in a laboratory). Evidence from *communication neuroscience*, or studies of brains engaged in the cognitive and social aspects of language processing within communicative contexts, helps to answer questions about the biological architecture supporting language use that behavioral studies may not be able to answer alone. Finally, computational approaches can implement models of language use (often using findings from these other approaches) to create working text generation programs or spoken dialogue systems.

With these basics in mind, this chapter aims to present an introduction to the cognitive science of discourse and dialogue. First, we will discuss two dominant experimental traditions, and then how information is packaged within discourse and how meaning is achieved within dialogue. We will provide a tour of some classic issues, findings, and theories in discourse processing. Then we will survey studies of language use in communicative contexts that shed light on how people plan, co-create, interpret, and coordinate language use within dialogue. We will touch on experimental techniques used in studies of the psycholinguistics of discourse and dialogue, including behavioral measures, referential communication tasks, eye-tracking in visual worlds studies, the use of experimental confederates, and other measures of dynamic coordination such as cross-recurrence gaze analysis. We will describe relevant examples of spoken dialogue systems for human–machine interaction. Along the way, we will highlight aspects of an ongoing debate about audience design, or the extent to which processing

language by speakers and addressees in dialogue is adapted to a specific conversational partner (i.e., when and how speakers tailor utterances for their addressees, and addressees tailor interpretations of utterances with speakers in mind). The identities and roles of dialogue partners is a key part of taking the context of language use into account. We will also cover some practical applications of research in discourse and dialogue: writing for a reader's comprehension and improving robustness in human interaction with spoken dialogue systems. We will close by considering recent findings about dialogue alongside research on the cognitive- and social-neuroscientific underpinnings of language use, and outline some future directions.

BACKGROUND AND CLASSIC ISSUES

Experimental Traditions in Discourse and Dialogue

Within psycholinguistics, there have been two long-standing experimental traditions relevant to the study of discourse and dialogue: the *language-as-product* and *language-as-action* traditions (for discussion, see H. H. Clark, 1992; Trueswell & Tanenhaus, 1995). The first tradition tends to focus on information processing (Miller, 1963) and to look for the effects of linguistic representations on either comprehension *or* production (but not both at once), with the assumption (since Chomsky, 1957) that language is for thinking rather than for communicating. Typically within this tradition, solitary subjects are asked by an experimenter to respond to stimuli such as fragments of language or idealized sentences (in comprehension studies) or to name words or pictures or describe stimuli (in production studies) while choice and reaction time data are collected. Although using stimuli outside of communicative contexts makes it easier to maximize experimental control and although such studies can provide useful data about linguistic products and processes, the findings do not necessarily scale up to a complete picture of discourse processing, particularly as it occurs in dialogue.

This language-as-product tradition has largely dominated the field of discourse comprehension, which commonly has subjects read short text discourses made up of grammatical sentences written expressly for a given experiment. Such texts may or may not be engaging to the subjects, and the absence of a relevant goal (apart from the subject's desire to get through the experiment quickly) may affect how they are processed. In these investigations, the emphasis has not been on what people do with or experience from discourse, but rather on factors leading to *cohesion* (Halliday & Hasan, 1976) or *coherence* (e.g., Sanders & Pander Maat, 2006). Cohesion captures the surface continuity between one sentence and the next, signaled by phenomena such as pronouns and other expressions that co-specify referents; ellipsis; discourse cues such as *well*, *oh*, and *so*; and syntactic choices that mark given and new information. Coherence is when utterances or sentences are perceived as relevant or semantically related to the topic or goal at hand, such that the discourse makes sense. In theories of reading comprehension (see, e.g., Gernsbacher, 1996; Graesser et al., 2003; Zwaan & Radvansky, 1998), both cohesion and coherence are predicted to affect the ease with which readers link and integrate the information they read into a mental model of the discourse. In the language-as-product tradition, a discourse is largely equated with its text transcript, so cohesion and coherence are viewed as properties of the discourse.

The second long-standing tradition, *language-as-action*, tends to focus on processing and behavior in communicative

contexts, with the assumption that language is for communicating and for doing things in the world. This tradition was initially inspired by philosophers of language, notably Austin's essay *How to do things with words* (1962), Searle's *speech act theory* (1969), and Grice's *cooperative principle* (1975). Within psychology, the language-as-action tradition was pioneered by experimentalists such as H. H. Clark (1992, 1996; H. H. Clark & Wilkes-Gibbs, 1986); Krauss (1987); Bavelas (Bavelas, Chovil, Coates, & Roe, 1995); Tanenhaus (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995); and their colleagues. Consider the following segment of spoken dialogue in a seminar room with a dozen people sitting around a table:

Herb: ok
now, next week
<looks at the people around the table and makes eye contact with Susan>
Susan: ok, I will
Herb: right.

What happened here? A research group meeting was winding down; it was time to determine who would make a presentation during the next meeting. With just eight words packaged into three speaking turns, along with the judicious use of eye gaze, Herb and Susan came to an agreement that was recognized as such and experienced as coherent by everyone in the room. According to the language-as-action tradition, these speakers understood one another with ease and experienced this exchange as coherent because they communicated within a context of mutual knowledge known as *common ground* (H. H. Clark, 1992; H. H. Clark & Brennan, 1991; H. H. Clark & Schaefer, 1989; H. H. Clark & Wilkes-Gibbs, 1986). Common ground can come from *perceptual co-presence* (as when a speaker says *give me the cup*, and the addressee hands over the only cup on the table between them), *linguistic co-presence* (when one partner says *where did you get it?* with both of them resolving the pronoun *it* to the previously mentioned cup), and *community co-membership* (when the other partner answers *at the pottery sale on campus*, which she can expect that her addressee has a high probability of understanding because they both know that they're part of the same campus community; see H. H. Clark & Marshall, 1981).

Discourse and dialogue are often presented as text transcripts, made up of grammatical sentences (in the case of published text) or transcribed utterances that can include single words, ungrammatical fragments, and disfluencies (such as interrupted words, mid-phrasal pauses, restarts, and repairs). Transcripts of spontaneous spoken dialogue show incremental evidence of both understanding and misunderstanding. Consider this snippet:

Amanda: have you got a new job yet—
<pause>
Brad: I don't know
I haven't heard yet
Amanda: what from *<pause>*
Brad: sorry
Amanda: have you heard about your interview with thingy
Brad: no I haven't heard
Amanda: *<continues>*

(adapted from the London-Lund corpus, Svartvik & Quirk, 1980)

Here, Brad thinks he's answered Amanda's question in his first turn, but the evidence that follows shows that Amanda is not confident that he has understood her question. She initiates a repair with a fragment and after a pause, Brad appears to have misunderstood what she is asking. She reframes her question more specifically, but with a

proxy expression (*thingy*) that can only be understood via their previously established common ground. Brad finally produces an answer that she appears to find satisfactory (as she continues on with the dialogue). Remarkably, even small misunderstandings such as this one are not experienced as incoherent, as people in conversation can seek and provide evidence until they conclude that they understand one another well enough for current purposes (Brennan, 2005; H. H. Clark & Brennan, 1991). We will return to this process of *grounding* in more detail presently.

Discourse analysts who study dialogue rely on transcripts to excerpt examples, code for types of speech acts, identify referring expressions that co-specify the same discourse entities, and count elements from language corpora such as words uttered or speaking turns taken; from such data, they posit rules or principles or examine distributions and sometimes test hypotheses to account for the form of discourse and dialogue. Some discourse analysts use made-up prototypical examples as part of an explanation (as do many linguists who are concerned with explaining grammatical phenomena). Ethnomethodologists and sociolinguists, on the other hand, are more concerned with the natural settings in which spontaneous conversations take place; they transcribe dialogues as faithfully as possible to analyze very fine-grained aspects of interaction, identifying and describing key structural phenomena in conversation such as turn-taking, repair, and the collaborative construction of utterances (with little concern about counting and comparing; see Levinson, 1983, for discussion of the differences between discourse analysis and conversation analysis approaches). What is abundantly clear to those who have ever transcribed spontaneous conversation is that the currency of social interaction is not idealized, grammatical sentences like those found in edited texts, but utterances that may be quite disfluent and fragmented. Even so, conversation is still orderly in that utterances show recognizable contingency in both form and timing with what comes before and after, reflecting the exchange of evidence as two people ground meanings.

However informative it may be, *a transcript is only an artifact*. It is not equivalent to a discourse itself, but provides one sort of evidence about the cognitive and social processes from which it emerges. Psycholinguistic studies of dialogue typically invite pairs of subjects to the lab to do a collaborative task together; this gives experimenters the ability to monitor for physical evidence about what they mean, understand, and misunderstand, as the subjects look at and manipulate task-relevant objects while communicating (see, e.g., H. H. Clark, 1992; Glucksberg & Weisberg, 1966; Schober & Brennan, 2003). This is one critical way in which experimenters differ from ethnomethodologists, who seek rich descriptive results that may come at the expense of summarizability and causality and who tend to approach the conversations they analyze with an open mind (rather than with a hypothesis). We find that all of these approaches can be complementary; for instance, conversation analysis is a good source of insights that can be developed into hypotheses and then wrestled into the lab for testing. Experimental studies have uncovered underlying mechanisms for phenomena such as conversational repair, fillers (*um* or *uh*), silent pauses, interruptions, lexical entrainment, perspective-taking, distribution of initiative, and audience design, which involves tailoring an utterance (or the processing of an utterance) to a particular partner (as we will discuss presently).

It may be tempting to equate (or at least to associate) the language-as-product

tradition with text discourse and the language-as-action tradition with interactive spoken dialogue, but that would not be an accurate mapping. Psycholinguists have done far more experimental work on comprehension than on production, as well as far more on production of speech than of text; however, a few have emphasized the effects of feedback or of imagining an audience's perspective upon production processes in writing (e.g., Traxler & Gernsbacher, 1992, 1993, 1995), and such work falls into the language-as-action tradition. The action tradition is also well represented by work on reading fiction by Gerrig and colleagues, with its emphasis on engagement and participatory responses (Gerrig, 1993) and on common ground between authors and readers or between characters (Gerrig, Brennan, & Ohaeri, 2001), as well as by work on layered perspectives in discourse by H. H. Clark (1996).

From Words to Discourse

Language is a system that people use to create meanings; these meanings emerge through discourse and dialogue. Words are combined into phrases, phrases are structured into sentences or utterances, with sentences arranged in written paragraphs formatted on a printed page or screen and with utterances accumulating into stretches of speech delivered within prosodic contours. (For discussion focusing on word and sentence processing, see Chapters 3 and 4, respectively, in this volume.) Each word and each syntactic constituent can be associated with linguistic knowledge that is conventionalized and shared by a language community. However, words do not function as little containers of meaning; *the meanings achieved by virtue of combining these elements within discourse are quite different from a simple sum of the parts*.

The Role of World Knowledge

Much of what one takes away from a discourse is not explicitly stated. Consider the word *approach*; this word yields quite different interpretations depending on who is doing the approaching and with what purpose in mind, within the discourse context (Morrow & Clark, 1988):

> I am standing on the porch of a farm house looking across the yard at a picket fence. A *tractor/mouse* is just approaching it. (p. 282)

> I am sitting in a jeep looking out the window at a lion lying beneath a tree. A game warden is just approaching it with a *rifle/hypodermic needle*. (p. 285)

With a tractor approaching, readers estimated the distance to the picket fence as 39.2 feet, whereas with a mouse, they estimated the distance as 2.1 feet. When the game warden approached the lion with a rifle, the distance between them was estimated as 67.5 feet, and with a hypodermic needle, as 23.5 feet. The point is that every word within a discourse can interact with other elements and alter the situation model that a reader constructs from the linguist input using world knowledge.

Compositionality, a useful principle by which semanticists account for meaning based on systematically combining smaller elements of language into larger constituents, does not apply in any strictly formulaic or deterministic manner when it comes to discourse comprehension (Fernando, 2012; Ginzburg & Cooper, 2004). The inferences needed for a reader to understand discourse in the way an author intended are drawn from at least four kinds of input (van den Broek, Young, Tzeng, & Linderholmal, 1999): the text currently being read (or speech currently being heard), the immediately previous text (or speech), the mental model or episodic memory representation of the situation so far, and the

knowledge base of a particular reader (or hearer).

Much of the meaning that a reader derives from a text is not expressed explicitly, but is achieved through bridging inferences (Haviland & Clark, 1974). Consider the following pairs of sentences (each forming a minimal discourse):

> I looked into the room. The ceiling was very high.
>
> I walked into the room. The windows looked out into a garden.
>
> I walked into the room. The chandeliers sparkled brightly. (H. H. Clark, 1977, p. 251)

In each case, the definite noun phrase underlined in the second sentence co-specifies information evoked by the situation introduced in the first sentence. The inferences that bridge from these referring expressions to *the room* differ in how direct or predictable the relationship is—the inference is obligatory in the first example (all rooms have ceilings), highly probable in the second (many rooms have windows), and forced in the third (H. H. Clark, 1977). This means that some information is activated automatically (before it is needed) by virtue of its strong association with the words in a discourse, whereas other information is computed only as needed; in the words of Lewis (1979), "Say something that requires a missing presupposition, and straightaway that presupposition springs into existence, making what you said acceptable after all" (p. 339). Readers expect that a text is intended by its author to be coherent (Grice, 1975), and so they tend to make the inferences that provide the best explanation for what they have read (a process known as *abduction*; see Hobbs, Stickel, Appelt, & Martin, 1988).

Much of the knowledge that readers use routinely to fill in the missing details and make sense of a text can be represented as schemas. A schema is a knowledge structure or concept in memory that captures the common attributes of a typical situation that has been experienced repeatedly; once evoked, schemas rapidly activate expectations and associations that make a situation easy to process and support the inferences needed to understand a text. An individual word (such as *approach* or *room*) can evoke a schema. As a discourse unfolds, slots within a schema can be filled with prototypical values (defaults), or else with varying information (variables) (Rumelhart, 1975); this makes the process of interpretation both efficient and flexible. Consider this four-sentence discourse (Rumelhart, 1979, p. 79):

> Business had been slow since the oil crisis. Nobody seemed to want anything really elegant anymore. Suddenly the door opened and a well-dressed man entered the showroom. John put on his friendliest and most sincere expression and walked toward the man.

The first sentence evoked a gas station for most readers, who reported discarding that schema after the second sentence (as a gas station schema is inconsistent with elegance). By the third sentence, readers reported considering a car dealership schema, with *the well-dressed man* in the customer slot; the fourth sentence confirmed that schema, with *John* filling the salesman slot. Experiments such as Rumelhart's made the point that discourse processing is incremental, and that for a text to be understood and experienced as coherent, constituents must be integrated into a mental representation or discourse model of the situation being described. When readers cannot evoke a schema, comprehension and memory for text is poor; those who read a detailed description of a situation or procedure understood and recalled it much better when they saw a meaningful title or graphical illustration beforehand than when they read it without

a title or illustration (Bransford & Johnson, 1972). Early studies of memory for text demonstrated that people recall the text they read (or hear) not as expressed verbatim, but consistent with (and distorted toward) the schemas evoked (e.g., Anderson, 1976; Bartlett, 1932; Sachs, 1967).

While early cognitive psychologists were conducting experiments about inferences in reading, artificial intelligence researchers were implementing models to create systems that could generate stories (Bobrow & Collins, 1975; Meehan, 1976; Schank & Abelson, 1977). Meehan's (1976) TALE-SPIN was programmed with information about simple characters and goals in order to support the automatic generation of fables. Here is one of its output stories:

> One day Joe Bear was hungry. He asked his friend Irving Bird where some honey was. Irving told him there was a beehive in the oak tree. Joe threatened to hit Irving if he didn't tell him where some honey was. (p. 127)

When Meehan added the proposition that beehives contain honey to TALE-SPIN's knowledge base, it generated this story:

> One day Joe Bear was hungry. He asked his friend Irving Bird where some honey was. Irving told him there was a beehive in the oak tree. Joe walked to the oak tree. He ate the beehive. (p. 128)

Early artificial intelligence researchers found that capturing the essential knowledge and inferences that come along with schemas (knowledge that readers deploy effortlessly) was more elusive than expected. Within the cognitive sciences, such computational efforts brought into sharp focus the complexities of human discourse processing.

The power and flexibility of inferences made during discourse processing—and made rapidly—is simply remarkable. In a study that measured neural activation to statements that either conformed to or violated lexical semantics (Nieuwland & van Berkum, 2006), readers rapidly integrated pragmatic information from fictional stories about inanimate objects that had emotions such that the readers did not show the typical N400 responses to statements that would be infelicitous in nonfictional contexts (e.g., *The girl comforted the clock* or *The peanut was in love*). The readers *did* show N400s to statements that ordinarily would not evoke this kind of response (e.g., *The peanut was salted*). We will discuss the neural basis for discourse and dialogue processing in the last section of this chapter.

Gricean Implicature

The topic of *pragmatics* focuses on the social context of language use. A powerful kind of pragmatic inference was captured by philosopher of language Paul Grice's influential *cooperative principle*. Grice proposed that speakers are rational, and as a result, conversations do not consist of disconnected remarks, but that speakers "make <their> conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which <they> are engaged" (Grice, 1975, p. 26). On this view, communication is by nature cooperative (note that this is true even when speakers are arguing, at least to the point where one of them gives up the intention to communicate and storms away). Grice further specified four maxims that underlie the indirect use of language in social interaction:

1. *Maxim of Quantity:*

   Make your contribution as informative as is required (for the current purposes of the exchange).

Do not make your contribution more informative than is required.

2. *Maxim of Quality:*

   Try to make your contribution one that is true.

   Do not say what you believe to be false.

   Do not say that for which you lack adequate evidence.

3. *Maxim of Relation:*

   Be relevant.

4. *Maxim of Manner:*

   Avoid obscurity of expression.

   Avoid ambiguity.

   Be brief (avoid unnecessary perspicuity).

   Be orderly. (Grice, 1975, pp. 26–27)

Many of the inferences that people make in conversation—known as conversational implicatures—can be explained by the presumption that speakers are following these maxims, or if they flout them, that they are doing so intentionally (that is, they intend their interlocutors to recognize that they are doing so). Consider the following:

Jeanne: Did Susan cook dinner last night?
Anna: Well, she placed a number of edible substances into a pot and then heated them until various chemical reactions took place.

This (made-up) example illustrates the notion of a standard (or particularized) implicature (see Levinson, 1983), where Anna violated the maxim of manner by going to some length to answer what Jeanne may have meant as a simple yes/no question (leading to inferences about Susan's questionable cooking skills). Sometimes an implicature is needed to bridge from one utterance to the next, as in this example adapted from Grice (1975):

Jeanne: I am out of gas.
Anna: There is a gas station around the corner.

The implication is that in order for Anna's response to be relevant, the station must be open and have fuel for sale; Anna would be rightfully surprised if she arrived there to find it boarded up and out of business. The next (sadly, naturally occurring) example comes from a *Time Magazine* story following the Hurricane Katrina tragedy, reporting an interview with a former employer of the then-FEMA head:

> "Yes. Mike Brown worked for me. He was my administrative assistant. He was a student at Central State University," recalls former city manager Bill Dashner. "Mike used to handle a lot of details. Every now and again I'd ask him to write me a speech. He was very loyal. He was always on time. He always had on a suit and a starched white shirt." (Fonda & Healy, 2005)

This speaker appears to be trying not to violate the maxim of quality, leading to an implicature that casts doubt on the qualifications of the individual in question. Such implicatures are shaped by an understanding of the context in which speakers and addressees find themselves. There are generalized implicatures too, in which context is argued to be unnecessary (Levinson, 1983). For instance, *I broke a finger last year* suggests that the speaker broke her own finger and is not a loan shark exacting revenge; *He's not unintelligent* suggests that he is not exactly intelligent either; *Anna has two children* suggests that she has two and only two.

Grice's cooperative principle and maxims have been used by many to explain phenomena about indirect language use, such as irony, politeness, and humor. However, the maxims have been argued to be culturally bound, to interact in unpredictable ways, and to be difficult to interpret when interlocutors have dueling goals. Grice's cooperative principle and maxims do not constitute a psychological

theory or model from which clear predictions can be made, so we have described them here rather than in the upcoming section about models of discourse and dialogue.

Linguistic Variability: Every Difference Makes a Difference

Not only does a given word evoke different associations and inferences in different situations, but speakers and writers have many expressive choices concerning what words to use, how to package information into grammatical forms, and what perspectives to take in a particular context of language use. The average high school graduate may know as many as 60,000 words (or twice as many if they are avid readers according to Pinker, 1994; such estimates are difficult to verify and depend, of course, on what is counted as a word). The point is that with such abundance in an individual's mental lexicon, the potential for variability in word choice in discourse is simply enormous.

The Vocabulary Problem

Such variability in word choice was dubbed *the vocabulary problem* by Furnas, Landauer, Gomez, and Dumais (1987), who were at the time trying to explain why it was so difficult for software design engineers to anticipate what words people would generate spontaneously while interacting with a command language interface to an unfamiliar software application (note that this investigation was conducted before personal computing popularized the direct manipulation of graphics and icons). At that time, typing in the wrong term for a command resulted in failure (and a cryptic error message). The researchers asked people to guess the name of the command to use for removing a file. Guesses included *remove*, *delete*, *erase*, *expunge*, *kill*, *omit*, *destroy*, *lose*, *change*, *rid*, and *trash*, with the likelihood that any two people would produce the same term for the same function ranging from only 7% to 18% (Furnas et al., 1987).[1] Designing command languages to accept multiple synonyms for the same command was proposed as a solution (e.g., Good, Whiteside, Wixon, & Jones, 1984), but even allowing as many as 20 synonyms for a single function did not guarantee success; the likelihood of two people choosing one or two of the allowable synonyms for a given function was only about 80% (Furnas et al., 1987). Moreover, allowing synonyms led to additional problems: In a text editor with only 25 commands, the likelihood that two people who used the same term were actually mapping it onto the same function was only 15% (Furnas et al., 1987).

The vocabulary problem is not unique to command languages; linguistic forms are even more variable when words are combined into syntactic constituents (Winograd, 1971). A group of computational linguists who were developing a natural language interface to a database query application made this point when they tried to list all possible variations of a query asking for *the set of programmers working for department managers*, using common words and syntax. Before they abandoned this enterprise, they managed to list no fewer than 7,000 well-formed queries (for a handful of examples, see Figure 5.1), all of which would seem perfectly natural in some discourse settings and less so in others.

Information Packaging and Flow

The structure of spontaneous spoken discourse (e.g., as reflected in a text transcript) reflects the thought processes that generate it, providing clues about the cognitive processing of information. With the *Pear Stories* project, Chafe (1980) pioneered the technique of having multiple speakers

[1] Since the command language was hypothetical, the likelihood that two people would guess the same term was used as a conservative estimate for how often a typical user might correctly guess a command when using a real command language.

List programmers department managers supervise.
What programmers work for department managers?
List programmers working for department managers.
List programmers who work for department managers.
List any programmers department managers supervise.
List all programmers working for department managers.
List each programmer a department manager supervises.
Which programmers work for managers of departments?
Which programmers do department managers supervise?
List all programmers who work for department managers.
List all programmers that department managers supervise.
List programmers whose supervisors manage departments.
Which of the programmers work for department managers?
Who are the programmers department managers supervise?
List every programmer any department manager supervises.
List every programmer supervised by a department manager.
List programmers with supervisors who manage departments.
Which programmers are supervised by department managers?
Who are the programmers working for department managers?
List programmers whose supervisors are department managers.
List each programmer that any department manager supervises.
List all of the programmers who work for department managers.
Who are the programmers who work for department managers?
List every programmer whom a department manager supervises.
List each programmer who is working for a department manager.
Which programmers are there working for department managers?
Which of the programmers are department managers supervising?
Which of the programmers are working for department managers?
List each of the programmers supervised by a department manager.
List the programmers who are supervised by department managers.
Which of the programmers do managers of departments supervise?
Who are all of the programmers working for department managers?
Which of the programmers are supervised by department managers?
List any programmer whose supervisor is a manager of a department.
Who are the programmers being supervised by department managers?
Who are all of the programmers that department managers supervise?
List any programmers there might be working for department managers.
List everyone who is a programmer supervised by a department manager.
List each of the programmers who is supervised by a department manager.
Which of the programmers are being supervised by department managers?
List any programmer with a supervisor who is the manager of a department.
Who are the programmers whose supervisors are managers of departments?
Which of the programmers are being supervised by managers of departments?
Which of the programmers have supervisors who are managers of departments?
List any programmer who has a supervisor who is the manager of a department.
List all programmers who work for anyone who is the manager of a department.
List all programmers working for supervisors who are managers of departments.
List each of the programmers who is supervised by anyone managing a department.
Which of the programmers have supervisors who are the managers of departments?
Who are all of the programmers who have supervisors who are department managers?

Figure 5.1 Excerpt from *7000 Variations on a Single Sentence*, the Hewlett Packard Natural Language Project.
SOURCE: Brennan (1990, p. 400).

describe the same film clip in order to investigate the "mentalism," or cognitive processing, that underlies discourse. This influential approach uncovered both commonalities and variability in narrative forms, both in English and across a variety of other languages (including the native American languages studied by Chafe and his colleagues), as well as across different kinds of language situations (Chafe & Danielewicz, 1987). Such relationships between language and mind, or "how the flow of consciousness affects the flow and shape of language" (Chafe, 2002, p. 254), contrast with Chomsky's generative syntactic approach or with compositional approaches to combining words.

Many phenomena that discourse and dialogue psycholinguists and other analysts seek to explain concern the choices that speakers make, such as the highly variable forms of the logical query shown in Figure 5.1. These include lexical, syntactic, and prosodic choices that can express topic, emphasis, perspective, co-reference, and relationships among the elements under discussion. The text transcript of a discourse is composed of multiple utterances or sentences that result from such choices, in which information is packaged and linked by cues that support inferences on the part of readers or addressees, either implicitly or explicitly.

Given and New Information. Speakers, as well as good writers, typically mark information as *given* (already mentioned in the discourse context or known to the addressee) or *new* (e.g., the punchline or point of the sentence; see Chafe, 1976; H. H. Clark & Haviland, 1977; Halliday & Hasan, 1976; Haviland & Clark, 1974). In English, given (or thematic, or known) information tends to appear early in the sentence (e.g., as the sentential subject or initial modifying phrase), whereas new information tends to appear at the end (Bock, 1977). Given and new information can also be marked by other syntactic means. Although English is considered an SVO (subject-verb-object) language (Greenberg, 1963) with verbs typically preceded by sentential subjects and followed by sentential objects (as in *I love ice cream*), syntactic structure can be manipulated in order to package information effectively for a particular discourse context. Syntactic resources include marked syntactic structures such as fronting (*Ice cream, I love*), clefting (*It's ice cream that I love*, along with *ice cream* receiving prosodic stress), and extraposition (*It's obvious that I love ice cream*). Another resource is the choice of active versus passive voice. Although developing writers are often advised to prefer the active over the passive, the wisdom of following this advice depends on the discourse context. The passive voice, when used effectively, can achieve thematic continuity, allowing the writer to mark a discourse entity as given by expressing it as a grammatical subject even when it is not the agent of the verb's action. Moreover, agentless passives avoid any need to attribute responsibility for actions (desirable in some discourse contexts, as in *data were collected* or *mistakes were made*).

In spoken discourse, given and new information is marked intonationally; the first time a word is mentioned or read aloud, it is typically pronounced more clearly and longer or accented, whereas subsequent mentions are typically shortened and are less intelligible (this has been demonstrated using words excised from running speech; Bard et al., 2000; Fowler & Housum, 1987; Samuel & Troicki, 1998). Information that is predictable in a discourse is often attenuated as well. For instance, people were asked to read one of two variants of this short discourse aloud, with either *man* or *thief*:

The [man/thief] looked back as he ran.

The police were not far behind.

The readers pronounced *police* clearly when it followed *man*, but attenuated *police* after *thief* (Chase, 1995).

In spontaneous speaking, especially in question answering, where parallelism in syntax and wording is expected between a question and its answer, given information may be left out altogether, with new information mentioned as a form of ellipsis (H. H. Clark, 1979; Levelt & Kelter, 1982; Malt, 1985).

As readers or hearers incrementally integrate sentences or utterances into a discourse model, they must establish which referring expressions are co-referential as well as make the necessary plausible bridging inferences about entities relevant to previously mentioned information. Information marked as given anchors the rest of the sentence; it helps readers or hearers identify the new information expressed in the sentence or utterance and know where to associate it within the discourse model under construction. This is the "given-new contract," proposed by Haviland and Clark (1974). Cues about information status in discourse are not only produced by speakers and writers, but also interpreted by hearers and readers. For instance, an accented noun rapidly signals to the hearer (even before the entire word has been heard) that it co-specifies a new discourse entity (or else one being contrasted with another entity), whereas a de-accented noun rapidly signals that it is anaphoric with another expression, that is, given (Dahan, Tanenhaus, & Chambers, 2002).

Referring Expressions. Another way in which speakers (and writers) mark the information status of entities is in the forms of referring expressions. Definite expressions (e.g., nouns following the determiner *the* and proper names) are used to specify entities potentially identifiable within a discourse context, whereas indefinite noun phrases specify entities that are new (not yet in common ground) or not identifiable (do not specify a specific referent). When speakers introduce a new referent into a discourse, they tend to use a full noun phrase, and when they mention the same referent again, they tend to use a shortened form such as a pronoun (Ariel, 1990; Chafe, 1976; Grosz, Joshi, & Weinstein, 1995). A hierarchy for givenness was proposed by Gundel, Hedberg, and Zacharski (1993), with the following labels ranging from least restrictive or identifiable to most restrictive or identifiable (examples adapted from Kehler & Ward, 2006):

> I couldn't sleep last night. A dog kept me awake. *Type identifiable*
>
> I couldn't sleep last night. This dog kept me awake. *Referential (indefinite)*
>
> I couldn't sleep last night. The dog (next door) kept me awake. *Uniquely identifiable*
>
> I couldn't sleep last night. That dog (next door) kept me awake. *Familiar*
>
> My neighbor has a dog. This dog kept me awake last night. *Activated*
>
> A dog was in the front yard last night. It kept me awake. *In focus*

The differences in meanings among these discourses are not captured in propositional or semantic representations. Yet speakers and hearers are sensitive to information status as expressed by the forms of referring expressions, including with respect to how such information has entered the discourse. Referring expressions mark whether information is currently salient or else previously evoked in the discourse but not currently salient; whether it is brand-new or else new to the discourse but known to the addressee; whether it is known and presumed; and whether it is predictable and therefore deletable (see Prince, 1981, for discussion of these and other information

statuses). The effects of these variations can be measured. For instance, one might ask: *Which kind of expression is easier to interpret, a pronoun or a full noun phrase?* The answer is that it depends on whether the referent is already contextually salient (that is, in the center of attention), or not. Pronouns are easily to interpret for discourse entities that are already salient (even when the entities have not been explicitly mentioned in prior discourse; McKoon, Gerrig, & Greene, 1996). Interpretation is actually slowed when a full noun phrase is used to refer to an entity already in the center of attention (this has been called a "repeated name penalty" by Gordon, Grosz, & Gilliom, 1993), whereas full noun phrases are faster to read when the referent is not salient (Hudson, Tanenhaus, & Dell, 1986; Hudson-D'Zmura, 1988). We will return to these ideas in the section on models of discourse and dialogue, in the discussion of centering theory.

Perspective in Discourse: Personal, Temporal, Spatial.

> The speaker-listener actively involves himself with a sentence by "getting inside it." (MacWhinney, 1977, p. 152)

Even though people do not usually recall verbatim the exact wording of discourse (Sachs, 1967), speakers' syntactic and referential choices *do* affect its comprehension, via cohesion, information flow, and the pragmatic inferences that addressees make about what speakers or writers are referring to. Such choices also affect the sequence and ease with which a discourse model is constructed by a reader or listener. Sentences (and utterances) are generated from *perspectives* taken by speakers or writers, which they express in their choices about *person* (first, second, or third person), *semantic roles* (including the choice of whether to express an action with an explicit agent or as agentless), *verb tense*, *spatial perspective*, and *lexical perspective*.

It is well established that readers and listeners can keep track of the perspectives of the protagonists in discourse, as well as the spatial and temporal perspectives associated with events. Consequently, readers and listeners are slowed by changes in perspective, especially by those that are unmotivated or incoherent. Take, for example, the sentence *Bill was sitting in the living room reading the paper when John* ***went*** *in*. This example is adapted from a study by Black, Turner, and Bower (1979), who found longer reading times for this sentence than for an otherwise identical sentence with *went* replaced by *came*. The explanation is that readers who have committed themselves to Bill's perspective in the living room experience a sort of narrative whiplash when the perspective suddenly switches from inside to outside the room (where the perspective of the implicit observer to *John went* is located). Abundant evidence has been found that readers also represent and keep track of spatial, temporal, and other goal-related information associated with the writer or protagonist (e.g., Gennari, 2004; Gernsbacher, 1996; Morrow, Greenspan, & Bower, 1987; Zwaan & Radvansky, 1998). This information shapes the emergent structure of discourse (Grosz, 1977; Linde, 1983).

Good writers take account of this and avoid changing perspectives without a reason.

Audience Design in Speaking and Writing

A matter of current debate in the psycholinguistics of discourse and dialogue concerns whether some variations in form reflect *audience design*, or tailoring language to a particular audience or partner. Though some theorists argue that speakers are able to mark and package information for the benefit of addressees (e.g., Galati & Brennan, 2010, 2014; Kraljic & Brennan, 2005), others argue that what appears to be audience design is

simply what is easiest (or automatic) for a speaker to produce, and this just happens to be easy for addressees as well (Brown & Dell, 1987; Ferreira & Dell, 2000; Pickering & Garrod, 2004). Establishing whether a choice or variation in spontaneous speaking is *for* the speaker or *for* the addressee requires an experimental design and task in which the speaker and the addressee have distinguishable perspectives, knowledge, or needs; this can be difficult to stage, as often interlocutors share significant context (see, e.g., Brennan & Williams, 1995; Keysar, 1997). Moreover, partners in conversation must be *aware* of their differences in order to adapt their utterances (that is, to design or interpret them) in partner-specific ways (see Horton & Gerrig, 2005a; Kraljic & Brennan, 2005, for discussion).

Entrainment in Spoken Dialogue. Linguists and psycholinguists concerned with pragmatics and communication have argued that there is no such thing as a synonym (e.g., Bolinger, 1977; E. V. Clark, 1987). Consider, for example, the abstract geometric object in Figure 5.2 (from a referential communication experiment by Stellmann & Brennan, 1993) and the expressions that 13 different pairs of speakers in 13 different conversations spontaneously produced to refer to it.

Each of these pairs of speakers were strangers and were separated by a barrier while they matched identical sets of cards displaying geometric objects or *tangrams* into the same order; one served as the director and the other, as the matcher. Tangrams, being unfamiliar, are not associated with a conventional label, and the card-matching task provides physical evidence of what interlocutors understand, so this task allows experimenters to uncover interactive processes in referential communication. When a pair finished matching the set of cards, the cards were reordered and matched again, for four rounds. Each pair arrived at different conceptualizations, as evident from the idiosyncratic expressions they used. In each case, successful referring was not a simple matter of the speaker producing an expression and the addressee immediately understanding it. Instead, meanings were achieved collaboratively, through exchanges like this one (*Note: overlapping speech appears between asterisks*):

A: ok this one, number 4—it looks kinda like almost like an airplane going down
B: it's ah straight down?
A: yeah it has it looks like it has like a point with like two triangles off the sides kind of like a wing or *wings*
B: *ok* I got it

"a bat"
"the candle"
"the anchor"
"the rocket ship"
"the Olympic torch"
"the Canada symbol"
"the symmetrical one"
"shapes on top of shapes"
"the one with all the shapes"
"the bird diving straight down"
"the airplane flying straight down"
"the angel upside down with sleeves"
"the man jumping in the air with bell bottoms on"

Figure 5.2 Referring expressions from 13 different conversations about the same tangram figure. SOURCE: Stellmann and Brennan (1993).

A: alright
B: yeah

In this exchange, person A began with a somewhat lengthy proposal about what the object resembled, marked as provisional with the hedges *kinda* and *almost*. B asked her for clarification, and A provided more detail. B ratified A's proposal as soon as he believed he understood. They each acknowledged that they believed they were talking about the same thing, and on they went to the next card. This exchange is fairly typical of how people in conversation collaborate to achieve a shared perspective through the *grounding* process (H. H. Clark & Brennan, 1991; H. H. Clark & Schaefer, 1989; H. H. Clark & Wilkes-Gibbs, 1986). The next time they referred to that object (after matching a dozen or so other objects in the set), they could rely on their mutual awareness of the common ground they had established previously; what had previously been a lengthy proposal was now ratified, allowing them to use a shorter and more efficient definite expression (this time, with B as director and A as matcher):

B: and number 2 is that plane going down
A: yup

This attenuation upon repeated referring depends to a large extent on the ability of partners to interact, occurring substantially less when speakers address a silent listener, an imaginary listener, or a tape recorder (Krauss & Weinheimer, 1966, 1967; Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992). In referential communication, interlocutors tend to maintain a perspective once they have grounded it, unless there is good reason to modify or abandon it; this leads to reusing the same expression upon repeated referring (often in a somewhat shortened version). This phenomenon is known as *lexical entrainment* (Brennan & Clark, 1996; Garrod & Anderson, 1987), and provides evidence that interlocutors believe that they share a conceptual perspective. Entrainment occurs not only in repeated referring to difficult-to-lexicalize referents like tangrams, but also for common objects. The following series of expressions from one pair was excised from the repeated matching rounds in a referential communication task that focused on shoes, dogs, cars, and fish (Brennan & Clark, 1996, p. 1488):

Round 1: "a car, sort of silvery purple colored"
Round 2: "the purplish car"
Round 3: "the purple car"

As the partners developed common ground, their referring expressions evolved from lengthy proposals (marked as such with hedges) that included descriptive information and needed to be explicitly accepted or modified by the partner, to shorter noun phrases used with confidence. This process of lexical entrainment strongly constrains the potential variation in referring expressions used *within* a conversation, relative to *between* conversations (where there is much greater variation, as illustrated in Figure 5.2).

Evidence for audience design in referring (that entrainment is partner specific) has been found in studies in which there is a partner switch after two people have entrained on labels for objects. In that situation, speakers take account of new partners by reconceptualizing their perspectives, providing more detail in referring expressions and reintroducing hedges into their utterances (Brennan & Clark, 1996; Horton & Gerrig 2005a). Some of these adjustments may be achieved as afterthoughts or repairs upon seeing cues that the new partner is puzzled; however other adjustments may well be

accomplished early in planning (we take this up presently in the section on dialogue structure and coordination). Not only speakers, but also addressees engage in partner-specific processing, in that they interpret referring expressions differently depending on who produced them (Metzing & Brennan, 2003); in a matching task in which naïve subjects wore an eye-tracker, a confederate directed them to place objects in an array, interacting spontaneously except for producing a total of eight referring expressions that were scripted in advance. After the matcher and director had entrained on labels in repeated trials, the confederate director left the room and then returned, or a second confederate returned, for the last trial. In that trial, the director used either the previous term or a new term. When an old partner used a new term (thereby departing from the precedent they had entrained upon earlier, that is, inexplicably *breaking a conceptual pact*), matchers were slow to interpret the term and appeared to search the display (perhaps looking for a new object that might have snuck in). The same new term spoken by the new partner involved no such expectations, and was interpreted just as quickly as the old term uttered by either partner (with rapid looks to the object that best matched the new term).

Note that Metzing and Brennan's (2003) experiment was inspired by Barr and Keysar's (2002) second experiment, which also employed a switch in speakers; that is, subjects entrained with a confederate speaker on labels for objects and then, in a critical trial, heard the previously mentioned label from the original confederate speaker or a new speaker. The third (final) cell in Barr and Keysar's design consisted of the new speaker producing a new label for the object. Because subjects were equally fast to look at the object when the old label was spoken by the new speaker as by the original speaker, Barr and Keysar concluded that precedents established during repeated referring are not represented in any partner-specific manner. However, that experiment was missing a key comparison with a cell in which the original speaker used a new label (as in Metzing & Brennan's broken conceptual pact). Other studies employing speaker switches in a variety of ways have argued for or against partner specificity in referential precedents, as we will discuss presently.

An Application of Audience Design: Writing for a Reader's Comprehension. In his advice on good writing, *Everyone Can Write Better (and You Are No Exception)*, H. H. Clark (2000) advises scientific writers to never write a word or phrase that they would not say aloud. This does not presume that written and spoken discourse are the same; they are not, as the costs and affordances associated with speaking and writing are quite different (see Carter-Thomas & Rowley-Jolivet, 2001; H. H. Clark, 2000; H. H. Clark & Brennan, 1991). Interactive dialogues, be they spoken or texted, are planned under social pressure—where the speaker or writer may sacrifice fluency or polish rather than risk losing the addressee's attention, whereas written monologues or text e-mails can ordinarily be edited without such risk. Speech is ephemeral, whereas text leaves a record that can be reviewed.

Despite the advantage of editable text, speaking spontaneously can help preserve the flow of ideas, since speakers often naturally mark given and new information in ways that addressees can process with ease (whereas overediting and rewriting can disrupt this natural flow of information). For this reason, we often advise students to read their papers aloud as part of the editing process.

Even though middle school teachers sometimes encourage young students to avoid using the same words over and over

in their writing assignments, and advise them to display their knowledge of vocabulary by using complex words rather than common ones (a strategy that seems to haunt some students through college, graduate school, and into academia), this is not helpful to readers (as H. H. Clark, 2000, observes). As research on entrainment suggests, introducing a new referring expression to specify something already mentioned can mislead readers, unless it is evident that the new referring expression is intended as an appositive (providing additional information about the same referent); a new referring expression is likely to suggest to the reader that a new referent is being introduced. Within a discourse context, writers should not switch terms without good reason; they should use the same term when they mean the same thing (as shown for speakers and addressees by Brennan & Clark, 1996; Metzing & Brennan, 2003); consistent with this principle, the same term should not be applied to distinctly different referents (Van Der Wege, 2009), or else readers or addressees will be led astray. If a discourse continues to be about a discourse entity that is presumed to be still salient in the mind of the reader, a pronoun or elliptical phrase should be used rather than a full noun phrase (Gordon et al., 1993; Hudson D'Zmura, 1988).

MODELS OF DISCOURSE AND DIALOGUE

Both discourse and dialogue exist in fundamentally social contexts. Both are locally structured by how information is packaged—as given and new, in referring expressions, and according to temporal, spatial, lexical, and personal perspective. However, monologues (whether text or spoken) are structured differently from dialogues (Fox Tree, 1999). Thus, we organize this section into different kinds of models: models of discourse structure in *monologue*, which is the product of a single mind (albeit one engaged to some degree in audience design), and models of structure in *dialogue*, which is shaped by the interaction of partners who coordinate their contributions within a particular communication medium (Brennan, Galati, & Kuhlen, 2010; H. H. Clark & Brennan, 1991).

Models of Discourse

Here, we briefly cover several influential models of discourse processing of text authored by an individual who is not engaged in interaction with a particular dialogue partner. Although we must limit our coverage due to space, there is much more work that could be included here, and indeed there has been more work on the psycholinguistics of text comprehension than on spoken dialogue.

The Construction-Integration Model of Reading

The *construction-integration model* of discourse comprehension developed by Kintsch and van Dijk (Kintsch, 1988; van Dijk & Kintsch, 1983) captures the impact upon reading comprehension of both the *form* of discourse, as it is structured locally and globally, and the *content* of discourse, whether realized explicitly or implicitly (integrated with world knowledge). According to this influential model, as linguistic input is encountered in text, it first activates the reader's knowledge, and then this knowledge is selectively integrated with a model of the text that the reader is building in working memory (see Graesser & Forsyth, 2013). Several levels of representation are implicated during this interpretive process.

First, there is a fleeting record in working memory of the verbatim surface form of text; this surface level of representation is assumed to be ephemeral, due to abundant evidence that people tend to recall the gist rather than the exact wording of material they hear or read (see Sachs, 1967). Next, readers represent the gist or text base, consisting of propositions or small units of conventional meaning that are extracted from the words in a text; also represented are the particular relationships between elements (as signaled by thematic relations such as semantic roles, or by connective words such as *and* or *however)*. Finally, an episodic mental representation is constructed as the reader interprets and integrates propositions into a model of the situation as described by the writer.

For good readers, this integration occurs automatically, with previously mentioned elements priming associated information to make it available in working memory. As noted earlier, an evolving discourse model is informed by at least four kinds of input (the state of the discourse model constructed so far, the new text being processed, the text read recently, and the reader's world knowledge; van den Broek et al., 1999).

Intentional Structure, Attentional State, and the Stack Model

In addition to the textual linguistic elements that are structured into a discourse (e.g., surface forms and text base as identified by the construction/integration model), discourse structure is shaped by what people are *doing* with language. Early work on task-oriented dialogue, such as where one person instructed another in how to assemble a pump (Grosz, 1977) or where people described apartments (Linde, 1983), demonstrated that physical tasks and goals can shape discourse structure. Building on this work, computational linguists working in artificial intelligence proposed a theory of discourse structure based on three interacting components: linguistic structure, intentional structure, and attentional state (Grosz & Sidner, 1986). These three component sources combine as inputs to a computational mechanism for determining the context and constraints with which referring expressions can be interpreted.

On this theory, "clues" expressed linguistically (e.g., from phrases like *by the way* or *first of all*, or changes of verb tense, person perspective, prosody, etc.) can organize discourse into segments, where each can be associated with a primary purpose intended by the speaker/author to be recognized and shared with the addressee (note that on this theory, discourse can have other, often implicit purposes as well). Subsegments with intermediate purposes that serve the larger purpose (such as the repair sequence of Brad and Amanda's misunderstanding from our early example) were captured by Grosz and Sidner's model using a stack metaphor that employed a first on, last off principle (like a stack of plates in a cafeteria dispenser). In their model, attention is deployed to new subsequences that are pushed atop previous ones and then popped off when their purposes are achieved (or abandoned). This model represents not only subsequences relevant to the primary purpose at hand like repairs or clarification subdialogues, but also digressions or interruptions as in the following example (from Polanyi & Scha, 1984, as quoted in Grosz & Sidner, 1986, pp. 192):

> John came by and left the groceries
> *stop that you kids*
> and I put them away after he left

In this case, presumably the speaker used a different tone of voice for the interruption, signaling that the discourse entity evoked by

you kids inhabited a context distinct from the surrounding utterances. After the interrupting imperative was popped off the stack followed by a return to the prior context, the *them* in the last utterance co-specified the groceries rather than the kids.

Grosz and Sidner's computational linguistic model was also appealing from a psychological standpoint, in that it did not focus solely on the surface text product, but related linguistic structure to intention and attention. In fact, the mid-1970s to mid-1980s were an exciting time in general for discourse and dialogue researchers from multiple disciplines who worked in the language-as-action tradition. These included Chafe (1980), a linguist who was examining commonalities and differences in the flow of narratives spontaneously produced in response to the Pear Stories; ethnomethodologists such as Goodwin (1979, 1981) and Sacks, Schegloff, and Jefferson (1974), who were documenting the details of conversation interaction; H. H. Clark and Wilkes-Gibbs (1986), psychologists who were examining the collaborative nature of spontaneous referential communication; and computer scientists such as Schmandt and Hulteen (1982) and Winograd (1971, 1983), who were creating automated systems that responded to natural language commands from human users. Before we turn to models that explicitly address coordination in two-person dialogues, we will cover a theory related to Grosz and Sidner's stack model, the centering theory of Grosz, Joshi, and Weinstein (1986).

The Centering Theory

Despite what one's middle-school English grammar teacher may have claimed, pronouns do not co-specify the most recently mentioned noun phrase that agrees in gender and number. Consider this made-up example:

> Brennan$_1$ drives an Alfa Romeo.
> She$_1$ drives too fast.
> Friedman$_2$ races her$_1$ on weekends.
> She$_2$ often beats her$_1$. (Brennan, Friedman, & Pollard, 1987, p. 157)

Whereas some readers find the pronouns in the last utterance to be genuinely ambiguous, many interpret the *she* to co-specify *Friedman* and *her*, *Brennan*. According to the centering theory (see Brennan et al., 1987; Grosz et al., 1986; Walker, Joshi, & Prince, 1988), realizing *Brennan* as the first sentential subject marks it as salient, so if that entity appears in the second sentence, it *must* be referred to with a pronoun, as opposed to a full name (which would sound awkward as well as lead to Gordon et al.'s [1993] *repeated name penalty* or slowing in reading time). Realizing *Friedman* as the subject of the third sentence marks it as salient (a forward-looking center), although *her* (*Brennan*) is still retained as the center of attention, or what that sentence is about (the backward-looking center, or center of attention). However, because of the salience accorded to *Friedman* as subject of the third sentence, the center of attention shifts to *Friedman* as *she* in the fourth sentence is taken to co-specify the most salient appropriate discourse entity. The predictions of this centering algorithm were supported by psycholinguistic findings from reading experiments (Hudson et al., 1986; Hudson-D'Zmura, 1988).

In Grosz et al.'s (1986) centering theory, as well as in Brennan et al.'s (1987) centering algorithm for resolving pronouns, context was defined largely as the surrounding words and sentences. Although many examples used by these researchers were made up, and the theory did not distinguish monologue from dialogue, there is evidence that the algorithm is consistent with spontaneous spoken dialogue (Brennan, 1995). In an experiment involving pairs of naïve speakers spontaneously discussing a basketball game,

a speaker describes the action to an addressee who cannot see the game:

> And now Wolverines have the ball . . .
> They're going down . . .
> Number thirty passes it off to forty-one$_1$ . . .
> Forty-one$_1$ goes up for the shot
> And he$_1$ misses. (Brennan, 1995, p. 142)

Here, the speaker repeats the full noun phrase *forty-one* when immediately re-referring to the same player rather than pronominalizing (even though the semantics of the situation are clear—one cannot shoot unless one has the ball). The pattern of findings was consistent with predictions from centering (Brennan et al., 1987; Grosz et al., 1986; Hudson et al., 1986), that in re-referring, speakers tended to repeat the full noun phrase rather than using a pronoun when a discourse entity was not currently salient due to having just been mentioned as a sentential object (Brennan, 1995). Also consistent with centering and the repeated name penalty was the finding that speakers prominalized directly after introducing an entity as a sentential subject (Brennan, 1995).

The influence of the centering theory may be due in part to its ascent at a time when psycholinguists were conceptualizing pronouns as memory cues that index entities that are salient and available to both speaker and addressee (rather than simply as placeholders that initiate a search process through text guided by recency and semantic knowledge). In short, when an entity is salient, a pronoun cues it rapidly without any need for "search" through previous discourse. In fact, the domain of interpretation is not restricted to explicitly mentioned antecedents in the prior discourse, as in this example of an unheralded pronoun:

> A: The set is a rip-off from "Gentlemen Prefer Blondes."
> B: Is that the one where she's standing over the grate and her dress blows up? (Greene, Gerrig, McKoon, & Ratcliff, 1994, p. 512)

Here, B's pronouns refer to Marilyn Monroe, successfully assuming that this is in common ground with A (even though B gets the movie wrong; Greene et al., 1994).

The centering theory encouraged further work on *shared attention* in dialogue, which had previously been almost entirely siloed within the field of child language acquisition (e.g., studies charting early word learning and the acquisition of pointing in infancy; see Baldwin, 1995). However, despite the predictive power of the centering theory (and its suitability as the basis for pronoun interpretation algorithms in natural language processing systems), this approach (at least that of the original formulation and algorithms based on it) is not entirely plausible as a psychological model, due to the limited definition of context as the preceding text and the presumption of discrete, successive discourse segments within which the salience of each discourse entity remains largely constant. Language processing runs on memory, with discourse entities waxing and waning in their activation (see, e.g., McKoon & Gerrig, 1998). The *intra*personal processing that takes place in the mind of an individual needs to be modeled at a finer grain than that addressed by the centering theory, as do the closely timed behaviors in *inter*personal coordination between two partners.

Models of Dialogue Structure and Coordination

As we have noted, language unfolds over time, and processing proceeds incrementally. A reader does not wait until the end of a sentence before interpreting it, but begins immediately to activate lexical items, build syntactic structures, and interpret meaning.

Likewise, during speech perception, a listener begins to activate and recognize words before they are fully pronounced and interprets utterances while they are still being spoken (e.g., Tanenhaus et al., 1995). Speakers begin to speak before they have finished planning what to say (e.g., Dell, 1986). Although audience design occurs in writing as well as in speaking, speaking affords opportunities to engage in partner-adapted processing and interacting at a fine grain on the part of both speakers and addressees. Dialogues are a particularly interesting sort of discourse, as they clearly reflect coordinated effort on the part of at least two partners who alternate as speakers and addressees, and who appear to achieve shared meanings . . . *somehow*.

The questions in this section include: What happens to language planning and interpretation when these processes must also be coordinated in both content and timing with another person? And: Do phenomena such as entrainment and audience design really rely on common ground or reflect adapting to the needs of a dialogue partner, or do they emerge simply as artifacts because two partners are similar in their characteristics and are primed by sharing the same perceptual environment?

Researchers' implicit assumptions or explicit theories about what constitutes the essence of dialogue affect the kinds of experimental designs and tasks they use, what factors they seek to control, and even how they interpret their findings (see Kuhlen & Brennan, 2013, for discussion). Such assumptions or theories also affect what kinds of human–computer interfaces or spoken dialogue systems researchers invent. Of course, not all researchers who study dialogue are driven by psychological questions about process; those interested in building spoken dialogue systems for telephony applications (whether for specific task domains like getting train schedules or account information, or for more domain-independent "intelligent" personal assistants like Apple's Siri) are primarily interested in generating realistic dialogue transcripts or supporting an experience that appears to be "natural" to a human user. Early on, researchers in artificial intelligence realized that even common words and referring expressions inspire considerable ambiguity within a particular context; this means that task-oriented dialogues must model the clarification of referring expressions as well as the repair of misunderstandings. Consider this early natural language dialogue from Winograd's SHRDLU program (1971), in which a simple graphic representing a virtual robot's arm responded to a speaker's commands to move objects around:

Person:	PICK UP A BIG RED BLOCK.
Computer:	OK. <*does it*>
Person:	GRASP THE PYRAMID.
Computer:	I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN. <*There are three pyramids in the scene, so the program does nothing. The person can see this by monitoring the graphic representation.*>
Person:	FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND PUT IT INTO THE BOX.
Computer:	BY "IT," I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.
Computer:	OK. <*does it*> (Winograd, 1971, pp. 35–39)

SHRDLU was a powerful demonstration of several hard problems that spoken dialogue systems would need to handle somehow—including conversational repairs and the interpretation of ambiguous words such as pronouns—but it was far from an actual working system, as it reliably generated this behavior only when the human user stuck to a limited script (Norberg, 1991).

Some of the early spoken dialogue systems that followed handled repairs more systematically, such as Put That There (Schmandt & Hulteen, 1982), which successfully modeled the resolution of simple indexical expressions by taking advantage of the ability to point and to share initiative in dialogue. In Put That There, the system relied on the lexical semantics of a small number of verbs (which require arguments such as subjects and objects) and implemented two ideas that were entirely innovative in human–computer interaction at the time: The system took the initiative for soliciting missing information from the user, and it did so multimodally, combining information from speech, graphics, and pointing:

User:	Put that <points at object on a screen>
System:	<*highlights object*> Where?
User:	There. <*points at location on the screen*>
System:	<*performs action*>

Another innovative system that was fairly robust and could actually be used by naïve users was invented by Davis (1989): Backseat Driver was the first spoken dialogue system to provide a driver with GPS directions. Its design was inspired by a task analysis and corpus of hours of directions given to drivers by passengers (Davis, 1989). Regardless of the motivation for how a speech interface should manage a dialogue, regardless of whether the partner is human or machine, and regardless of whether the creators of such systems are concerned with the psychological questions surrounding dialogues, it is not sufficient to model language use alone; it is also necessary to model coordination between partners.

Over the years, many different metaphors have been used to explain how dialogue structure emerges from coordination. Such metaphors have included passing messages, competing for a scarce resource (the conversational floor; Sacks et al., 1974), or participating in practiced routines that consist of dialogue moves (e.g., Larsson & Traum, 2000; Traum, 1994). The following subsections present some explanations that aim to address, at least in part, how coordination shapes dialogue structure.

The Message Model

Perhaps the most pervasive explanation for the structure of dialogue is one that inherits its assumptions from information theory (MacKay, 1983; Shannon & Weaver, 1949) and has been dubbed the *message model* by Akmajian, Demers, and Harnish (1987), or the *conduit metaphor* by Reddy (1979). The message model assumes that a speaker (or sender) encodes thoughts into words (which are presumed to be little packages that contain meanings) in order to produce a message, which is then conveyed through a communication channel to be received by others. These recipients then simply decode the message using the same linguistic rules for decoding that the message was encoded with. On this view, communication should succeed as long as the senders and recipients speak the same language and as long as there is not too much noise in the communication channel.

However, the message model fails in many ways. Simply knowing the same language (and being able to use the same encoding and decoding rules) is no guarantee of successful communication. In our early excerpt from the phone conversation between Brad and Amanda, Amanda asks, *Have you got a new job yet?*—a question that seems simple enough. However, it takes them five more speaking turns to get Amanda's question clarified and to come to a shared understanding of Brad's *no* answer. The message model fails to predict the need for such repairs.

Another assumption of the message model is that the important information is transmitted from sender to receiver. That assumption does not hold up either. Consider the following exchange between two students, A and B, who participated in a referential communication experiment in which they were separated by a barrier while matching duplicate sets of cards:

A: Ah boy this one ah boy alright it looks kinda like—on the right top there's a square that looks diagonal.
B: Uh huh
A: And you have sort of another like rectangle shape, the—like a triangle, angled, and on the bottom it's ah I don't know what that is, glass-shaped.
B: Alright I think I got it.
A: It's almost like a person kind of in a weird way.
B: Yeah like like a monk praying or something.
A: Right yeah good great.
B: Alright I got it. (Stellmann & Brennan, 1983)

Notice that the praying-monk perspective was actually proposed by B, the person who did not know the identity of the card they were discussing (and in fact, they ended up entraining on *the monk praying*, with both students using that phrase throughout the rest of the experiment). It is evident from examples like this one that speakers are *not* simply sending messages to addressees who are simply decoding on the receiving end; instead, interlocutors work together to achieve a shared perspective (H. H. Clark & Wilkes-Gibbs, 1986).

Another way in which the message model fails is that it presumes that brief listener responses (or what Yngve, 1970, called *backchannels*, such as A's *Right yeah good great)* regulate the flow of information through the channel, just as in an engineering application, where a feedback signal controls the speed of a servo motor (according to Rosenfeld, 1987, p. 584, "If the speaker is generating new information at an adequate rate the listener should be expected to signal the speaker to continue via a simple listener response"). On this view, such feedback signals are assumed to not contribute any content, and in fact to be unnecessary unless the channel is noisy (we will present an alternate view of such signals presently). Moreover, a dialogue need not result in any joint product achieved by participants working together, with both taking responsibility for mutual understanding; instead, the speaker should be ready to move on as soon as she has uttered the message, and the addressee, as soon as he has autonomously reached a state of understanding. Rosenfeld (1987) also used the metaphor of entering and exiting the flow of traffic to account for turn-taking in conversation; however, the implications are that drivers (speakers) need only to avoid collisions with other drivers (overlapping speech); they do not care whether the other driver ever gets anywhere (for discussion, see Brennan, 1990).

Although some version of the message model seems to be assumed within many research agendas from psychology, linguistics, artificial intelligence, and human-computer interaction, we argue that it is not the basis for a satisfying model of dialogue.

Adjacency Pairs and Turn-Taking in Dialogue

In spoken dialogue, words are structured into utterances, and utterances are structured into conversational turns. The message model fails to capture the relationships of relevance between adjacent utterances by different speakers. Consider the following exchange:

Susan: you don't have any nails, do you?
Bridget: no

Conversation analysts have observed that utterances are often produced in meaningful pairs of turns, or *adjacency pairs*. An adjacency pair accomplishes a collaborative task such as a question and an answer (where the two utterances perform complementary functions, with each speaker taking a different role), or a closing (where each party reciprocates in bidding the other goodbye). The idea of adjacency is taken loosely, as adjacency pairs can be nested within other adjacency pairs (Schegloff, 1972; Schegloff & Sacks 1973).

The related phenomenon of *turn-taking* has likewise been extensively documented by conversation analysts, with other disciplines applying insights from that work to both human and human–machine interaction. The conversational floor has been viewed by conversation analysts as a limited resource that needs to be managed by interacting speakers in order to avoid significant stretches of overlapping speech (Sacks et al., 1974). Sacks et al. proposed that utterances are constructed of turn-constructional units (consisting of words, phrases, clauses, and sentences) and proposed a set of rules by which the economy of turn-taking is managed. For example, their rule *current speaker selects next* is consistent with the observation that a speaker (especially when more than two other people are present) will often suspend speaking and look at the person who begins to speak next. If that person does not speak, then another person may self-select, or the first speaker may continue.

The problem with the proposal that turn-taking behavior is rule generated is its presumption that the purpose of conversation is to hold or manage the floor and minimize overlaps, rather than to reach a point where interlocutors believe that they understand one another (this assumption also underlies some modern research in turn-taking; e.g., Levinson & Torreira, 2015; Wachsmuth, 2008). However, it is likely that what Sacks et al. (1974) called *rules* do not actually generate turn-taking behavior; rather, turn-taking is generated from the need to ground meanings. When a speaker returns her gaze to an addressee, she is looking for evidence of understanding or uptake; the addressee may then speak to provide such evidence, else the speaker may rephrase in order to be clear (Brennan, 1990). Rather than turn-constructional units determined a priori by linguistic structure, the primitives of dialogue can be considered to be constituents presented provisionally and needing to be grounded by a speaker and addressee working within the constraints of a particular medium, to some criterion. Sacks et al. do acknowledge at the end of their article, "It is a systematic consequence of the turn-taking organization of conversation that it obliges its participants to display to each other, in a turn's talk, their understanding of other turns' talk" (p. 728). But this seems backward; in spoken conversation, a systematic consequence of grounding is that it obliges participants to take turns.

Interactive Alignment and Other Two-Stage Models

Many studies have demonstrated that people build up common ground over the course of a conversation; what is under debate is the extent to which they *really* are taking one another's perspective, knowledge, or needs into account, as opposed to just *appearing* to be doing so (as we foreshadowed in the section on *applying audience design to speaking and writing*). Some theories assume that similarity between interlocutors is enough to ensure that they will understand one another; as Sperber and Wilson asserted about their *relevance* theory, "Clearly, if people share cognitive environments, it is because they share physical environments and have similar cognitive abilities"

(Sperber & Wilson, 1986, p. 41). To the extent that two people are similar in their cognitive abilities and experiences, share the same perceptual environment, and are representing their recent conversation in working memory, what is easiest for the speaker is often easiest for the addressee (Brown & Dell, 1987; Dell & Brown, 1991). The telling case is when partners in a conversation hold distinctly different perspectives (Keysar, 1997). To distinguish speaker's and addressee's perspectives, several kinds of tasks have been used. These include making information available to one partner but not to the other, usually via lack of perceptual co-presence where some objects are occluded or missing from one partner's display; changing the partner at some point during the session so the new partner would not have access to prior linguistic co-presence; and ensuring that two interacting partners have distinct viewpoints in a visuo-spatial task.

The two-stage models cited here hypothesize that any aspects of utterance planning or interpretation that are specifically adapted to a partner are resource intensive and thus require extra processing. Such models include *monitoring and adjustment* (Horton & Keysar, 1996), *perspective adjustment* (Keysar, Barr, & Horton, 1998); and *interactive alignment* (Pickering & Garrod, 2004). These theories posit an initial egocentric stage that is modular (informationally encapsulated); that is, initially speakers or addressees do not take each other's perspectives into account, but process language in a way that is fast, automatic, and inflexible. Following that, processing may be adjusted to a partner via a slower and more computationally expensive, inferential process; at that second stage, what Pickering and Garrod call "full common ground" is either deployed optionally or invoked only when necessary for a repair ("Normal conversation does not routinely require modeling the interlocutor's mind"; Pickering & Garrod, 2004, p. 180). Another two-stage model, *anticipation integration* (Barr, 2008; Bögels, Barr, Garrod, & Kessler, 2014), proposes that common ground can have an early, anticipatory effect *before* an addressee hears an utterance (based on Barr and colleagues' evidence that the addressee tends to look more at objects that are in common ground with the speaker), while questioning whether addressees can use common ground in the online processing of the utterance (for discussion, see Brown-Schmidt & Hanna, 2011, as well as the upcoming section on neural evidence for mentalizing and perspective-taking).

On Pickering and Garrod's interactive alignment model (2004), language processing in dialogue differs from language processing in monologue because in dialogue, the speech production and comprehension systems are both active at once, with the two systems working off the same mental representations of dialogue context (known as *representational parity*). This and other two-stage models propose that interlocutors routinely come to achieve shared mental representations directly, through priming (as opposed to any sort of partner-specific processing or mentalizing). Priming has been offered by the authors of the interactive alignment model as an explanation for phenomena such as entrainment, as when speakers and addressees in conversation come to use the same referring expressions repeatedly to refer to the same thing (for challenges to this view, see the commentaries following Pickering & Garrod, 2004).

Some of the experimental evidence presented in support of a two-stage view has come from egocentric errors in perspective-taking. In one investigation (Keysar, Barr, Balin, & Brauner, 2000), an experimental confederate used a referring expression (e.g., *Pick up the small candle*) that ambiguously matched not only a large- and medium-sized

candle in common ground (visible to both confederate speaker and addressee subject) but also a smaller candle occluded from the speaker's view and thus privileged to the addressee (who wore an eye-tracker). That addressees did not ignore privileged information that only they could see, but included the small candle in their early looks around the display, was interpreted as evidence for egocentricity. However, this behavior could also be explained by lexical competition, especially since the privileged object was the best match for the referring expression compared to the other objects of the same type (for discussion, see Brown-Schmidt & Hanna, 2011).

In order to provide a fair test of whether partners in dialogue are able to take account of one another's perspectives early in processing, an experiment must not only distinguish between the partners' perspectives, knowledge, or needs (Keysar, 1997, but also set up a situation in which they are *fully aware* of their distinct perspectives, knowledge, or needs (Horton & Gerrig, 2005a; Kraljic & Brennan, 2005).

Partner-Specific Processing

The alternative explanations to two-stage models that we describe in this section do not posit an early, egocentric stage and a late inferential stage where partner-specific processing can be achieved, but argue that partner-specific processing in dialogue is simply a function of ordinary memory processes. That is, any information about a partner's perspective that is currently activated in working memory can be used early in planning and interpreting referring expressions (Brown-Schmidt & Hanna, 2011; Horton & Brennan, 2016; Horton & Gerrig, 2005b). Such memory traces that become active during grounding vary in their strength and accessibility, and so combine probabilistically to constrain utterance planning and interpretation. This means that sometimes speakers and addressees take one another's perspectives into account early in processing, and sometimes they do not. A constraint-based model views referring *not* as a deterministic process where speakers provide only what is strictly necessary to pick out one object from a set, nor as a process where addressees come to inflexibly associate an object with a single expression. Rather, representations in memory that link referents and expressions can wax and wane, as well as be updated abruptly when the pragmatics of the situation change, such as when the speaker changed in the Metzing and Brennan (2003) study discussed earlier. Critically in that study, the old speaker's identity and the common ground established with that speaker during grounding shaped addressees' early processing of a new referring expression, compared to when that same new expression was presented by a new speaker.

It would, of course, be computationally expensive to maintain and tailor processing to an elaborate model of a dialogue partner (Horton & Keysar, 1996; Polichak & Gerrig, 1998). But contrary to predictions from two-stage models, processing language in dialogue can be quite flexible. Even though inferences about a partner's perspective, knowledge, or needs take measurable processing time to make, it appears that once an inference has been made (meaning that common ground or other pragmatic, partner-specific information is active in working memory), it can be reused without cost (Hwang, Brennan, & Huffman, 2015). And considering privileged information upon hearing a speaker's referring expression does not always represent an egocentric error; in fact, when the expression forms part of a *wh-* question, an addressee who is taking the speaker's knowledge into account *should* first consider what the speaker does *not*

already know. This point was compellingly made in studies by Brown-Schmidt and colleagues. *Wh-* questions about objects unseen by speakers but that had visible competitors in a display led addressees to gaze more at the objects that speakers did not know about than ones that both partners could see (Brown-Schmidt, 2009b; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008). Findings such as these demonstrate that interpretation is a highly flexible process rather than simply a matter of low-level "dumb" priming (as assumed by two-stage models).

One particularly innovative study by Brown-Schmidt and Fraundorf (2015) demonstrated that addressees do not simply respond inflexibly to the form of an utterance (as either a statement or a question). When the speaker asked a question with falling intonation, which is typical in questions about what is unknown (e.g., *What's above the bear that's wearing a flower?*), addressees fixated objects that were unknown to the speaker, whereas when the question was asked with high intonation, suggesting that the speaker knew but had just forgotten the answer (e.g., *What's? above the bear? that's wearing a flower?* pronounced as in *What was that again?*), addressees fixated objects that were in common ground and known to the speaker. Such fixations were produced rapidly, contradicting predictions from two-stage models that partner-specific processing must result from a slow, inferential process or from the repair following an egocentric error (Brown-Schmidt & Fraundorf, 2015).

Studies that provide the clearest evidence for the rapid and flexible use of common ground tend to use tasks in which the working memory load on interlocutors is relatively low; partners' perspectives can be distinguished with one or just a few clear and relevant perceptual cues or well-established factors, sometimes binary in nature. These situations include whether two partners can both see what they are talking about (Brennan, 1990; Lockridge & Brennan, 2002); whether they have previously discussed what they are discussing now (Brennan & Hanna, 2009; Galati & Brennan, 2006, 2010, 2014; Matthews, Lieven, & Tomasello, 2010; Metzing & Brennan, 2003), whether a speaker can reach or is gazing at the object she is referring to (Hanna & Brennan, 2007; Hanna & Tanenhaus, 2004); whether previous speech was interrupted before a referring expression was fully grounded (Brown-Schmidt, 2009b); or whether an item in a matching task needs to be distinguished from a similar adjacent item (Hwang et al., 2015). Such simple situations or one-bit models (as opposed to elaborate models of a partner's knowledge) can serve as cues that are relatively easy to monitor or keep track of, especially when evidence from the dialogue context keeps them strongly activated in working memory (see, e.g., Brennan & Hanna, 2009; Galati & Brennan, 2006; Horton & Brennan, 2016).[2] The first time such an inference is made, extra time is needed, but once the partner-specific information has been computed and is available, it can be used rapidly, such that partner-specific processing is essentially automatic (Hwang et al., 2015).

Another characteristic common to experiments that show clear evidence for rapid partner-specific processing is that they involve pairs of naïve subjects, or else subjects interacting with a confederate

[2]Note that by *one-bit model*, we mean only that the partner-relevant cue or constraint is binary (that is, simple) and therefore easy to perceive or represent, such as that something is visible or not. By *binary*, we do *not* mean that the information or constraint is deterministic or associated with 100% or 0% confidence. The representation cued by a binary cue (and the evidence for it) can presumably be associated with a particular confidence value or strength in memory, and is therefore gradient (see Brown-Schmidt, 2012; H. H. Clark & Schaefer, 1989).

who has actual informational needs in the experimental task. It has been argued that partner-specific information is likely to be more strongly activated and thus easier to use in interactive versus noninteractive situations (Brown-Schmidt, 2009a, 2012); in addition, subjects cannot interact with prerecorded speakers to ground utterances, and so they may behave as if they are participating in quite a different language game than in an interactive dialogue. In general, people appear to be quite sensitive to odd behavior in dialogues, including behavior that involves nonverbal cues. As an example, similar experiments by Lockridge and Brennan (2002) and Brown and Dell (1987) paired, respectively, two naïve subjects versus a naïve subject and a confederate (who acted as addressee in the same task an average of 40 times), finding different results concerning partner-specific processing (evidence *for* in the former and *against* in the latter). It is feasible that subjects can detect when their partner knows too much, and that they would adapt to the partner's needs only when there are actual needs. For this reason, it is wise for experimenters to pay attention to the ways in which confederates are deployed as speakers or as addressees in an experiment, as their nonverbal behavior (Brennan & Williams, 1995; H. H. Clark, 1996) can have unintended influences on the dialogue context. See Kuhlen and Brennan (2013) for discussion of the risks and benefits of using confederates.

The Need to Model Coordination in Dialogue

Dialogue structure is an emergent product of the coordination between interlocutors; coordination shapes dialogue from the start. This means that coordination should be modeled as an essential process that drives language use, during both comprehension and production (rather than as a late stage that considers partner-specific information only after a speaker has planned all or part of an utterance, or only after an addressee has reached an initial egocentric interpretation).

Moreover, the fact that misunderstandings occur from time to time does not warrant the conclusion that processing is egocentric. People have many demands on their attention, and they make mistakes. They must often trade off speaking fluently with initiating speaking in a timely fashion (H. H. Clark & Brennan, 1991). And the very fact that even young children engage in conversational repairs (E. V. Clark, 2014) suggests that (at times) children want to adapt to their partners, even if repairs are not always successful. Let's revisit the previous example of Susan's question and Bridget's answer in more detail. This dialogue occurred between officemates who were students in the same graduate program. What really happened is that one entered their shared office, dumped her books on her desk, leaned two large framed posters against the couch, and asked the other:

Susan:	you don't have any nails, do you?
Bridget:	<*pause*>
	fingernails?
Susan:	no, nails to nail into the wall
	<*pause*>
	when I get bored here I'm going to go put up those pictures
Bridget:	no

From the perspective of the first speaker, the abrupt initial question does seem egocentric; Susan could have led with her intention: *I want to hang these pictures—do you have any nails?* At this moment, Susan's priority was to create a plan to dispatch the large, bulky posters, and it was possible that Bridget might have noticed these unusual objects and inferred the intention behind the question. At any rate, there was little cost to Susan's

carelessness in designing an utterance, as she and Bridget were co-present and could repair the problem easily.

But according to both the message model and the interactive alignment model, Bridget should have understood Susan's question with ease. After all, they were both native English speakers, speaking face-to-face, and discussing a concrete situation using simple, common words. Despite this apparently optimal situation, as well as their highly similar backgrounds and shared perceptual affordances, *nails* evoked a different homophone for Bridget than the one intended by Susan. Bridget noticed the problem first, and provided evidence about a tentative interpretation: *fingernails?* This evidence (and the delay before Bridget's utterance) made Susan aware of the problem a moment later; the additional pause after her attempt at clarification (that is, Bridget's lack of uptake) suggested that Bridget still didn't understand (*Did Susan really mean nails to nail into the wall, or was she being sarcastic?*). Bridget was not able to answer Susan's question until she understood what Susan was up to. As is evident from the pauses, making these sorts of complex inferences takes up processing time (especially the first time an inference is used).

This example illustrates the importance of recognizing a partner's intention and making other inferences that are fundamental to successful communication; these are lacking in the all-too-simple message model, as well as in interactive alignment (which pushes information from inferences into a second stage of processing). Further, it demonstrates that similarity between conversational partners gets them only part of the way to mutual understanding. It is this very process of *grounding*, or seeking and providing evidence about intention and understanding (Brennan, 1990, 2005), that shapes the form that spoken dialogue takes, allows interlocutors to achieve meanings incrementally, and provides a safety net that makes communication likely to succeed, despite the ambiguity and distraction present in real-world communication. On H. H. Clark and Wilkes-Gibbs's (1986) original formulation of the grounding theory, a principle of *mutual responsibility* specifies that not only speakers, but also addressees, take responsibility for grounding by working together to minimize the effort that they put in collectively (following a principle of *least collaborative effort)*. Finally, they attempt to meet a *grounding criterion* or standard of evidence that there is enough certainty that they understand one another to satisfy current purposes (so dialogues between air traffic controllers and pilots should [and in fact are required] to use a much higher grounding criterion than strangers chatting in line at a store, seeking and providing different strengths of evidence about an understanding).

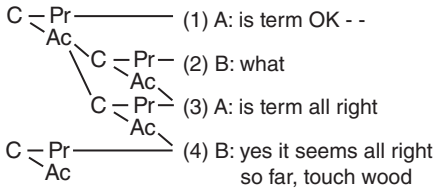
Collateral Signals. In dialogue, meanings are coordinated through grounding, When Susan asked Bridget for nails, she did so provisionally by marking her question with an expression of doubt and using a tag question (*You don't have any nails, do you?*). She could not count on the request she had just uttered being in their common ground until she had evidence that Bridget had attended to it, heard it, understood it, and taken it up (H. H. Clark & Brennan, 1991); such evidence could take such forms as a simple *no*, or else Bridget opening her desk drawer and handing over the nails. But as Bridget could not be sure about Susan's intention, she provided evidence of a lack of understanding in the form of a clarification question: *fingernails?* Likewise, in our previous examples from tangram-matching tasks, the directors' initial referring expressions were also marked as provisional, with multiple attempts at proposing a perspective, or with hedges, as in partner A's utterance to partner B, *OK this one, number 4—it looks kinda like almost like an airplane going down*.

A model of grounding in conversation known as the *contribution model* was proposed by H. H. Clark and Schaefer (1989) to formally capture the provisional nature of utterances. This model holds that the status of any utterance in the dialogue (i.e., whether it is a bona fide contribution to common ground) is uncertain until sufficient evidence is available in the form of a partner's response. In this case, the response *It's ah straight down?* provided evidence that the partner (B) recognized that she did not yet understand and that more work was needed. This clarification question from B could not be depended on to be in common ground until it was accepted by A, with a reconceptualized perspective: *Yeah, it looks like it has a point . . . kind of like a wing or wings*. The reconceptualized perspective then became part of the pair's common ground after the confirming evidence: *OK I got it—alright—yeah*. On the contribution model and the theory of grounding that underlies it, these small pieces of language (Yngve's backchannels, 1970) are clearly not empty verbalizations that control the rate of information flow (as presumed by Rosenfeld, 1987), but are specific metalinguistic signals that support the task of grounding (Brennan, 1990, 2005; Clark, 1996). If the official business of a dialogue is considered to be in Track 1, then Track 2 contains what H. H. Clark calls *collateral signals*:

> The claim is this: Every presentation enacts the collateral question *Do you understand what I mean by this?* The very act of directing an utterance to a respondent is a signal that means *Are you hearing, identifying, and understanding this now?* (H. H. Clark, 1996, p. 243)

Speaking spontaneously in dialogue includes a number of tasks such as getting an addressee's attention, ensuring that an utterance can be not only heard but also understood, and ascertaining whether intentions have been recognized and taken up (Bavelas et al., 1995; Brennan, 1990; H. H. Clark, 1996; H. H. Clark & Brennan, 1991; H. H. Clark & Schaefer, 1989). The phenomena illustrated in this chapter such as self-interruptions, ungrammatical stretches of speech, pauses, hedges such as *kinda like*, backchannels such as *uh huh*, interjections such as *right yeah good great*, and mutual gaze are generally considered uninteresting by many linguists and outside of the kind of language worth modeling by many psycholinguists. However, these elements are deployed in a way that is actually quite orderly (a point made early and often by conversation analysts) and are resources for the grounding process.

The Contribution Model. In the contribution model, each contribution to a conversation has a presentation phase (the utterance) and an acceptance phase (the evidence from one or more subsequent utterances about a partner's understanding and uptake that follows). A speaker evaluates the evidence provided by an addressee's response in comparison to the response that was expected; she can then revise her utterance and try again. The contribution model is a significant structural improvement over adjacency pairs, as its graph notation not only structurally pairs the two relevant utterances that form a joint action, but also nests them into the larger structure in which they play a role, as in this example that includes a nested repair (Cahn & Brennan, 1999):

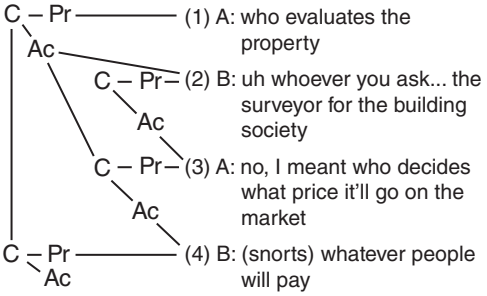


SOURCE: Cahn and Brennan (1999, p. 25).

The contribution notation was critiqued as being difficult for discourse analysts to

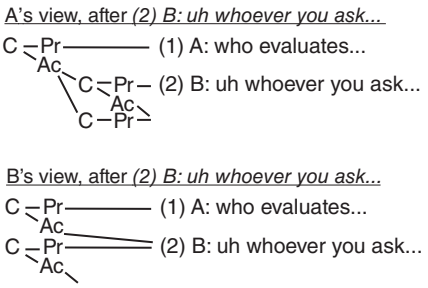
apply to an existing transcript (Traum, 1994); although each utterance participates in the structure as both presentation and acceptance at the leaf node level, it can be hard to tell what the role an utterance plays within the graph structure at a higher level. The source of this problem is that H. H. Clark and Schafer's (1989) original notation confounds the perspectives of both partners into a single representation that captures, post-hoc, only the transcript's product, and therefore fails to adequately represent the incremental nature of repairs as the coordination of the mental states of two distinct partners (Cahn & Brennan, 1999). In any dialogue in which a repair becomes necessary, one interlocutor will typically notice the need for a repair before the other does. In the preceding diagram, B appears to not have heard A, so B is the first to notice the problem; after he utters *what*, then A becomes aware of the need for a repair. Thus, as a dialogue unfolds, the interlocutors' private models will regularly be out of sync.

Extending the Contribution Model. Consider the next example, which includes a repair marked by *No, I meant*:



SOURCE: H. H. Clark and Schaefer (1989, p. 277).

This contribution graph represents two disjointed perspectives at once (as pointed out in Cahn & Brennan, 1999). The node that links utterance (2) to (3) is unrooted because B believes that by uttering (2), he has accepted (1), whereas upon hearing (2), A realizes that B has not understood her question and is preparing to initiate a repair with (3), so the acceptance phase will be more complex than just (2). A more accurate representation would capture this difference as two distinct, incremental representations held by A and B, in which the contributions' structures of presentation and acceptance phases are revised moment by moment as the evidence rolls in. The notation for the contribution model needs to be extended to represent the distinct, incremental perspectives of each partner, as follows.



SOURCE: Cahn and Brennan (1999), p. 27.

The metaphor of a process of *joint hypothesis testing* (Brennan, 1990, 2004) extends the contribution model to capture these momentarily divergent representations. A speaker's utterance represents her hypothesis about what an addressee is likely to understand and take up, whereas an addressee forms incremental (repeatedly revised) hypotheses about what the speaker intends, as the utterance unfolds. Both parties then test and revise their hypotheses as more evidence accrues. The hypothesis-testing metaphor (and representing the two partners' perspectives as separate graphs that are incrementally revised based on the evidence) accommodates the nondeterministic, probabilistic nature of grounding, as two partners can never be completely sure that their perspectives are aligned, but only that there is evidence that they converge closely enough for current purposes.

A further challenge for modeling contributions (at least for implementing a discrete notation), one that needs to be handled in order to achieve a psychological account of dialogue, is the problem of incrementality. The contribution notation requires that dialogue be segmented into constituents that can be grounded (not unlike the turn-constructional units hypothesized to play a role in turn-taking by Sacks et al., 1974). Language processing studies that use continuous measures (e.g., mouse cursor movements in Brennan, 1990, 2005; poising gestures in H. H. Clark & Krych, 2004; and eye gaze in Tanenhaus et al. 1995) have confirmed that the interpretation of spoken utterances is highly incremental. In a study of grounding in which a director could (or could not) see the matcher's progress as she moved a cursor across a map, contributions were structured quite differently when the director had visual information; matchers provided frequent spoken backchannels when they knew directors could *not* see the cursor, and were often silent when they knew the director *could* see the cursor (Brennan, 1990, 2005). In this way, visual evidence that comes from one party's actions in a joint task can count as an acceptance in the grounding process. This means that, depending on whether the communication medium allows for visual evidence, utterances can be presented and accepted to some degree in parallel, or at least in very fine-grained increments.

Conversation analysts provide many descriptions of such fine-grained coordination. In this next example (adapted from Goodwin, 1981, p. 60), Ethyl starts to speak while Barbara is looking elsewhere; so Ethyl interrupts and restarts herself mid-word, presumably in order to get Barbara's attention. At that point, Barbara begins to move her gaze (dotted line), which arrives at Ethyl's face at the point in time marked by X, and Barbara continues to gaze at Ethyl while Ethyl continues speaking (solid line).

Ethyl: So they st- their clas ses start
around (0.2) in
|
Barbara: ••••••••••••••X_____

It is unclear what the limits may be for coordination between speakers and addressees. Concerning the grain of interaction and the extent to which one partner's representation may lag another's, a method called *cross-recurrent gaze analysis* has been used to document the dynamics and temporal lag between interlocutors' producing and interpreting speech in dialogue. Two interlocutors' eye gaze over a shared display is most coupled when compared at a delay of two seconds (with the speaker's gaze leading the addressee's); those pairs with more closely coupled gaze were shown to have communicated more successfully than those with less closely coupled gaze (Richardson & Dale, 2005). Such continuous measures collected in parallel from interacting partners may hold promise for uncovering any limits to the grain at which interlocutors can coordinate (which appears to depend in part on a conversation's context, purpose, and communication channel; H. H. Clark & Brennan, 1991).

FUTURE DIRECTIONS, EMERGING TRENDS

Research in discourse and dialogue encompasses topics and findings from multiple disciplines that have wide application—not only to improving writing for the reader's comprehension, as we described earlier, but also to real-world problems in law, education, cross-cultural relationships, human–machine interaction, technology for remote or multimedia communication, and improving our

understanding of how the brain supports language and communication. Here we will cover two directions along which research has been advancing at a particularly rapid rate: spoken dialogue systems and the neural bases of discourse and dialogue.

Spoken Dialogue Systems

Chatterbots are automated dialogue systems that do not perform useful tasks, but simulate dialogue behavior. In 1966, Weizenbaum created the first chatterbot Eliza, a text-based system that had no capacity for intelligent or task-based behavior whatsoever. Users typed freely to a terminal and received responses modeled loosely on the response style of a Rogerian therapist (where the therapist reflects back to the patient what the patient has just said, in order to create an impression of understanding, e.g., *User: I feel depressed. Eliza: Why do you feel depressed?*). Weizenbaum reacted with alarm when he found his secretary so engaged with Eliza that she asked him to leave the room. He concluded that anthropomorphic dialogue interfaces could deceive users into thinking that the underlying systems were intelligent, and that such systems were possibly unethical and might be capable of harm (Weizenbaum, 1976), a position taken up later by Shneiderman (1987). However, critics who fear the consequences of attributing intelligence to machines may be making the error of failing to attribute it to people (who are probably less naïve than they expect). Further, most people can adapt to a wide variety of dialogue partners, including artificial ones (for an early instance of this debate, see Don, Brennan, Laurel, & Shneiderman, 1992). There is great entertainment value in playing with and testing the boundaries of such systems, whose limitations are obvious enough to prevent them from passing the Turing test (in this case, not passing for human).

Chatterbots aside, spoken dialogue systems that can actually accomplish tasks have come a long way since Put That There (Schmandt & Hulteen, 1982). Most people regularly encounter automated telephone systems when seeking travel information, doing banking or checking credit card balances, or struggling with situations that require customer support. Such systems present prompts in the form of recorded or synthesized utterances and allow (or require) users to speak in response (although many offer keypad input options also). Virtually all of these systems are constrained to a small set of tasks and are scripted to follow a specific menu; the most tedious among them recite exactly what a user can opt to say, and accept no responses that depart from those options. More usable systems allow people to take the initiative to enter information in flexible order to create a query, or invite the user to "Tell me what you're calling about" in order to categorize and channel the topic to a particular subscript. In addition to the typical components of spoken dialogue systems (automated speech recognition, natural language processor, natural language generation, and speech synthesis), today's systems also include dialogue managers that track the state of the dialogue in order to generate next moves within a well-understood task domain (Larsson & Traum, 2000; Traum, 1994; Wachsmuth, 2008; Williams, Raux, & Henderson, 2016).

One relatively successful telephone dialogue system designed for the public and implemented in 2001 was Amtrak's Julie, a simulated customer service agent described as "unshakably courteous and tirelessly chipper. . . . Many riders say that she sounds and acts so lifelike that they did not immediately realize that she was just a computer program" (Urbina, 2004). Julie embodied an informal conversational style and elicited information for queries about train travel in a

breezy but scripted way (*Let's get started!*), grounding the conversation by accepting users' utterances with feedback such as *Got it* and *I think you said 5 o'clock, am I right?* This was natural enough for this task; but delightfully, soon there were televised parodies by *Saturday Night Live* depicting Julie on a date or at a cocktail party (e.g., retaining her chipper customer service register while interacting with a potential romantic interest, who seemed strangely unaware that anything was off, despite Julie's oddly high grounding criterion and insistence on repeating things back to him). Despite the entertainment value inherent in using an anthropomorphized system such as this one, it did serve the information access needs of many users from the general public.

Another example of an agent that was used by the general public was that of Max, an animated guide to a museum in Paderborn, Germany (Kopp, Gesellensetter, Krämer, & Wachsmuth, 2005). Although the primary task of this agent was to provide visitors with information about the museum, its exhibitions, or other topics of interest, it appears that many of the dialogues were inspired by the system's novelty, as opposed to the task (Max's creators reported that users frequently tried to flirt with Max; Kopp et al., 2005).

Today's dialogue systems, though still largely limited to routine tasks within a specific domain, have improved substantially in both usability and naturalness over the past two decades. Improvements have been made due to better speaker-independent speech recognition technology as well as context-specific feedback that is relevant to a user's previous utterance (Brennan & Hulteen, 1995; Mizukami, Yoshino, Neubig, Traum, & Nakamura, 2016) and helps the user to identify errors and initiate repair sub-dialogues. Systems have been programmed to become more adept dialogue partners, with an improved ability to detect and recover from errors (e.g., Marge & Rudnicky, 2015) or predict and interpret feedback (Hough, & Schlangen, 2016; Morency, de Kock, & Gratch, 2010) or engage partners (Yu, Nicolich-Henkin, Black, & Rudnicky, 2016); grounding strategies may be implemented in disembodied audio dialogues, by animated agents, or by robots. Some current systems use natural prosody as by inserting small hesitations or pauses where a human partner would have had to look up information, as well as by using spoken prompts that appropriately stress new information and de-stress given information (for discussion and additional strategies, see Cohen, Giangola, & Balogh, 2004).

New spoken dialogue systems in use by the public include smartphone agents such as Apple's Siri, Amazon's Alexa, Facebook's M, Google's Google Now, or Windows' Cortana. These agents, though still limited to well-defined tasks, are more open-ended in the topics and kinds of utterances they can handle, and can sometimes perform tasks from more than one application domain, such as booking a flight, finding a nearby restaurant, answering a question about the weather, or invoking the Internet to search for information in response to a general knowledge question (even if they do not yet seamlessly connect smaller tasks from these domains into a higher level goal). Some of these systems are being designed to learn from their experience with a particular task or user. They are sometimes programmed to provide entertaining responses to questions they cannot interpret or act upon, although they have far to go in their pragmatic knowledge (being able to behave in socially authentic ways).

Advanced spoken dialogue systems that are being developed or simulated in laboratories include so-called intelligent personal assistants intended to build rapport as well as perform tasks; for instance, a humanoid animated agent, Sara, makes recommendations

to conference goers as well as small talk (Matsuyama et al., 2016). The next frontier is likely to be populated by flexible systems that can make inferences across multiple tasks (e.g., Lee & Stent, 2016), derive general knowledge from corpora and online information sources (e.g., Rahimtoroghi, Hernandez, & Walker, 2016), or learn perceptually grounded concepts from their human interlocutors (Y. Yu, Eshghi, & Lemon, 2016).

In sum, though having a conversation feels effortless for most people, conversational behavior is still quite a challenge to achieve in human–machine interaction. Many practical issues remain to be addressed in spoken dialogue systems, including the simultaneous processing of social and task-related goals; the establishment of trust and rapport; concerns about technology and privacy; the ability to represent, learn about, and understand real-world contexts; and the determination of what sort of interactive partner a spoken dialogue system should model.

Neural Bases of Discourse and Dialogue

> Turning finally to social factors that influence the shape of language, we need to keep in mind that conversation consists of interactions between separate minds and separate selves. Language is the preeminent way of compensating for the fact that our separate brains lack direct neural links. But our brains are the properties of separate selves, each with its own self-centered agenda. How communication between these separate selves is managed, both collaboratively and not-so-collaboratively, is more than a matter of taking turns. (Chafe, 2002, p. 258)

From Isolated Words to Coherent Text: Processing Discourse Involves Additional Neural Structures

The number of neuroscience studies in the field of communication and discourse processing has grown tremendously in recent years (e.g., for reviews, see Bornkessel-Schlesewsky & Friederici, 2007; Ferstl, 2010; Ferstl, Neumann, Bogler, & von Cramon, 2008; Mar, 2004; Mason & Just, 2006). These studies have gone beyond the processing of single words or sentences, in order to investigate the neural underpinnings of reading or listening to complete narratives, and to capture the neural activity that ensues during spoken or textual communication with another person. Comprehending narratives or other genres of connected discourse is more than processing a sequence of individual sentences; as we discussed previously, discourse processing requires making inferences for connecting these sentences, interpreting referring expressions, building and maintaining situation models, using background knowledge and discourse context, addressing discourse to specific audiences, and interpreting pragmatic cues to metaphors, irony, or indirect speech.

Converging evidence from neuroscience studies of discourse processing has identified a network of brain areas that include, but also go beyond, the left perisylvian areas traditionally associated with language: Broca's and Wernicke's. In comparisons of processing of coherent text versus incoherent or isolated text (e.g., scrambled sentences or word lists), an *extended language network* (Ferstl et al., 2008) appears to consistently engage the bilateral anterior temporal lobes, the superior temporal sulcus, the left inferior frontal gyrus, and the right-hemisphere counterpart of Wernicke's area. In addition, several medial regions have been proposed to be involved, such as the dorsomedial prefrontal cortex (dmPFC), and the precuneus (PC). The individual functional contributions of these areas are still being uncovered (e.g., see Ferstl, 2010). But it has become clear already that studying language in a context that captures more closely how language

is used in everyday life has revised and extended our understanding of language and the brain.

The vast majority of these studies focus on language comprehension; there has been little neuroscientific work on speech production during communication. This is presumably due not only to the methodological challenges inherent in achieving sufficient control over spontaneous speaking (given the enormous variability in speakers' expressive choices), but also to the specific limitations imposed by imaging technology such as motion artifacts due to speaking in the scanner or while EEG (electroencephalography) signals are being recorded. Moreover, although an increasing number of behavioral studies of discourse processing focus on language use during social interaction, few neuroscience studies have done so. Typically in the latter, linguistic material is presented without information about who is speaking or who is being addressed, and subjects have no opportunity to formulate a reply or to engage in actual interaction (just as in behavioral language-as-product studies). Using and processing language during social interaction and for the purpose of communicating is likely to impose an additional set of processing constraints, so the extended language network may need to be functionally refined or further extended. The neuroscience of communication, especially in dialogue contexts and including speech production, is a rather new, emerging field. In the following we will discuss some recent advances.

Speaking for the Purpose of Communicating

We return to the topic of audience design. To what extent might the neural resources that support linguistic processing without an intention to communicate be *distinct* from those that support communication? A functional magnetic resonance imaging (fMRI) study by Willems et al. (2010) had subjects in the scanner play the game Taboo with a confederate partner located outside of the scanner. In this popular game, speakers describe a basic term (e.g., *beard*) without being able to use a predefined set of associated words (e.g., *hair*, *man*, *shave*). Crucially, linguistic difficulty of the task and communicative intent were manipulated independently from each other. Linguistic difficulty was varied by how closely related the target was to the banned taboo words. And communicative intent was manipulated by varying what speakers assumed about their partner's needs: In the communicative condition, speakers were told their partners would have to guess the target word, whereas in the noncommunicative condition, speakers were told that their partners already knew the target word. The result was that distinct brain areas were activated by linguistic difficulty versus communicative intentent: Whereas the manipulation of linguistic difficulty engaged the inferior frontal and inferior parietal cortex, the manipulation of communicative intent engaged the dmPFC. This latter area has frequently been shown to be involved in perspective-taking and the ability to infer the mental states of another person (mentalizing; for reviews, see Amodio & Frith, 2006; Bzdok et al., 2013). However, one area of the brain, the left posterior superior temporal sulcus (pSTS), was responsive to both linguistic and communicative factors and activated more strongly in linguistically difficult trials with communicative intent.

Willems and colleagues' (2010) study suggests that the communicative and linguistic requirements of speech production draw upon some distinct mechanisms. Mentalizing appears to be an essential skill in the planning of communicative actions (as we will discuss presently), particularly when speech is adapted to what a conversational partner is presumed to believe and know.

This interpretation is complemented by a recent study of ours in which we found the medial prefrontal cortex (mPFC) to be a core neural structure that encodes information about the upcoming speech context (Kuhlen, Bogler, Brennan, & Haynes, 2017). In this study, subjects in the scanner gave simple spatial instructions to a (confederate) partner located outside of the scanner on how to place colored pieces on a game board of large colored squares (e.g., *red on blue*) via a live video stream. In half of the trials, subjects were told that instead of addressing the partner they would need to "test the new MRI-proof microphone." Hence, subjects either communicated with a conversational partner (and could witness the partner executing the instructions), or they produced virtually identical speech but not for communicative purposes (and could also observe the partner via the video stream). In both conditions, data were collected during the preparation phase, just before speaking began, when subjects knew the context under which they would be speaking (to a partner or to test the microphone), but did not know yet which instruction they would need to give. This allowed us to separate processes associated with preparing to speak in a communicative versus noncommunicative context from processes associated with speech production. We applied a pattern classification technique known as multivariate searchlight analysis that combines information across multiple voxels of the brain (as opposed to the more conventional mass-univariate analysis that compares single voxels). This technique enables insight into brain regions that encode information on the task condition (see e.g., Haynes, 2015; Haynes & Rees, 2006).

Our analyses revealed that the ventral bilateral prefrontal cortex (vlPFC) encoded information that differentiated the two upcoming tasks; even more relevant to the question of audience design, the ventromedial prefrontal cortex (vmPFC) was also involved. The vlPFC has previously been associated with prospective task representation (e.g., Momennejad & Haynes, 2012), and the vmPFC has been found to be engaged in tasks that required person-specific mentalizing that is tailored toward the idiosyncratic characteristics of a particular individual (Welborn & Lieberman, 2015). Patients with lesions in the vmPFC have been reported to show an inability to tailor communicative messages to specific characteristics of their conversational partner (Stolk, D'Imperio, di Pellegrino, & Toni, 2015). Our findings suggest that the brain engages in preparatory neural configurations (task sets) that may support adaptation to a conversational partner early in speech planning, as a form of audience design (as opposed to during late, strategic repair processes). Moreover, our study corroborates the role of the mPFC during language use in communication.

Together, these and other studies (e.g., Rice & Redcay, 2015; Sassa et al., 2007) suggest that processing language for the purpose of communicating with a conversational partner shows patterns of neural activation that are distinct from those involved in processing language outside of a communicative context. One core area of the so-called mentalizing network, the mPFC, seems consistently engaged when communicating with, and possibly adapting to, a conversational partner. Next we will consider in more detail the possible role of the mentalizing network in partner-adapted communication.

Mentalizing: Perspective Taking in Partner-Adapted Communication

Mentalizing, the ability to take into account others' perspectives and draw inferences about their mental states, is of central interest in the study of communication (Brennan et al., 2010). Typically, the neural basis of mentalizing, the so-called mentalizing

network, has been associated with the mPFC, the bilateral temporoparietal junction (TPJ) and the PC (for an overview, see Van Overwalle & Baetens, 2009). Areas of this mentalizing network have been shown to be engaged while readers or listeners process ironic utterances (e.g., Spotorno, Koun, Prado, Van Der Henst, & Noveck, 2012), comprehend indirect speech acts (e.g., Bašnáková, van Berkum, Weber, & Hagoort, 2015; Bašnáková, Weber, Petersson, van Berkum, & Hagoort, 2014), or interpret indirect requests (e.g., van Ackeren, Casasanto, Bekkering, Hagoort, & Rueschemeyer, 2012). These findings suggest that inferences about another person's mental state are needed to correctly interpret the communicative intention behind indirect utterances.

The ability to take another person's perspective not only seems to facilitate one's own comprehension, but is also engaged when making judgments about a conversational partner's understanding of utterances. In a recent study using EEG, subjects were presented with short narratives either with or without a confederate partner present (Rueschemeyer, Gardner, & Stoner, 2015). In the critical condition, the target sentence was rendered plausible only in conjunction with the preceding context sentence (context sentence: *In the boy's dream, he could breathe under water*; target sentence: *The boy had gills*). Crucially, the first (context) sentence was presented exclusively to subjects via headphones (but not to their partners). Thus, though target sentences were plausible to subjects, subjects knew that the targets were implausible to their partners. After hearing the second sentence subjects were asked to judge how well they and their partner had understood the target sentence. The comparison condition had a similar target sentence that was plausible without a contextualizing sentence (context sentence: *The fishmonger prepared the fish*; target sentence: *The fish had gills*). Subjects showed a more pronounced negativity 350 to 550 milliseconds after onset of the critical sentence final word (known as the N400 effect) when listening to the sentences with a partner compared to without a partner. Hence, they showed the well-known electrophysiological marker of semantic integration difficulties occurred in reaction to their *partner's* inability to understand the target sentence, even though they themselves had no difficulties understanding the sentence. This implies that a conversational partner's knowledge and level of understanding can be tracked by processes closely related to those aiding one's own language comprehension. What remains unclear from this study is whether others' understanding is tracked routinely or only when explicitly queried by the (experimental) task.

As for whether conversational partners routinely draw upon perspective-taking when engaged in conversation, the evidence from neuroscience is mixed. A recent study (Bögels et al., 2015) recorded subjects' brain activity using magnetoencephalography (MEG) while the subjects interacted with two confederate partners in a referential communication game using a design similar to Metzing and Brennan's (2003) study described previously. In the first phase of live interaction, subjects and one of the confederate partners established a precedent of using particular terms to refer to reoccurring objects in the game ("grounding phase"). In a second "test" phase, subjects' brain activity was recorded while the initial confederate partner or a new confederate partner either referred to objects that had previously been established using a different term (e.g., calling an object *sofa* although it had been called *couch* during the grounding phase) or else referred to objects that had not been established in the grounding phase. In this way, the confederate partners either abandoned an

established term (amounting to the sort of pragmatic violation that Brennan and Clark, 1996, called "breaking a conceptual pact" when this was done by the confederate who had established the precedent) or referred to an entirely new object (no precedent). An analysis of subjects' brain activity just after seeing the object, but prior to hearing the referring expression, identified activity in brain areas related to language processing and episodic memory. These areas were more strongly activated when subjects expected the confederate they had previously interacted with to name an object for which they had an established precedent (that is, a conceptual pact with that partner), than when they expected the previous confederate to name an object for which no precedent had been established at all). In contrast, no difference in activation was found between these two naming conditions when subjects expected to interact with the new confederate. None of the conditions found any activation of the mentalizing network during this first time period.

The authors interpret this finding as evidence that basic cognitive processes such as retrieval from memory of previously used terms underlie partner-specific processing, and that contextual influences from mentalizing and assessing common ground do not. Activity in the mentalizing network, most notably the vmPFC, the right TPJ, and the PC, was detected only in the second time period, 200–800 ms after the initial partner used a new term (i.e., broke the conceptual pact that the partner had previously established with the subject). The authors argue that mentalizing becomes engaged only on demand in reaction to a pragmatic violation, but is not engaged spontaneously or in advance to guide listeners' expectations about their common ground with the current speaker. This finding is characteristic of accounts that assign perspective-taking and common ground only a peripheral role in communication (see e.g., Kronmüller & Barr, 2015; Pickering & Garrod, 2004), and that assume that reasoning about another person's perspective is effortful and done only as a kind of repair when needed.

However, Bögels et al. (2015) does not establish that mentalizing occurs only in response to pragmatic violation (especially since there was no experimental condition in which precedents were maintained)—only that it occurs in response to hearing a linguistic expression (as opposed to *anticipating* hearing one). In addition, the expectation that the partner should continue to use the established term would have been significantly weakened, since subjects experienced 80 instances over the course of the experiment in which their initial partner departed from this expectation (actually, twice as often as the partner maintained the precedent). This matters because infelicitous behavior on the part of a dialogue partner has been shown to change the nature of the language game in which subjects are engaged, making the findings of questionable generalizability to spontaneous spoken dialogue. In the original study of conceptual pacts (Metzing & Brennan, 2003), each subject experienced a total of only two broken pacts. Subsequently, a replication of that study in 3- and 5-year-olds (Matthews et al., 2010) found that the effect was smaller for the second broken pact than for the first one, suggesting that people (or children, at least) are highly sensitive to infelicity. As for whether mentalizing is involved only on demand as claimed by Bögels et al. (2015), this does not seem to be supported by the rapidity with which the mentalizing network was activated in the second time interval. Other neuroscientific findings suggest that another person's perspective can be taken into account automatically and without effort (Ramsey, Hansen, Apperly, & Samson, 2013; Rice & Redcay, 2015).

Mirroring: Simulating a Partner's Communicative Intention

Apart from the mentalizing network, another network has been hotly debated in the context of social interaction and communication. The so-called mirroring network typically involves the pSTS, the premotor cortex, and the anterior intraparietal sulcus (Van Overwalle & Baetens, 2009). Mirroring is said to facilitate social interaction by simulating another person's motor actions, thereby providing a neural mechanism for understanding and predicting the actions of others. By engaging processes that are comparable to those engaged when performing the action oneself, the mirroring network is proposed to encode not only others' actions (what they are doing), but also the intention behind others' actions (why they are doing what they're doing; Iacoboni et al., 2005). In the context of communication, it has been suggested that the mirror neuron system may enable mutual understanding by means of an automatic sensorimotor resonance between the sender of a message and its receiver (Rizzolatti & Craighero, 2004). Though this proposal ties in with some theories of dialogue (e.g., Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Pickering & Garrod, 2004), it has been criticized by others for being insufficient to bridge the gap between a given linguistic code and the meaning intended by the speaker (e.g., Hagoort & Levinson, 2014; Noordzij et al., 2009; Stolk et al., 2014; Stolk, Verhagen, & Toni, 2016; see also the previous sections in this chaper on the vocabulary problem and message models).

Many scholars have proposed that mentalizing and mirroring may be complementary mechanisms that each contribute to successful communication. But how exactly the mentalizing and mirroring systems may work together is still a matter of debate (for discussion see, e.g., Brennan et al., 2010; de Lange, Spronk, Willems, Toni, & Bekkering, 2008; Keysers & Gazzola, 2007; Spunt & Lieberman, 2012; Waytz & Mitchell, 2011; Zaki, Hennigan, Weber, & Ochsner, 2010). On one hand, a meta-analysis of over 200 fMRI studies came to the conclusion that the two networks are rarely active at the same time (Van Overwalle & Baetens, 2009), suggesting that they may be functionally independent from each other (however, it is important to note that most if not all of the tasks involved in the fMRI studies surveyed did not involve actual communication with a social partner). On the other hand, several studies have explicitly investigated the relationship between these two networks and have proposed a more integrative account (de Lange et al., 2008; Keysers & Gazzola, 2007; Lieberman, 2007; Olsson & Ochsner, 2008; Spunt & Lieberman, 2012; Spunt, Satpute, & Lieberman, 2011; Thioux, Gazzola, & Keysers, 2008). One such proposal has been that mirroring supports the perceptual encoding of observable motor behavior, whereas mentalizing supports the interpretation of this behavior with respect to another person's underlying mental states (e.g., Spunt & Lieberman, 2012).

This proposal is in line with an fMRI study investigating how nonverbal displays of a person's knowledge state are processed in the brain of an observer (Kuhlen, Bogler, Swerts, & Haynes, 2015). As we discussed earlier, drawing inferences about another person's knowledge is generally considered to be an important prerequisite for partner-adapted communication. Previous behavioral work has identified verbal markers (e.g., speech disfluencies) as well as nonverbal markers (e.g., length of a pause or facial displays) that are informative about speakers' mental states, reflecting how committed speakers are to an utterance they are producing (Brennan & Williams, 1995; Smith & Clark, 1993; Swerts & Krahmer, 2005). These markers are also used by observers to make

reliable judgments about speakers' mental states (Brennan & Williams, 1995; Swerts & Krahmer, 2005). In Kuhlen, Bogler, et al.'s (2015) fMRI study, subjects watched silent video recordings of nonverbal facial displays of people responding to general knowledge questions (e.g., *What is the capital of Switzerland?*; see Swerts & Krahmer, 2005). After each video, subjects were asked to indicate how confident the person in the video seemed in their answer.

While watching these videos, subjects showed activation in brain areas associated with both the mentalizing and mirroring networks. Crucially, however, only activity in areas of the mentalizing network was modulated by the content of mental state inferences: The less confident that subjects perceived the respondent to be, the more active were core areas of the mentalizing network, namely, the bilateral TPJ and the mPFC. No modulation of the mirroring network was observed in response to subjects' perception of the respondent's confidence. This finding suggests that the mirroring and mentalizing networks are distinct but not independent, and are able to serve complementary functions that facilitate inferences about another person's mental state. Whereas mirroring may assist the perceptual encoding of overt motor behavior such as nonverbal facial displays, mentalizing may be instrumental in making sense out of the observed behavior.

Multi-Brain Approaches to Language Processing in Social Settings

There has been a recent movement in social neuroscience toward investigating how multiple brains coordinate with each other in social interaction (for review, see, e.g., Hari, Henriksson, Malinen, & Parkkonen, 2015; Konvalinka & Roepstorff, 2012). This movement accompanies a call for more ecologically valid experimental paradigms in which subjects are directly addressed during a dialogue, or are actually engaged in interaction (Holler et al., 2015; Schilbach, 2015; Schilbach et al., 2013) instead of being passive observers of social stimuli. This movement has also influenced neuroscientific approaches for studying language use in communication (for review, see Kuhlen, Bogler, et al., 2015; Willems et al., 2015).

In one pioneering study, a speaker's brain activity was recorded in an fMRI scanner while the speaker spontaneously told an autobiographical story (Stephens, Silbert, & Hasson, 2010). An audio recording of this narration was then presented to listeners while their brain activity was recorded. Neural activity recorded during speaking was then compared to neural activity recorded during listening with the goal of detecting coordination of neural activity between speaker and listeners. Indeed, correlational analyses between the speaker's and listeners' neural data showed that brain areas engaged during the production of the narration were also engaged during its comprehension. Brain areas that showed coordination between speaker and listeners were those associated with low-level auditory processes and areas related to language processing (e.g., Wernicke's and Broca's area) as well as areas related to mentalizing (e.g., PC, mPFC), suggesting that speaker–listener coordination took place across different levels of processing. In most of these areas, activity in the listeners' brain lagged up to three seconds behind the speaker's brain activity. Remarkably, this lag is not far off from the two-second lag between the gaze of an optimally communicating speaker and an addressee while discussing a visual display, as detected by the cross-recurrent gaze analysis technique (Richardson & Dale, 2005). And just as with gaze, the degree of coordination between speaker's and listeners' brain activity was related to communication

success, measured by listeners' performance in a subsequent knowledge questionnaire testing their comprehension of the narration. This implies a functional link between interbrain coordination and successful communication. Such an interpretation was further corroborated by a lack of significant interbrain coordination when monolingual English subjects listened to a narrative told in Russian (Stephens et al., 2010).

Studies like Stephens et al. (2010), together with other multi-brain studies on nonverbal communication (Anders, Heinzle, Weiskopf, Ethofer, & Haynes, 2011; Bilek et al., 2015; Schippers, Roebroeck, Renken, Nanetti, & Keysers, 2010), have led to a proposal of a brain-to-brain coupling mechanism for transmitting information between communicating individuals (comparable to a coupling between action perception and action execution in one individual's brain; Prinz, 1990). Based on a parity of representations in sender and receiver, shared understanding presumably occurs by evoking similar patterns of brain activity in the person listening and the person speaking, achieved entirely through speech (or other communicative) signals (Hasson et al., 2012). As with mirroring (or alignment through priming) accounts of communication like the interactive alignment model (Pickering & Garrod, 2004), it remains to be shown how or whether such a largely automatic coupling mechanism can account for the rich pragmatic inferences that are made seemingly effortlessly in naturally occurring social interaction.

Interbrain coordination between speakers and listeners is not limited to coordination of identical brain areas in speakers and listeners. In an EEG study listeners' brain activity was recorded while listening to a speaker telling stories of about two minutes in length (Kuhlen, Allefeld, & Haynes, 2012). Crucially, the video that listeners were presented with consisted of two superimposed speakers narrating simultaneously two different types of stories. One group of listeners was instructed to attend to one story, and the other group of listeners was instructed to attend to the other story. Although both groups of listeners were presented with comparable low-level perceptual input and viewed the same superimposed video image, they attended to different higher level, discourse-related aspects of the video. Listeners who attended to the same story had more similar EEG to each other than they did to listeners who attended to the other story. Moreover, the correlation between the listeners' and the attended speaker's EEG revealed that their brain activity was coordinated, but with the listeners lagging at about 12.5 seconds. The authors propose that speaker–listener coordination corresponds to processing linguistic information at different grains. The rather long time lag at which speaker and listeners coordinated in this study may correspond to the production and comprehension of larger units of linguistic information, possibly at the level of a situation model. Coordination on smaller units of information (e.g., words) may have been hampered by the difficulty of comprehending every single word due to the superimposing of videos. Notably, coordination involved not only spatially corresponding brain regions, but also distinct areas in the brain of the speakers and the brain of the listeners, including activation at medial-frontal electrode locations only in listeners. This suggests an involvement of mentalizing rather than simply a mirroring mechanism in the coordination of speaker–listener neural activity.

Studies like those reviewed in this section have succeeded in investigating two or more brains producing and comprehending the same complex, naturalistic linguistic material. Results have produced

interesting insights into the temporal and spatial dimension of interbrain coordination between verbally communicating individuals. However, so far there have been no comparable investigations of language use in truly interactive settings (which would, of course, be methodologically challenging; for discussion, see Kuhlen, Allefeld, Anders, & Haynes, 2015). Thus, it is unclear how these findings will scale up to scenarios in which conversational partners take turns speaking and listening and can interact with each other in a fully contingent fashion (note that interactive social encounters have been investigated in the context of nonverbal or motor coordination; see, e.g., Bilek et al., 2015; Dumas, Nadel, Soussignan, Martinerie, & Garnero, 2010; Konvalinka et al., 2014). Another limitation of the studies surveyed in this section is that the measures of interbrain coordination uncovered correspond to coordination over a larger time period. To our knowledge, temporally more fine-grained analyses that could reveal a moment-by-moment coordination of neural activity have not been developed or implemented yet. It will be interesting to see how future studies on multi-brain coordination will be able to address the interactive and incremental nature of language processing in dialogue context.

One promising advance in this direction is a recent dual-brain fMRI study that measured neural activity simultaneously in two individuals engaged in a computer-mediated nonverbal communication game (Stolk et al., 2014). In this game one person described to the partner where and how to position a target token on a grid simply by moving a mouse cursor (for a comparable experimental task see Noordzij et al., 2009). In some trials, interlocutors could use prior established strategies for solving a specific constellation of this communicative challenge, whereas in others they had to establish mutual understanding anew (*known* vs. *novel* trials). Interbrain coordination occurred in pairs with a shared communicative history, but not in pairs without a shared history. Specifically, in pairs with a shared communicative history the vmPFC (an area central to many of the studies reviewed above) and an anterior portion of the superior temporal gyrus (rSTG) were more active when using known compared to novel communicative strategies. Notably, increasing activity in the rSTG corresponded to pairs' increasing communicative success in establishing new conventions for solving the task. The pair-specific interpersonal coordination in this region may therefore reflect the process by which communicating pairs establish common ground and converge on a shared conceptual space. The fact that the observed pattern of interbrain coordination corresponded to shared communicative history, but did not time-lock to specific communicative events, speaks against theories that propose priming and automatically shared sensorimotor processes as the basic mechanisms for communication (for additional discussion, see Stolk et al., 2016). Instead, this study supports theories that emphasize the incremental and partner-specific processes by which conversational partners establish shared understanding through grounding.

CONCLUSION

Discourse and dialogue are the outcomes of a fundamental, complex, and universal kind of human experience—interaction that uses language as its primary currency. In this chapter, we have covered some basic findings, described some contrasting accounts and controversies, and highlighted approaches from multiple disciplines, including the cognitive sciences (psycholinguistics,

artificial intelligence, linguistics, and computational linguistics), as well as the disciplines of sociolinguistics, neuroscience, and human–computer interaction. These approaches employ a wide variety of measures as well as different grains of analysis. Each makes different assumptions about the forces that shape the production and interpretation of discourse and dialogue.

The methods typical within each discipline bring their own strengths and weaknesses. Spoken dialogue systems may simulate dialogue behavior and model joint actions in the service of one or several task domains, or possibly even learn from a context, corpus, or partner, but today's systems are only as good as their underlying data, domain model, and architecture. Ethnographic observations of spontaneous dialogue provide rich descriptive data that set the bar high for what needs to be modeled and explained, but such data resist generalizing and can be subjective and unreliable. Experiments can be replicated and can test detailed hypotheses about cognitive representations or processes, but if the need for control renders social context either inauthentic or missing entirely, then the object of study risks being transformed into a different kind of language game. Neuroimaging can sometimes succeed in distinguishing the underlying neural circuitry of two experimental conditions that may appear otherwise identical in the behaviors or reaction times they yield. However, subjects in neuroimaging experiments are highly constrained, as they may be required to lie motionless in a noisy scanner or wear an EEG cap while they listen to speech or communicate remotely (and as far as we know, no neuroscience study has produced any interesting transcripts of spontaneous language use).

We conclude that there is not only value in synthesizing an interdisciplinary approach to discourse and dialogue, but that such an approach is essential. No one discipline or approach provides a complete picture; taken together, the different approaches provide a wealth of insights about the cognitive, computational, social, and biological nature of discourse and dialogue.

REFERENCES

Akmajian, A., Demers, R. A., & Harnish, R. M. (1987). *Linguistics: An introduction to language and communication* (2nd ed.). Cambridge, MA: MIT Press.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268–277. doi:10.1038/nrn1884

Anders, S., Heinzle, J., Weiskopf, N., Ethofer, T., & Haynes, J.-D. (2011). Flow of affective information between communicating brains. *NeuroImage*, *54*(1), 439–446. doi:10.1016/j.neuroimage.2010.07.004

Anderson, J.R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum. doi: 10.4324/9780203780954

Ariel, M. (1990) *Accessing noun-phrase antecedents*. London, United Kingdom: Routledge.

Austin, J. L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press. doi:10.1093/acprof:oso/9780198245537.001.0001

Bakhtin, M. (1986). The problem of speech genres. In C. Emerson & M. Holquist (Eds.), *Speech genres and other late essays* (pp. 60–102). Austin: Texas University Press.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development*. Hillsdale, NJ: Erlbaum.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*(1), 1–22. doi:10.1006/jmla.1999.2667

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of*

Memory and Language, *46*(2), 391–418. doi:10.1006/jmla.2001.2815

Barr, D. J. (2008). Pragmatic expectations at linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*, 18–40.

Bartlett, F. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, United Kingdom: Cambridge University Press. doi:10.1017/CBO9780511759185.012

Bašnáková, J., van Berkum, J., Weber, K., & Hagoort, P. (2015). A job interview in the MRI scanner: How does indirectness affect addressees and overhearers? *Neuropsychologia*, *76*, 79–91. doi:10.1016/j.neuropsychologia.2015.03.030

Bašnáková, J., Weber, K., Petersson, K. M., van Berkum, J., & Hagoort, P. (2014). Beyond the language given: The neural correlates of inferring speaker meaning. *Cerebral Cortex*, *24*(10), 2572–2578. doi: 10.1093/cercor/bht112

Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and social Psychology Bulletin*, *21*, 394–405.

Bilek, E., Ruf, M., Schäfer, A., Akdeniz, C., Calhoun, V.D., Schmahl, C., . . . Meyer-Lindenberg, A. (2015). Information flow between interacting human brains: Identification, validation, and relationship to social expertise. *Proceedings of the National Academy of Sciences, USA*, *112*(16), 5207–8212.

Black, J. B., Turner, T. J., & Bower, G. H. (1979). Point of view in narrative comprehension, memory, and production. *Journal of Verbal Learning and Verbal Behavior*, *18*(2), 187–198. doi: 10.1016/s0022–5371(79)90118-x

Bobrow, D. G., & Collins, A. (1975). *Representation and understanding: Studies in cognitive science*. New York, NY: Academic Press.

Bock, J. K. (1977). The effect of a pragmatic presupposition on syntactic structure in question answering. *Journal of Verbal Learning and Verbal Behavior*, *16*, 723–734.

Bögels, S., Barr, D. J., Garrod, S., & Kessler, K. (2014). Conversational interaction in the scanner: Mentalizing during language processing as revealed by MEG. *Cerebral Cortex*, *25*(9), 3219–3234. doi: 10.1093/cercor/bhu116

Bolinger, D. (1977). *Meaning and form*. London, United Kingdom: Longman.

Bornkessel-Schlesewsky, I., & Friederici, A. D. (2007). Neuroimaging studies of sentence and discourse comprehension. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 407–424). Oxford, United Kingdom: Oxford University Press. doi:10.1093/oxfordhb/9780198568971.013.0024

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 717–726. doi:10.1016/s0022–5371(79)80006–9

Brennan, S. E. (1990). *Seeking and providing evidence for mutual understanding* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.

Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, *10*, 137–167. doi: 10.1080/01690969508407091

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-action traditions* (pp. 95–129). Cambridge, MA: MIT Press.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Brennan, S. E., Friedman, M. W., & Pollard, C. J. (1987). A centering approach to pronouns. In *Proceedings of 25th Annual Meeting of Association for Computational Linguistics* (pp. 155–162). Stroudsburg, PA: Association for Computation Linguistics.

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialogue: Coordinating speaking and understanding. In B. H. Ross (Ed.) *Psychology of learning and motivation: Advances in research and theory* (Vol. *53*, pp. 301–344). Cambridge, MA: Academic Press. doi:10.1016/s0079–7421(10)53008–1

Brennan, S. E. & Hanna, J. E. (2009). Partner-specific adaptation in dialogue. *Topics in Cognitive Science (Special Issue on Joint Action)*, *1*, 274–291.

Brennan, S. E., & Hulteen, E. (1995). Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-Based Systems*, *8*, 143–151.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*(3), 383–398.

Brown, P., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*(4), 441–472. doi:10.1016/0010–0285(87)90015–6

Brown-Schmidt, S. (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190.

Brown-Schmidt, S. (2009b). The role of executive function in perspective-taking during online language comprehension. *Psychonomic Bulletin and Review*, *16*, 893–900.

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, *27*, 62–89. doi:10.1016/j.jml.2015.05.002

Brown-Schmidt, S. & Fraundorf, S. (2015). Interpretation of informational questions modulated by joint knowledge and intonational contours. *Journal of Memory and Language*, *84*, 49–74. doi:10.1016/j.jml.2015.05.002

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*, 1122–1134. doi:10.1016/j.cognition.2007.11.005

Brown-Schmidt, S., & Hanna, J.E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue and Discourse 2*, 11–33. doi:10.5087/dad.2011.102

Bzdok, D., Langner, R., Schilbach, L., Engemann, D. A., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2013). Segregation of the human medial prefrontal cortex in social cognition. *Frontiers in Human Neuroscience*, *7*. doi:10.3389/fnhum.2013.00232

Cahn, J. E., & Brennan, S. E. (1999). A psychological model of grounding and repair in dialog. *Proceedings of the AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems* (pp. 25–33). North Falmouth, MA: American Association for Artificial Intelligence.

Carter-Thomas, S., & Rowley-Jolivet, E. (2001). Syntactic differences in oral and written scientific discourse: The role of information structure. *ASp—La revue du GERAS* (31–33), 19–37.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–56). New York, NY: Academic Press.

Chafe, W. (Ed.). (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Chafe, W. (2002). Searching for meaning in language: A memoir. *Historiographia Linguistica*, *29*(1), 245–261. doi:10.1075/hl.29.1.21cha

Chafe, W., & Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral and written language* (pp. 83–113). San Diego, CA: Academic Press.

Chase, P. (1995). *Given and new information* (Unpublished doctoral dissertation). Stony Brook University, Stony Brook, NY.

Chomsky, N. (1957). *Syntactic structures*. The Hague, Netherlands: Mouton.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Erlbaum.

Clark, E. V. (2014). Pragmatics in acquisition. *Journal of Child Language*, *41*(A1), 105–116. doi:10.1017/So3o5ooo91400017

Clark, H. H. (1977). Inferences in comprehension. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension* (pp. 243–263). Hillsdale, NJ: Erlbaum.

Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive Psychology*, *11*(4), 430–477. doi:10.1016/0010–0285(79)90020–3

Clark, H. H. (1992). *Arenas of language use*. Chicago, IL: University of Chicago Press.

Clark, H. H. (1996). *Using language*. Cambridge, United Kingdom: Cambridge University Press.

Clark, H. H. (2000). Everyone can write better (and you are no exception). In K. A. Keough & J. Garcia (Eds.), *Social psychology of gender, race, and ethnicity* (p. 379). New York, NY: McGraw-Hill.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association. doi:10.1037/10096–006

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given/new contract. In R. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Norwood, NJ: Ablex.

Clark H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62–81.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, United Kingdom: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*(2), 259–294. doi:10.1207/s15516709cog1302_7

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

Cohen, M. H., Giangola, J. P., & Balogh, J. (2004). *Voice user interface design*. New York, NY: Addison-Wesley.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314. doi:10.1016/s0749-596x(02)00001-3

Davis, J. R. (1989). *Back seat driver: Voice assisted automobile navigation* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering, H. (2008). Complementary systems for understanding action intentions. *Current Biology: CB*, *18*(6), 454–457. doi:10.1016/j.cub.2008.02.057

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener." In D. Napoli & J. Kegl (Eds.), *Bridges between psychology and linguistics* (pp. 105–129). San Diego, CA: Academic Press.

Don, A., Brennan, S., Laurel, B., & Shneiderman, B. (1992). Anthropomorphism: From Eliza to Terminator 2. In P. Bauersfeld, J. Bennett, & G. Lynch (Eds.), *Proceedings, CHI '92, human factors in computing systems* (pp. 67–70). New York, NY: ACM.

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLOS ONE*, *5*(8), e12166. doi:10.1371/journal.pone.0012166

Dumontheil, I., Küster, O., Apperly, I. A., & Blakemore, S. J. (2010). Taking perspective into account in a communicative task. *NeuroImage*, *52*(4), 1574–1583. doi:10.1016/j.neuroimage.2010.05.056

Fernando, T. (2012). Compositionality in discourse from a logical perspective. In W. Hinzen, E. Machery, & M. Werning (Eds.), *The Oxford handbook of compositionality*. Oxford, United Kingdom: Oxford University Press. doi:10.1093/oxfordhb/9780199541072.013.0013

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340. doi:10.1006/cogp.1999.0730

Ferstl, E. C. (2010). The neuroanatomy of discourse comprehension: Where are we now? In V. Bambini (Guest Ed.), *Neuropragmatics, Special Issue of Italian Journal of Linguistics*, *22*, 61–88.

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, *29*(5), 581–593. doi:10.1002/hbm.20422

Fonda, D., & Healy, R. (2005, September 8). How reliable is Brown's resume? *TIME*. http://content.time.com/time/nation/article/0,8599,1103003,00.html

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*(5), 489–504. doi: 10.1006/cogp.1999.0730

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

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communications. *Communications of the ACM*, *30*(11), 964–971. doi:10.1145/32206.32212

Galati, A., & Brennan, S. E. (2006). Given-new attenuation effects in spoken discourse: For the speaker, or for the addressee? In *Abstracts of the Psychonomic Society, 47th annual meeting* (p. 15), Austin, TX: Psychonomic Society Publications.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*(1), 35–51. doi:10.1016/j.jml.2009.09.002

Galati, A., & Brennan, S. E. (2014). Speakers adapt gestures to addressees' knowledge: Implications for models of co-speech gesture. *Language, Cognition, and Neuroscience*, *29*(4), 435–451. doi:10.1080/01690965.2013.796397

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218. doi:10.1016/0010-0277(87)90018-7

Gennari, S. P. (2004). Temporal reference and temporal relations in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 877–890. doi: 10.1037/0278-7393.30.4.877

Gernsbacher, M. A. (1996). Coherence cues mapping during comprehension. In J. Costermans & M. Fayol (Eds.), *Processing interclausal relationships in the production and comprehension of text* (pp. 3–21). Mahwah, NJ: Erlbaum.

Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. New Haven, CT: Yale University Press. doi: 10.2307/3684839

Gerrig, R. J., Brennan, S. E., & Ohaeri, J. O. (2001). What characters know: Projected knowledge and projected co-presence. *Journal of Memory and Language*, *44*(1), 81–95.

Ginzburg, J., & Cooper, R. (2004). Clarification, ellipsis, and the nature of contextual updates. *Linguistics and Philosophy*, *27*(3), 297–365. doi: 10.1023/b:ling.0000023369.19306.90

Glucksberg, S., & Weisberg, R. W. (1966). Verbal behavior and problem solving: Some effects of labeling in a functional fixedness problem. *Journal of Experimental Psychology*, *71*(5), 659–664. doi:10.1037/h0023118

Good, M. D., Whiteside, J. A., Wixon, D. R., & Jones, S. J. (1984). Building a user-derived interface. *Communications of the ACM*, *27*(10), 1032–1043. doi:10.1145/358274.358284

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.). *Everyday language: Studies in ethnomethodology* (pp. 97–121). New York, NY: Irvington.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York, NY: Academic Press.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive science*, *17*(3), 311–347. doi:10.1207/s15516709cog1703_1

Graesser, A. C., & Forsyth, C. M. (2013). Discourse comprehension. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 475–491). Oxford, United Kingdom: Oxford University Press. doi:10.1093/oxfordhb/9780195376746.013.0030

Graesser, A. C., Gernsbacher, M. A., & Goldman, S. R. (2003). *Handbook of discourse processes*. London, United Kingdom: Routledge.

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. 73–113). London, United Kingdom: MIT Press.

Greene, S. B., Gerrig, R. J., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language*, *33*(4), 511–526. doi: 10.1006/jmla.1994.1024

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics III: Speech acts* (pp. 41–58). New York, NY: Academic Press,

Grosz, B. J. (1977). *The representation and use of focus in a system for dialogue understanding* (Doctoral thesis). University of California–Berkeley, Berkeley, California. *Tech. Note 151, Artificial Intelligence Center*. Menlo Park, CA: SRI International.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational linguistics*, *12*(3), 175–204.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, *21*(2), 203–225.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274–307. doi: 10.2307/416535

Hagoort, P., & Levinson, S. C. (2014). Neuropragmatics. In M. S. Gazzaniga & G. R. Mangun (Eds.), *The cognitive neurosciences* (5th ed., pp. 667–674). Cambridge, MA: MIT Press.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London, United Kingdom: Longman. doi:10.4324/9781315836010

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615. doi:10.1016/j.jml.2007.01.008

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*(1), 105–115. doi:10.1207/s15516709cog2801_5

Hari, R., Henriksson, L., Malinen, S., & Parkkonen, L. (2015). Centrality of social interaction in human brain function. *Neuron*, *88*(1), 181–193. doi:10.1016/j.neuron.2015.09.022

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, *16*(2), 114–121. doi:10.1016/j.tics.2011.12.007

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, *13*(5), 512–521.

Haynes, J. D. (2015). A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives. *Neuron*, *87*(2), 257–270. doi:10.1016/j.neuron.2015.05.025

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), 523–534. doi:10.1038/nrn1931

Hobbs, J. R., Stickel, M., Martin, P., & Edwards, D. (1988). Interpretation as abduction. In *Proceedings of the 26th annual meeting of Association for Computational Linguistics* (pp. 95–103). Stroudsburg, PA: Association for Computation Linguistics.

Holler, J., Kokal, I., Toni, I., Hagoort, P., Kelly, S. D., & Ozyurek, A. (2015). Eye'm talking to you: Speakers' gaze direction modulates co-speech gesture processing in the right MTG. *Social Cognitive & Affective Neuroscience*, *10*(2), 255–261. doi:10.1093/scan/nsu047.

Horton, W. S., & Brennan, S. E. (2016). The role of metarepresentation in the production and resolution of referring expressions. *Frontiers in Psychology*, *7*, 1–12. doi:10.3389/fpsyg.2016.01111

Horton, W. S., & Gerrig, R. J. (2005a). Conversational common ground and memory processes in language production. *Discourse Processes*, *40*, 1–35.

Horton, W. S., & Gerrig, R. J. (2005b). The impact of memory demands on audience design during language production. *Cognition*, *96*, 127–142.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*, 91–117.

Hough, J., & Schlangen, D. (2016). Investigating fluidity for human-robot interaction with real-time, real-world grounding strategies. In *Proceedings of the SIGDIAL 2016 conference* (pp. 288–298). Los Angeles, CA: Association for Computational Linguistics.

Hudson, S., Tanenhaus, M., & Dell, G. (1986, August). *The effect of discourse center on the local coherence of a discourse*. Paper presented at Eighth Annual Conference of the Cognitive Science Society, Amherst, MA.

Hudson-D'Zmura, S. B. (1988). *The structure of discourse and anaphor resolution: The discourse center and the roles of nouns and pronouns* (Unpublished doctoral dissertation). University of Rochester, Rochester, NY.

Hwang, J., Brennan, S. E., & Huffman, M. K. (2015). Phonetic adaptation in non-native spoken dialogue: Effects of priming and audience design. *Journal of Memory and Language*, *81*, 72–90. doi:10.1016/j.jml.2015.01.001

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*(3), e79. doi:10.1371/journal.pbio.0030079

Jefferson, G. (1973). A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica*, *9*(1), 47–96. doi:10.1515/semi.1973.9.1.47

Kehler, A., & Ward, G. (2006). Referring expressions and conversational implicature. In B. J. Birner & G. Ward (Eds.), *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn* (pp. 177–193). Philadelphia, PA: John Benjamins. doi:10.1075/slcs.80.11keh

Keysar, B. (1997). Unconfounding common ground. *Discourse Processes*, *24*, 253–270.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, *4*, 46–50.

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: From self to social cognition. *Trends in Cognitive Sciences*, *11*(5), 194–196. doi:10.1016/j.tics.2007.02.002

Kintsch, W. (1988). The use of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182. doi: 10.1037/0033–295x.95.2.163

Konvalinka, I., Bauer, M., Stahlhut, C., Hansen, L. K., Roepstorff, A., & Frith, C. D. (2014). Frontal alpha oscillations distinguish leaders from followers: Multivariate decoding of mutually interacting brains. *NeuroImage*, *94*, 79–88. doi:10.1016/j.neuroimage.2014.03.003

Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: How can mutually interacting brains teach us something about social interaction? *Frontiers in Human Neuroscience*, *6*, 215. doi:10.3389/fnhum.2012.00215

Kopp, S., Gesellensetter, L., Krämer, N., & Wachsmuth, I. (2005). A conversational agent as museum guide–design and evaluation of a real-world application. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, & T. Rist (Eds.), *Intelligent Virtual Agents, LNAI 3661* (pp. 329–343). Heidelberg, Germany, Springer-Verlag.

Kraljic, T., & Brennan, S. E. (2005). Using prosody and optional words to disambiguate utterances: For the speaker or for the addressee? *Cognitive Psychology*, *50*, 194–231. doi:10.1016/j.cogpsych.2004.08.00

Krauss, R. M. (1987). The role of the listener: Addressee influences on message formulation. *Journal of Language and Social Psychology*, *6*(2), 81–98.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*, 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*, 359–363.

Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language*, *83*, 1–19. doi:10.1016/j.jml.2015.03.008

Kuhlen, A. K., Allefeld, C., Anders, S., & Haynes, J. D. (2015). Towards a multi-brain perspective on communication in dialogue. In R. Willems (Ed.), *Cognitive neuroscience of natural language use* (pp. 182–200). Cambridge, United Kingdom: Cambridge University Press. doi:10.1017/cbo9781107323667.009

Kuhlen, A. K., Allefeld, C., & Haynes, J.-D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in Human Neuroscience*, *6*, 266. doi:10.3389/fnhum.2012.00266

Kuhlen, A. K., Bogler, C., Brennan, S. E., & Haynes, J.-D. (2017). Brains in dialogue: Decoding neural preparation of speaking to a conversational partner. *Social, Cognitive, and Affective Neuroscience*, *12*(6), 871–880. doi:10.1093/scan/nsx018.

Kuhlen, A. K., Bogler, C., Swerts, M., & Haynes, J.-D. (2015). Neural coding of assessing another person's knowledge based on nonverbal cues. *Social Cognitive and Affective Neuroscience*, *10*(5), 729–734. doi:10.1093/scan/nsu111

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, *20*(1), 54–72.

Larsson, S., & Traum, D. R. (2000). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering 6 (*3–4*)*, 323–334.

Lee, S., & Stent, A. (2016). Task lineages: Dialog state tracking for flexible interaction. In *Proceedings of the SIGDIAL 2016 conference* (pp. 11–21). Los Angeles, CA: Association for Computational Linguistics.

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, *14*(1), 78–106. doi:10.1016/0010-0285(82)90005-6

Levinson, S. C. (1983). *Pragmatics*. Cambridge, United Kingdom: Cambridge University Press.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology: Language Sciences*, *6*, 731. doi:10.3389/fpsyg.2015.00731

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*(1), 339–359. doi:10.1007/bf00258436

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, *58*(1), 259–289. doi:10.1146/annurev.psych.58.110405.0856

Linde, C. (1983). A framework for formal models of discourse: What can we modely and why. *Text & Talk: An Interdisciplinary Journal of Language, Discourse & Communication Studies*, *3*, 217–276.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, *9*(3), 550–557. doi:10.3758/bf03196312

MacKay, D. M. (1983). The wider scope of information theory. In F. Machlup (Ed.), *The study of information: Interdisciplinary messages* (pp. 485–492). Hoboken, NJ: Wiley.

MacWhinney, B. (1977). Starting points. *Language*, *53*(1), 152–168. doi:10.2307/413059

Malt, B. C. (1985). The role of discourse structure in understanding anaphora. *Journal of Memory and Language*, *24*, 271–289.

Mar, R. A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia*, *42*(10), 1414–1434. doi:10.1016/j.neuropsychologia.2003.12.016

Marge, M., & Rudnicky, A. I. (2015). Miscommunication recovery in physically situated dialogue. In *Proceedings of the SIGDIAL 2015 conference* (p. 22). Prague, Czech Republic: Association for Computational Linguistics.

Mason, R., A., & Just, M. A. (2006). Neuroimaging contributions to the understanding of discourse processes. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 765–799). Amsterdam, Netherlands: Elsevier.

Matsuyama, Y., Bhardwaj, A., Zhao, R., Romero, O. J., Akoju, S. A., & Cassell, J. (2016). Socially-aware animated intelligent personal assistant agent. In *Proceedings of the SIGDIAL 2016 conference* (pp. 224–227). Los Angeles, CA: Association for Computational Linguistics.

Matthews, D., Lieven, E., & Tomasello, M. (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Developmental Psychology*, *46*(4), 749–760. doi:10.1037/a0019657.

McKoon, G., & Gerrig, R. J. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes*, *26*, 67–86.

McKoon, G., Gerrig, R. J., & Greene, S. B. (1996). Pronoun resolution without pronouns: Some consequences of memory-based text processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 919–932.

Meehan, J. R. (1976). *The metanovel: Writing stories by computer* (Unpublished doctoral dissertation). Yale University, New Haven, CT.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213. doi:10.1016/s0749-596x(03)00028-7

Miller, G. A. (1963). *Language and communication*. New York, NY: McGraw-Hill.

Mizukami, M., Yoshino, K., Neubig, G., Traum, D., & Nakamura, S. (2016). Analyzing the effect of entrainment on dialogue acts. In *Proceedings of the SIGDIAL 2016 conference* (pp. 310–318). Los Angeles, CA: Association for Computational Linguistics.

Momennejad, I., & Haynes, J.-D. (2012). Human anterior prefrontal cortex encodes the "what" and "when" of future intentions. *NeuroImage*, *61*(1), 139–148. doi:10.1016/j.neuroimage.2012.02.079

Morency, L. P., de Kok, I., & Gratch, J. (2010). A probabilistic multimodal approach for predicting listener backchannels. *Autonomous Agents and Multi-Agent Systems*, *20*(1), 70–84.

Morrow, D. G., & Clark, H. H. (1988). Interpreting words in spatial descriptions. *Language and Cognitive Processes*, *3*(4), 275–291.

Morrow, D., Clark, H. H., Lee, A., & Rodvold, M. (1990). Collaboration in controller-pilot communication. *Abstracts of the Psychonomic Society, 31st annual meeting* (p. 494) New Orleans, LA: Psychonomic Society Publication.

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language*, *26*(2), 165–187.

Nieuwland, M. S., & van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*, 1098–1111.

Noordzij, M. L., Newman-Norlund, S. E., de Ruiter, J. P., Hagoort, P., Levinson, S. C., & Toni, I. (2009). Brain mechanisms underlying human communication. *Frontiers in Human Neuroscience*, *3*, 14. doi:10.3389/neuro.09.014.2009

Norberg, A. L. (1991). *Oral history interview with Terry Allen Winograd/Interviewer Arthur L. Norberg*. Charles Babbage Institute, Center for the History of Information Processing University of Minnesota, Minneapolis. Retrieved from http://conservancy.umn.edu/handle/11299/107717

Olsson, A., & Ochsner, K. N. (2008). The role of social cognition in emotion. *Trends in Cognitive Sciences*, *12*(2), 65–71. doi:10.1016/j.tics.2007.11.010

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02), 169–190. doi:10.1017/S0140525X04000056

Pinker, S. (1994). *The language instinct*. New York, NY: HarperCollins.

Polanyi, L., & Scha, R. (1984, July). A syntactic approach to discourse semantics. In *Proceedings of the 10th international conference on computational linguistics* (pp. 413–419). Los Angeles, CA: Association for Computational Linguistics.

Polichak, J. W., & Gerrig, R. J. (1998). Common ground and everyday language use: Comments on Horton and Keysar (1996). *Cognition*, *66*, 183–189.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–56). New York, NY: Academic Press.

Prinz W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action: Current approaches* (pp. 167–201). Berlin/Heidelberg, Germany: Springer. doi:10.1007/978-3-642-75348-0_7

Rahimtoroghi, E., Hernandez, E., & Walker, M. A. (2016). Learning fine-grained knowledge about contingent relations between everyday events. In *Proceedings of the SIGDIAL 2016 conference* (pp. 350–359). Los Angeles, CA: Association for Computational Linguistics.

Ramsey, R., Hansen, P., Apperly, I., & Samson, D. (2013). Seeing it my way or your way: Frontoparietal brain areas sustain viewpoint-independent perspective selection processes. *Journal of Cognitive Neuroscience*, *25*(5), 670–684. doi:10.1162/jocn_a_00345

Reddy, M. J. (1979). The conduit metaphor: A case of frame conflict in our language about language. In A. Ortony (Ed.), *Metaphor and thought* (pp. 284–310). Cambridge, United Kingdom: Cambridge University Press.

Rice, K., & Redcay, E. (2015). Interaction matters: A perceived social partner alters the neural processing of human speech. *NeuroImage*, *129*, 480–488. doi:10.1016/j.neuroimage.2015.11.041

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, *29*(6), 1045–1060. doi:10.1207/s15516709cog0000_29

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192. doi:10.1146/annurev.neuro.27.070203.144230

Rosenfeld, H. M. (1987). Conversational control functions of nonverbal behavior. In A. W. Siegman & S. Feldstein (Eds.), *Nonverbal behavior and communication* (pp. 563–601). Hillsdale, NJ: Erlbaum.

Rueschemeyer, S.-A., Gardner, T., & Stoner, C. (2015). The social N400 effect: How the presence of other listeners affects language comprehension. *Psychonomic Bulletin & Review*, *22*(1), 128–134. doi:10.3758/s13423-014-0654-x

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 211–236). New York, NY: Academic Press.

Rumelhart, D. E. (1979). Some problems with the notion of literal meaning. In A. Ortony (Ed.), *Metaphor and thought* (pp. 71–82). Cambridge, United Kingdom: Cambridge University Press. doi:10.1017/cbo9781139173865.007

Sachs, J. D. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, *2*(9), 437–442.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735. doi:10.1353/lan.1974.0010

Samuel, A. G., & Troicki, M. (1998). Articulation quality is inversely related to redundancy when children or adults have verbal control. *Journal of Memory and Language*, *39*(2), 175–194.

Sanders, T., & Pander Maat, H. (2006). Cohesion and coherence: Linguistic approaches. In K. Brown (Ed.-in-Chief) & A. H. Anderson, L. Bauer, M. Berns, G. Hirst, & J. Miller (Coordinating Eds.), *Encyclopedia of language and linguistics* (2nd ed., pp. 591–595). London, United Kingdom: Elsevier. doi:10.1016/b0-08-044854-2/00497-1

Sassa, Y., Sugiura, M., Jeong, H., Horie, K., Sato, S., & Kawashima, R. (2007). Cortical mechanism of communicative speech production. *NeuroImage*, *37*(3), 985–992. doi:10.1016/j.neuroimage.2007.05.059

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. *Studies in Social Interaction*, *75*, 75–119.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica*, *8*(4), 289–327.

Schilbach, L. (2015). Eye to eye, face to face and brain to brain: Novel approaches to study the behavioral dynamics and neural mechanisms of social interactions. *Current Opinion in Behavioral Sciences*, *3*, 130–135. doi:10.1016/j.cobeha.2015.03.006

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *The Behavioral and Brain Sciences*, *36*(4), 393–414. doi:10.1017/S0140525X12000660

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., & Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proceedings of the National Academy of Sciences, USA*, *107*(20), 9388–9393. doi:10.1073/pnas.1001791107

Schmandt, C., & Hulteen, E. A. (1982). The intelligent voice-interactive interface. In *Proceedings of the 1982 Conference on Human Factors in Computing Systems* (pp. 363–366). New York, NY: ACM.

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123–164). Hillsdale, NJ: Erlbaum.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232. doi:10.1016/0010-0285(89)90008-x

Searle, J. (1969). *Speech acts: An essay in the philosophy of language*, Cambridge, United Kingdom: Cambridge University Press.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Champaign: University of Illinois Press.

Shneiderman, B. (1987). *Designing the user interface: Strategies for effective human-computer interaction*. Addison-Wesley, Reading, MA.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*(1), 25–38. doi10.1006/jmla.1993.1002

Sperber, D., & Wilson, D., (1986). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J.-B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, *63*(1), 25–39. doi:10.1016/j.neuroimage.2012.06.046

Spunt, R. P., & Lieberman, M. D. (2012). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *NeuroImage*, *59*(3), 3050–3059. doi:10.1016/j.neuroimage.2011.10.005

Spunt, R. P., Satpute, A. B., & Lieberman, M. D. (2011). Identifying the what, why, and how of an observed action: An fMRI study of mentalizing and mechanizing during action observation. *Journal of Cognitive Neuroscience*, *23*(1), 63–74. doi:10.1162/jocn.2010.21446

Stellmann, P., & Brennan, S. E. (1993). Flexible perspective-setting in conversation. In *Abstracts of the Psychonomic Society, 34th annual meeting* (p. 20). Washington, DC: Psychonomic Society.

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences, USA*, *107*(32), 14425–14430. doi:10.1073/pnas.1008662107

Stolk, A., D'Imperio, D., di Pellegrino, G., & Toni, I. (2015). Altered communicative decisions following ventromedial prefrontal lesions. *Current Biology: CB*, *25*(11), 1469–1474. doi:10.1016/j.cub.2015.03.057

Stolk, A., Noordzij, M. L., Verhagen, L., Volman, I., Schoffelen, J.-M., Oostenveld, R., . . . Toni, I. (2014). Cerebral coherence between communicators marks the emergence of meaning. *Proceedings of the National Academy of Sciences, USA*, *111*(51), 18183–18188. doi:10.1073/pnas.1414886111

Stolk, A., Verhagen, L, & Toni, I. (2016). Conceptual alignment: How brains achieve mutual understanding. *Trends in Cognitive Sciences*, *20*(3), 180–191. doi: 10.1016/j.tics.2015.11.007

Svartvik, J., & Quirk, R. (1980). *A corpus of English conversation*. Lund Studies in English 56. Lund, Sweden: Gleerup.

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*(1), 81–94. doi:10.1016/j.jml.2005.02.003

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Thioux, M., Gazzola, V., & Keysers, C. (2008). Action understanding: How, what and why. *Current Biology*, *18*(10), R431–R434. doi:10.1016/j.cub.2008.03.018

Traum, D. R. (1994). *A computational theory of grounding in natural language conversation* (Unpublished doctoral dissertation). University of Rochester, Rochester, NY.

Traxler, M. J., & Gernsbacher, M. A. (1992). Improving written communication through minimal feedback. *Language and Cognitive Processes*, *7(*1), 1–22. doi:10.1080/01690969208409378

Traxler, M. J., & Gernsbacher, M. A. (1993). Improving written communication through perspective-taking. *Language and Cognitive Processes*, *8*(3), 311–334. doi:10.1080/01690969308406958

Traxler, M. J., & Gernsbacher, M. A. (1995). Improving coherence in written communication. In M. A. Gernsbacher & T. Givón (Eds.), *Coherence in spontaneous text* (pp. 215–238). Philadelphia, PA: John Benjamins.

Trueswell, J. C., & Tanenhaus, M. K. (2005). *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.

Urbina, I. (2004, November 24). Your train will be late, she says cheerily. *New York Times*. http://www.nytimes.com/2004/11/24/nyregion/your-train-will-be-late-she-says-cheerily.html

Van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rueschemeyer, S.-A. (2012). Pragmatics in action: Indirect requests engage theory of mind areas and the cortical motor network. *Journal of Cognitive Neuroscience*, *24*(11), 2237–2247. doi:10.1162/jocn_a_00274

Van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Erlbaum.

Van Der Wege, M. M. (2009). Lexical entrainment and lexical differentiation in reference phrase choice. *Journal of Memory and Language*, *60*(4), 448–463. doi:10.1016/j.jml.2008.12.003

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*(3), 564–584. doi:10.1016/j.neuroimage.2009.06.009

Wachsmuth, I. (2008). 'I, Max'—communicating with an artificial agent. In I. Wachsmuth & G. Knoblich (Eds.), *Modeling communication with robots and virtual humans* (pp. 279–295). Berlin, Germany: Springer.

Walker, M. A., Joshi, A. K., & Prince, E. F. (1998). Centering in naturally-occurring discourse: An overview. In M. A. Walker, A. K. Joshi, & E. F. Prince (Eds.), *Centering theory in discourse* (pp. 1–28). Oxford, United Kingdom: Clarendon Press.

Waytz, A., & Mitchell, J. P. (2011). Two mechanisms for simulating other minds dissociations between mirroring and self-projection. *Current Directions in Psychological Science*, *20*(3), 197–200. doi:1010.1177/0963721411409007

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the Association for Computing Machinery*, *9*, 36–45.

Weizenbaum, J. (1976). *Computer power and human reason*. San Francisco, CA: W.H. Freeman.

Welborn, B. L., & Lieberman, M. D. (2015). Person-specific theory of mind in medial pFC. *Journal of Cognitive Neuroscience*, *27*(1), 1–12. doi:10.1162/jocn_a_00700

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*, 183–194.

Willems, R. M. (Ed.). (2015). *Cognitive neuroscience of natural language use*. Cambridge, United Kingdom: Cambridge University Press.

Willems, R. M., de Boer, M., de Ruiter, J. P., Noordzij, M. L., Hagoort, P., & Toni, I. (2010). A dissociation between linguistic and communicative abilities in the human brain. *Psychological Science*, *21*(1), 8–14. doi:10.1177/0956797609355563

Williams, J., Raux, A., & Henderson, M. (2016). The dialog state tracking challenge series: A review. *Dialogue and Discourse*, *7*(3), 4–33.

Winograd, T. (1971). *Procedures as a representation for data in a computer program for understanding natural language* (Report No. 235). Cambridge, MA: MIT Artificial Intelligence Laboratory.

Winograd, T. (1983). *Language as a cognitive process*. Reading, MA: Addison-Wesley.

Yngve, V. (1970). On getting a word in edgewise. In *Papers from the sixth regional meeting, Chicago Linguistics Circle*. Chicago, IL: Chicago Linguistics Circle.

Yu, Y., Eshghi, A., & Lemon, O. (2016). Training an adaptive dialogue policy for interactive learning of visually grounded word meanings. In *Proceedings of the SIGDIAL 2016 conference* (pp. 339–349). Los Angeles, CA: Association for Computational Linguistics.

Yu, Z., Nicolich-Henkin, L., Black, A. W., & Rudnicky, A. (2016). A Wizard-of-Oz study on a non-task-oriented dialog systems that reacts to user engagement. In *Proceedings of the SIGDIAL 2016 conference* (pp. 55–63). Los Angeles, CA: Association for Computational Linguistics.

Zaki, J., Hennigan, K., Weber, J., & Ochsner, K. N. (2010). Social cognitive conflict resolution: Contributions of domain-general and domain-specific neural systems. *The Journal of Neuroscience*, *30*(25), 8481–8488. doi:10.1523/JNEUROSCI.0382-10.2010

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 386–397. doi:10.1037/0278-7393.21.2.386

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185. doi:10.1037/0033-2909.123.2.162