Selected Published Research on Modeling Face-to-face Conversation

Justine Cassell, Matthew Stone and David Traum
ESSLLI 2005

The following list contains a survey of some important and recent research in modeling face-to-face conversation. The list below is a presented as a guide to the literature by topic and date; we include complete citations afterwards in alphabetical order. For brevity, research works are keyed by *first author* and date only (we use these keys on the slides as well as in this list). Of course, most papers are multiply authored.

The list is not intended to be exhaustive. Our primary aim is simply to provide bibliographic information for all the research that we will refer to during the ESSLLI class itself. The entries also provide a sampling from ongoing research projects so that you can get an overall sense of the state of the field and begin to follow up topics of particular interest to you.

Useful Frameworks

Theoretical principles

- [Grosz 1986] A programmatic discussion showing how goal-directed analyses of dialogue can be realized computationally in conversational agents. Introduces coordinated representations of the linguistic structure of discourse, the attentional state of the discourse, and the hierarchical relationships among communicative intentions across discourse.

- [Clark 1996] A book-length survey of pragmatics in philosophy and psychology, using collaboration as unifying principle. Gives a systematic, intuitive and detailed account of how people naturally use language in spontaneous face-to-face communication and other settings for language use.

- [Cassell 2000f] An edited collection of chapters describing a range of research projects in embodied conversational agents. Includes [Cassell 2000c], an overview of face-to-face conversation for conversational agents, and [Cassell 2000d], a description of design and architectural principles for realizing the behaviors and functions of face-to-face conversation. Other chapters report on specific approaches to the design, implementation, applications and evaluation of embodied conversational agents.

- [Larsson 2000] The standard published reference for the information-state approach to dialogue management, as implemented in the TRINDI project. Briefly surveys a down-to-earth knowledge-representation methodology for formalizing context and context change. A longer technical report about the TRINDI framework is also available on the web [TRINDI Consortium 2001].

- [Stone 2004a] and [Stone 2004b] An investigation of intentional agency as a framework for implementing language processing modules in conversational agents. Sketches how linguistic representations of utterance interpretation can be understood as encoding communicative intentions and how modules for utterance understanding, utterance generation and dialogue management can construct and exchange these representations.

Methodology

- [Power 1977] A clear and still useful illustration of computer simulation as a methodology for dialogue research. Power designs, implements and analyzes two computational agents that talk to each other and get stuff done. Provides a strong test of the dialogue model these agents use, because it allows us to rigorously compare the predictions of the theory, as illustrated by the agents' dialogues, with the organization of natural human–human dialogue.

- [Brachman 1990] A hands-on description of the practical methodology for knowledge representation—developing a formal representation of an arbitrary domain for reasoning in a computational model. Portrays knowledge representation as an appealing mix of the techniques of logic, software engineering, and plain common sense. To the extent that realizing computational models of face-to-face conversation involves substantial formalization and implementation, the lessons described in this paper are not to be ignored!

- [Dahlbäck 1993] A clear statement of the rationale for carrying out Wizard-of-Oz studies—analyses of users interacting with interactive computer systems in situations where, unbeknownst to users, some of the apparent system functionality is actually provided by a person acting behind the scenes. Also provides principles for designing such studies and carrying them out successfully.

- [Carletta 1997] Explores the methodological issues in annotating conversational data with theoretically-motivated deep descriptions. In this case, the issue is the underlying structure of a dialogue, and a key issue is reliability [Carletta 1996]—making sure that different judges describe the same dialogues in a consistent and meaningful way.

- [Walker 1997] Explores the methodological issues in evaluating dialogue systems. Particularly concerned with finding statistical relationships among a range of dialogue performance measures, including task completion and user satisfaction, so that a system will be able to assess its own performance automatically. The methodology is applied and illustrated in [Walker 1998].

Behaviors in Conversation

General

- [Ekman 1969] A classic paper, introducing a number of useful distinctions for thinking about human behavior in conversation. Particularly clear on the distinctions between intentionally communicative actions that accompany speech and other kinds of behavior, including signals of emotional state.

- [Duncan 1977] An early, exhaustive statistical analysis of patterns of verbal and nonverbal behavior in face-to-face dialogue. Documents how the events in conversation are an emergent property of the interaction between two people, and

thereby sets the stage for a collaborative view of conversational interaction.

- [Goodwin 1981] Studies of conversation that clearly show how behaviors in conversation across modalities work together, both for the speaker and for the hearer. For example, in looking at gaze in turn-taking, Goodwin argues for considering not just the gaze of the speaker but also the gaze of the hearer and the coordination of the gaze of conversational participants.

- [Kendon 1990] Adopts a social science perspective to offer a wide-ranging look at specific dialogue phenomena analyzed in detail. A good source of data to motivate a range of actions for models of face-to-face conversation.

Gaze

- [Argyle 1976] A descriptive survey of the role of gaze in conversation. Motivates a range of functions for gaze with a particular focus on mutual gaze—where both parties in a dyad look into one another's eyes. This seems to help interlocutors coordinate their efforts to seek evidence for mutual understanding, to show interest, and to manage the floor.

- [Novick 1996] Combines an empirical study with a computational model to characterize the role of gaze in turn-taking in dyadic conversation. Emphasizes two patterns: first, a *mutual-break* pattern in which interlocutors establish mutual gaze at the end of one utterance before the next speaker looks away and begins to speech; and second, a *mutual-hold* pattern (perhaps associated with difficulty) where the new speaker maintains gaze while speaking.

- [Lee 2002] Describes a data-driven statistical model designed to portray believable eye-movements for virtual characters; the model is conditioned on conversational state as well as other factors, and thereby reproduces the results of observational studies of conversational signals.

Facial expressions

- [Ekman 1975] An accessible book for improving one's skills at reading the face, with particular emphasis on the emotions. Photographic collages highlight the different possible movements of the face and their involvement in paradigmatic emotional expressions.

- [Ekman 1978] The actual manual for Ekman and colleagues' Facial Action Coding System—the authoritative source for learning to recognize (produce!) and annotate the movements that comprise the expressive repertoire of the human face.

- [Ekman 1979] Highlights the use of eyebrows in conversational signals as well as emotional expressions, and observes the use of eyebrows as underliners, in sync with important phrases, and batons, accompanying important words.

- [Faigin 1990] Another accessible book about the emotional face, this one aimed at artists. A scientifically-informed but aesthetically-motivated guide for looking at the face, drawing the face, and appreciating expressive portraiture.

- [Chovil 1991a] Analyzes facial behaviors in conversation, documents a wide range of displays besides felt emotions, and speculates about the diverse descriptive and interactive functions these displays could serve. Many of these observations are easy to implement; see [Nagao 1994] under Systems–Embodied Conversational Agents.

- [Chovil 1991b], [Buck 1991], and [Chovil 1991c]. An instructive debate on whether interaction (Chovil, Fridlund) or emotion (Buck) best explains most facial displays in conversation. Interesting in part because it reveals deep unknowns in the theory of emotion and interaction, and key methodological difficulties in resolving these unknowns.

- [Pelachaud 1996] Uses a model of information structure and conversational state to animate head nods, eyebrow movements and other facial displays to accompany automatically-generated speech for a virtual character.

- [Cave 1996] An instrumental study on English speech linking emphasis in speech (measured by pitch) with accompanying eyebrow movements.

- [Poggi 2000] A survey from a computer graphics perspective of the expressive resources required for the eyes and eyebrows in embodied conversational agents. Argues that computer graphics design and architecture must address a range of functions the upper face serves, from coordinating the conversation to identifying points in space or even to disambiguating the kind of move being performed in dialogue.

- [Pelachaud 2002] A state-of-the-art illustration of the potential complexity of controlling an animated face from a process model of communication and emotion. Because the face can do so much, it becomes necessary to balance potential conflicts among actions and to infer a consistent pattern of actions for the face to animate.

- [Krahmer 2002] Introduces an ongoing research program on the interpretation of conversational signals: the experimental results here show that viewers use raised eyebrows to help reconstruct a speaker's point of emphasis on utterances whose prosodic focus is otherwise ambiguous.

Gesture

- [McNeill 1992] A book-length exposition of McNeill's empirical and theoretical results on gesture in conversation, including programmatic discussions of the forms of gestures, the meanings of gestures, and the role of gestures in cognition and communication.

- [Bavelas 1995] An empirical study of deictic and metaphorical gestures directed at one's interlocutor, emphasizing the frequency of these gestures and their diverse roles in managing conversational interaction.

- [Cassell 2000e] Describes a computational model of the coordinated generation of speech and accompanying iconic gesture as an integrated process. The generator takes multiple communicative goals as input and plans a complex, composite communicative action. The process is able to explain naturalistic choices of gesture in utterances by representing the gestures' discourse function, content and synchrony with speech.

- [Kopp 2004a] Presents an empirically-based computational model of gesture morphology, which links features of gesture such as hand shape and trajectory with abstract, qualitative semantic constraints. Decomposing a gesture into these morphological features makes it possible to to generate novel iconic gestures without drawing on a predefined gestural lexicon. Utterances are generated in an integrated microplanning process, much as in [Cassell 2000e].

Posture

- [Condon 1971] and [Kendon 1972]. These two descriptive studies argue, from qualitative analysis of individual dialogues, that shifts in posture during a conversation index important shifts in the content or direction of discourse.

- [Cassell 2001b] Correlates posture shifts in human monologues and dialogues with newer theories of discourse structure and conversation structure. Provides empirical evidence that posture shifts may come at the beginning of new high-level discourse segments or accompany interactive efforts to actively take a turn. Shows that these patterns of behavior can be directly realized in a computational agent with a model of discourse and floor management.

Intonation

- [Pierrehumbert 1990] Surveys the varied tunes possible in English intonation, and attempts to factor the meanings of these tunes compositionally into separate contributions carried by pitch targets on accented syllables and tones on phrases and boundaries. A great introduction to English intonation and the problems of analyzing it. Published with instructive comments by Jerry Hobbs [Hobbs 1990].

- [Hirschberg 1993] Uses the theory of intonation to design a system for speech synthesis that automatically determines which words in a text should be accented when the text is read. The general idea is to predict communicative functions or behaviors that are missing from a transcript using reliable surface cues. This idea is very effective for intonation and a useful starting point for other aspects of communication; see [Cassell 2001a].

- [Hirschberg 1996] Analyzes the way intonation covaries with discourse structure. A specific contribution is its demonstration that speakers systematically vary pitch range—the difference between highest peak and deepest low across a whole phrase—as a function of the placement of each phrase within the overall structure of discourse.

- [Steedman 2000] A detailed linguistic analysis of the relationship between syntactic structure, prosodic structure, and information structure in utterances. Notable for its proposal that intonation is precisely matched with syntactic units in utterances and transparently characterizes the status of the accompanying information in the ongoing discourse—an elegant analysis and a model for other communicative modalities that can accompany speech.

Markup

- [Silverman 1992] The published reference for the Tones and Break Indices (ToBI) system, commonly used as a standard for describing the qualitative organization and tune of English intonation in synchrony with simultaneous speech.

- [Chi 2000] [Badler 2002] and [Byun 2002] Describes the EMOTE model for specifying the quality of movement in animation. Where markup for conversational animation typically focuses on *what* the character is to do, it's also crucial to address *how* the character is to do it. The EMOTE model is inspired by Laban's analysis of movement in terms of shape, "the changing form the body makes in space", and effort, "how the body concentrates its exertion when performing movements".

- [Cassell 2001a] Describes the BEAT system for automatically generating an animated conversational delivery of input text. The architecture is based on a cascade of processes that mark up text for the likely communicative functions planned by its author and for communicative behaviors that could have complemented a spoken delivery and helped convey these functions. The system thus has access to a rich markup including not only behaviors such as intonation, gesture and facial expression but also descriptions of information structure units, discourse structure relationships, and the status of entities and properties in the conversation.

- [Beskow 2002] Present a high-level formalism for specifying verbal and nonverbal output from a multimodal dialogue system. The formalism is designed for retargeting utterances for different platforms, characters, and contexts, so the output characterizes the communicative functions of the output without specifying the actual behaviors that realize them. Realization decisions in animating the utterance are made by individual character models.

- [Piwek 2002] Describes representation of communicative functions and behaviors in the NECA system. Their proposed RRL (rich representation language) is not just a way of specifying conversational action, but actually an eclectic and flexible formalism designed to help structure the interfaces among all phases of generation of embodied conversation.

- [DeCarlo 2004] Proposes a tiered representation (linked with the Scheme format used by the Festival speech synthesizer) for describing conversational movements of the head and face in synchrony with simultaneous speech. The representation is not general but builds in constraints about the relationships between gesture and speech. Also describes an implemented (and freely available) system for synthesizing animations based on the markup. A preliminary version of this work appears as [DeCarlo 2002].

- [Kopp 2004b] Proposes a tiered representation in XML format for describing the form of coverbal gesture and its synchrony with simultaneous speech. Also describes a powerful implemented system for synthesizing corresponding animations. A preliminary version of this work appears as [Kopp 2002].

Functions in Conversation

General

- [Poesio 1997] This paper presents the theory behind the information-state approach to dialogue management. It provides a general way of thinking about context and formalizing the different context-management functions of natural conversation. The overall model reflects insights from work on reference resolution, intention recognition, and dialogue management, and is able to describe speech acts, conversational moves, turn-taking and grounding.

- [Bavelas 2000] An integrative survey of research on the psychology of communication that argues for an *integrated message model* of contributions to dialogue. This model views utterances as coordinated ensembles of gesture, speech and other behaviors that speakers produce and interlocutors interpret as presenting one consistent description of objects and events and achieving a coherent set of communicative goals.

- [Engle 2000] An analysis of human speakers' spontaneous use of gesture, diagrams and demonstration to accompany

spoken explanations (of the workings of a mechanical lock). Illustrates how the coordination of timing between words and other communicative actions helps bind them into multimodal ensembles, and shows how information from multiple modalities works together to help the speaker precisely signal their one overall intended interpretation.

- [Traum 2002] An overview of an information-state approach to multi-party conversations between human users and characters in immersive virtual worlds. Illustrates the use of the information-state approach to describe face-to-face dialogue, including issues of embodiment, attention, and the perceptual context available to interlocutors in a shared environment. Describes conversation through a series of layers describing different subsystems of communication; each subsystem is described in terms of its state and the actions that change the state. The layers include established ones such as turn-taking and grounding, as well as several novel layers to which describe multi-party conversations as wholes as well as overlapping interactions of pairs of participants.

- [Stent 2002] Presents an architecture for generating contributions to spoken dialogue. Interfaces with approaches to dialogue management where utterances must achieve a diverse array of conversation acts or interactional functions. Specifically considers how complex utterances can be constructed to explicitly address a diverse range of conversation acts for dialogue, including actions for attention, turn-taking and grounding as well as conveying propositional information.

- [Stone 2003] Presents SPUD (sentence planning using description), a computational framework for constructing utterances as coordinated ensembles of complex action. SPUD explores a search space for utterances described by a linguistic grammar. At each stage of search, SPUD uses a model of interpretation, which characterizes the potential links between the utterance and the domain and context, to assess its progress towards constructing a satisfactory utterance. SPUD realizes an integrated message model: it constructs the syntax, semantics and pragmatics of an incomplete utterance simultaneously, and can work incrementally to achieve a range of communicative functions.

Opening and closing

- [Schegloff 1973] Investigates how interlocutors end conversations; finishing up requires mutual consent, and Schegloff finds that interlocutors summarize and agree on what's happened in a conversation before they move it to a close.

Floor management

- [Duncan 1974] Calls attention to several cues that the speaker employs to indicate the end of a turn or invite the hearer to take a turn. These cues include not only verbal signs that an utterance is ending but also nonverbal cues, such as the speaker's looking away from the hearer as an utterance begins and toward the hearer as the utterance ends.

- [Sacks 1974] An insightful analysis of turn-taking in which pragmatic knowledge is seen as providing opportunities for assigning and taking a speaking turn; thus, the rules of turn-taking are subject to the control of the participants and actual turns result from the interaction of the rules with the goals and beliefs of interlocutors.

- [Novick 1988] A remarkably forward-thinking computational model of conversation that describes dialogue in terms of actions affecting the state of the communication on several layers, including layers for attention and turn-taking. Evaluated by simulations that demonstrate that the model enabled artificial agents to communicate with one another while reproducing characteristics of natural human-human dialogue.

- [Jarmon 1996] A dissertation—distributed on CD rom, with movies!—that features an instructive collection of examples of floor management in action, with a particular analysis of the role of eye gaze and head orientation in signaling expectations and opportunities for changes of speaker.

- [Cassell 1999b] An evaluation of floor management in GANDALF; see also [Thórisson 1997] under embodied conversational agents. The agent's nonverbal cues to turn-taking contributed to more fluid interactions, in which users repeated themselves less, hesitated less, and got frustrated less (compared both to a control version of the system without any nonverbal cues and to an alternative with emotional feedback about the progress of the interaction). Users also notice this lifelikeness and fluidity, as revealed in their responses to an exit questionnaire.

Grounding

- [Clark 1989] Argues that participants in conversation generally work to achieve mutual belief about the effects of their actions (putting in as much effort on coordination as their purposes require). Analyzes utterances in dialogue as a special case; most utterances in conversation are taken up in two stages, a first where the speaker presents it and a second where the hearer accepts it by giving evidence that they understand.

- [Clark 1991] A general look at grounding—establishing content as part of the common ground well enough for the purposes of the conversation. Emphasizes that grounding is a collaborative effort that interlocutors try to accomplish efficiently, and that patterns of grounding correspondingly change depending on the content that has to be grounded and the communicative behaviors available to the interlocutors.

- [Allwood 1992] A great paper on the use of feedback in dialogue. While developed independently of Clark, this work comes to much the same conclusions about the need to acknowledge contributions to dialogue at levels of attention, understanding, and acceptance. Also discusses different kinds of communication (inspired by Peircean semiotics), and evocative and expressive functions of utterances (basically the important distinction between utterance functions that look backward at the previous moves in the conversation and those that look forward to the pending goals of the conversation).

- [Traum 1994a] A computational investigation of grounding. Outlines how the process of reaching mutual understanding can be modeled algorithmically, and describes the design of conversational agents that plan and recognize the grounding actions that can move this process forward. In this model, ungrounded contributions to a task-oriented dialogue are provisional, and only grounding actions update the agreed state of the conversation.

- [Dillenbourg 1996] An empirical study of grounding in human–human multimodal computer-supported collaborative problem-solving dialogues. Pairs of subjects performed a diagnosis task (solving a murder mystery in a simulated environment), and communicated by typing and drawing. Interestingly, subjects do grounding through words, drawing, and

action in the virtual environment, and often use one modality to ground information presented in another.

- [Matheson 2000] Implements an information-state model of grounding. Takes the idea that moves in dialogue trigger obligations for grounding from [Traum 1994b]—see Negotiation, below—and shows how to formalize this: obligations are ingredients of the dialogue state, new utterances come with updates that introduce grounding obligations and grounding moves come with updates that discharge these obligations.

- [Nakano 2003] Investigates grounding in face-to-face conversation and and describes an agent that combines verbal and non-verbal signals to establish common ground in human–computer interaction. A major result is that listeners do give negative feedback when they do not understand adequately, and that speakers can take lack of negative feedback as evidence that they have been understood.

- [Purver 2004] A fine-grained empirical and computational study of clarification requests and elliptical responses. These constructions are the bread-and-butter methods interlocutors use to ground problematic utterances in spoken dialogue. Purver shows how these utterances can be handled in their natural complexity with a declarative grammar and an elegant information-state dialogue manager.

Rapport

- [Brown 1987] A book-length survey of politeness phenomena in language, explaining in detail how polite language in conversation smooths interpersonal relationships by supporting people's self-image and self-presentation.

- [Tickle-Degnen 1990] Describes rapport in face-to-face conversation as a complex phenomenon with contributions from the coordinated attention of interlocutors (as evidenced by eye movements or spoken backchannel feedback), positive engagement (as evidenced by encouraging words or behaviors such as leaning forward) and coordination (as evidenced by smooth transitions between speaking turns and an evolving shared conceptualization and vocabulary). Significantly, the construct of rapport applies uniformly both to verbal and nonverbal signals in conversation.

- [Svennevig 1999] A sociolinguistic investigation of the use of small talk to establish trust in human–human interactions.

- [Cassell 2002] Explores the modeling of relationships and rapport in embodied conversational agents. Formulates a taxonomy of verbal and nonverbal strategies (such as small talk) which can be used to move the relationship in a desired direction, and describes a dialogue planner which can plan conversational strategies that work towards the achievement of both task goals and relational goals. Evaluation shows that the relational talk leads some users to trust the system more.

Negotiation and collaboration

- [Pollack 1990] A clear characterization of what mental attitudes an agent must have before committing to a plan. A plan sets out what the agent is to do, when and in what circumstances the agent is to act, and what outcome the agent will thereby achieve. To adopt a plan, the agent must must believe that the circumstances laid out in the plan will obtain, must expect to carry out the actions in the plan, must believe that the outcome spelled out by the plan will occur, and must desire that outcome. This simple idea has deep consequences for plan recognition, for dialogue management and for collaboration generally.

- [Grosz 1990] Extends Pollack's mental-state characterization of plans to shared plans involving multiple agents, and uses them to account for patterns of negotiation between collaborators in dialogue. The central question is the mutual beliefs, goals and commitments that individual agents must adopt in a coordinated way as part of agreeing to do something together.

- [Cohen 1991] Describes the responsibilities that agents have when they join a team that is working collaboratively on shared goals. For example, an agent must not only carry out the actions it commits to do as part of a collaboration, but it must do so in a way that allows its collaborators to recognize the contribution it is making to the joint activity. Similarly, an agent must report problems it encounters as well as confirming successes it achieves.

- [Walker 1990] Explores the transfer of initiative between participants in human–human conversation. Shows that speakers have ways of ceding initiative to their interlocutors and ways of taking control of the dialogue for themselves; argues that these dynamics of initiative help explain both the content people provide in discourse and the attentional and intentional structures that tie discourse together.

- [Traum 1994b] Presents a computational model of contributing to dialogue in which grounding is taken to be part of agents' obligations for dialogue. In this model, obligations represent constraints on interactions that agents simply adhere to; obligations are treated separately from the flexible processes of general goal-directed deliberation normally considered in planning models. The model thus realizes Clark's idea that grounding—making sure that utterances are understood and that their contributions to conversation are agreed—is simply part of what conversation is.

- [Lochbaum 1998] Extends Grosz and Sidner's shared plan model to describe problem-solving dialogue. Characteristic problem-solving activities include identifying goals that need to be achieved, identifying subtasks to perform and selecting suitable parameters for them, allocating them to individual agents, and jointly assessing the results once agents have acted. A key insight of the work is that all these problem-solving processes are collaborative—just as real-world actions are. So all these processes can and should be described in the same framework of shared plans.

- [Carberry 1999] A flexible model of negotiation that emphasizes the indirect role utterances play in achieving real-world goals. For example, problem-solving dialogue may play out over several turns before one party's underlying domain goal is fully specified. The model therefore allows different representations and inference processes for plan recognition at the utterance level, problem-solving level and domain level.

- [Blaylock 2003] Describes individual utterances in natural conversation in terms of a model of collaboration in a dynamic domain that combines the indirection and collaboration of previous approaches. It factors the course of collaboration into primitive joint steps, such as agreeing to adopt an action into a plan, or agreeing to carry out an action immediately. It characterize utterances by linking them with abstract communicative moves which specify one agent's contribution to one of these joint steps, such as initiating or completing one.

- [Traum 2003] Describes a negotiation model for multi-party dialogues, including agents in a dynamic virtual environment. Shows how to formalize negotiation moves in terms of their effect on the state of the interaction and how to model agents' obligations to address and ground these moves.

Reference

- [Clark 1986] An empirical study of what people naturally do in dialogue when one needs to identify an object for another. Finds that speakers are prepared to give many alternative descriptions, and listeners not only show whether they understand each description but often actively help the speaker find one they do understand. These patterns of 'collaborative reference' provide clear motivation for modeling language use as joint activity.

- [Kronfeld 1986] While most computational work focuses on descriptions that identify objects, it's important to note that people have other goals in formulating descriptions, such as characterizing an object or making an argument. Kronfeld's work (one of the few on this topic) emphasizes that all of these communicative goals are compatible with cognitive models of language use based on joint activity and with plan-based approaches to natural language generation in particular.

- [Garrod 1987] An empirical study of communication in dialogue—using a maze task where interlocutors must identify their positions. Highlights the coordinated conceptual and linguistic representations that interlocutors come to achieve; for example, speakers come to alignment in whether they refer to places using coordinates in a grid like *A4* or qualitiative descriptions like *T-shape*.

- [Dale 1992] A computational study of reference in recipes, a challenging domain involving sets of objects, quantities of stuff, and processes that create, destroy and radically transform their raw materials. Still a useful survey of generating a wide variety of noun phrases from rich underlying knowledge of the world.

- [Prince 1992] A linguistic study of the pragmatics of reference. Investigates the form and information status of subjects, using the correlations to explore hypotheses about information structure and the organization of discourse. In particular, subjects turn out to be preferentially discourse old, suggesting a role for a subject or topic position in establishing links among propositions in discourse.

- [Gundel 1993] Argues that the English form that is used to describe a referent corresponds to its status in the information state of the discourse. This paper postulates a hierarchy with six different levels that indicate progressively stronger levels of prominence and correspond to specific categories of form. For example, a referent that is *in focus* can be referenced with a pronoun; it is not only *familiar* (available for reference with *that N*) and *active* (available for reference with *this N*), but because of the recent discourse or the mutual environment it is maximally prominent.

- [Heeman 1995] A computational model of reference as goal-directed activity in dialogue. Explains speakers' references by attributing detailed plan representations and corresponding mental states to speakers. Shows how these kinds of representations can support interactive refinements to reference plans, so that interlocutors are able to work together to achieve mutual understanding.

- [Dale 1995] An investigation of how speakers and hearers coordinate on referring expressions and avoid the Gricean implicatures that may accompany marked referring forms. Argues from an analysis of empirical data and considerations of computational complexity that hearers should attribute only a narrow, local motivation to speakers' descriptive choices. Shows that a simple and fast procedure can in fact generate referring expressions that meet this requirement.

- [Brennan 1996] An experimental study of the referring forms speakers use across extended conversation, demonstrating that interlocutors come to agree on the way they describe objects, and in addition providing suggestive evidence that interlocutors associate these agreements with the specific participants in a conversation.

- [Brown-Schmidt 2002] A corpus study of reference in human-human task-oriented dialogue (a directed assembly task); one participant in each dialogue was eye-tracked. The results document speakers' abbreviated referring forms in context and hearers' unproblematic interpretations of these forms—suggesting that interlocutors precisely coordinate attention in face-to-face conversation based on a range of criteria including recent mention, spatial prominence and task relevance.

Spatial descriptions

- [Landau 1993] A suggestive and wide-ranging position paper exploring possible connections between spatial language, spatial concepts and our abilities to perceive and act in space. Suggests that linguistic universals about the expression of the locations and relationships of objects in space—see Talmy below—may have their origin in a modality-independent conceptual level of spatial representation that (perhaps because of the constraints of brain structure) only has access to limited kinds of spatial information.

- [Di Eugenio 1996] A computational model of spatial and causal reasoning in understanding natural language instructions. Shows that instructions can tell you what to do in part by telling you why to do it, and therefore concludes that the meanings of spatial descriptions—even those realized exclusively in language—should be represented in terms of ensembles of semantic constraints that function together to achieve a speaker's communicative goals.

- [Talmy 2000] A two volume anthology of Talmy's work exploring the varied ways language can present the causal and spatial properties of real-world events and states. Considers the meanings languages use to portray relationships in space, to distinguish figure from ground, to describe the forces with which entities interact and to portray relationships between events. Includes both useful descriptive crosslinguistic taxonomies and detailed analyses for English.

- [Emmorey 2000] Investigates how communicators describe space in different systems, including spoken language, signed language and coverbal gesture. A number of interesting choices are always available, including whether to adopt a survey perspective, giving an overview of space as though in a map, or to adopt a route perspective, portraying what it's like to move around a space.

Extended discourse

- [Halliday 1976] The classic description of cohesion in English—the use of repeated and elliptical forms across dis-

course in such a way that the interpretation of discourse involves a dense network of mutually interconnected references and concepts.

- [Preece 1992] Extended discourse is collaborative too, as this study of children's everyday storytelling nicely demonstrates. Preece's study found that children's interactions with each other contributed to the modification, expansion, increased coherence, and complexity of their anecdotes and stories, and revealed that children are active, alert, engaged, and even aggressive listeners. In everyday storytelling, children become collaborators and facilitators of peer narrations, egging one another on, and also act as critics and correctors, pointing out flaws and taking issue with peers' stories.

- [Kehler 2001] A recent integrative survey of coherence, the relationships that we infer to connect successive units of discourse. Argues for systematizing these relationships in terms of three broad categories: explanatory relationships between ideas, resemblance of similar ideas, and extended descriptions of a particular situation. Each category involves distinct consequences for linguistic structure and attentional state. A nice introduction and summary of Kehler's perspective on discourse is his contribution to Jurafsky and Martin's textbook [Kehler 2000]

- [McNeill 2001] An exploration of the *catchment* through state-of-the-art instrumental analysis of natural conversation. The catchment is a recurrent "image" in gesture linked with a specific topic in discourse, with a specific kind of gestural depiction, and with specific units of spoken discourse. Catchments thus provide a way of analyzing patterns of gesture as contributing to the coherence of discourse.

Integrative Systems

Embodiment without conversational interaction

- [Bates 1992] An introduction to the OZ project at CMU, an early effort to combine ideas from computer graphics, artificial intelligence, and theater to create interactive characters for entertainment. The research focuses on how the tools of AI could support artists in the design of graphical characters that convey the impression of awareness of and responsiveness to the world around them, convey their goals and their affective state, and engage the people they interact with.

- [Perlin 1996] Presents IMPROV, an influential procedural model for animating characters which emphasizes the author's design of character behavior. In IMPROV a designer can define a character's repertoire of actions and decision-making, and can also create procedures that give the character a consistent but variable manner of motion (and thereby help to portray distinctive affect or personality).

- [Vilhjálmsson 1998] Describes the BODYCHAT system, which automatically creates visualizations to accompany text chat for online communities. Uses models of nonverbal communication in conversational opening, turn-taking, and closing in order to realize a believable animation of characters engaging in chat dialogue.

- [Sengers 1999] Describes the design of the INDUSTRIAL GRAVEYARD, a virtual world for entertainment. In effect, agents are designed to communicate their internal state to the user—the architecture makes sure that the animated actions characters perform make the characters' choices visible. This involves designing behaviors that signal what the agent is doing, why the agent is doing it, and when the agent's plans and goals change; it also involves making sure the user can see and notice these behaviors.

- [Cassell 2000a] Describes SAM, a character that uses models of nonverbal communication to act as a supportive listener to children's stories; SAM also has its own stories to tell and its own agenda for play. Note that while SAM listens it does not understand.

- [André 2000] Describes an approach to generating graphical interactions in which teams of agents interact with one another to present information to the user—embodied conversation as theater rather than user interaction. They demonstrate the approach by realizing a team of commentators for a robocup soccer game, and a presentation team involving a car salesman and a potential buyer.

- [Breazeal 2002] Describes KISMET, an interactive physical robot. Takes the regulation of conversation as a starting point, and explores how a robot can be designed to interact with people and reproduce the back-and-forth interactive style of fluid conversation. No language or collaboration...yet.

Systems for natural spoken dialogue

- [Wahlster 2000] Describes VERBMOBIL, an interactive system for speech-to-speech translation that supported task-oriented dialogues for meeting scheduling. A key research theme in VERBMOBIL is achieving robustness by combining modules for shallow and deep processing of language and dialogue within a single architecture.

- [Theune 2001] Describes the natural language generation process in D2S, a generic speech planning system that has been used in music database and travel domains. Illustrates how discourse planning, reference generation, information structure and prosodic control can be staged to yield high-quality output for applied spoken language systems.

- [Allen 2001] A position paper motivating the current conversational architecture from James Allen's lab at Rochester. Emphasizes the importance of incremental processing in interactive dialogue and the need to manage domain action and domain problem solving in tandem with task-oriented dialogue moves. Grows out of influential work on the TRAINS [Allen 1995] and TRIPS [Ferguson 1998] systems.

- [Johnston 2002] Describes MATCH (multimodal access to city help) a guide to New York City realized as a state-of-the-art multimodal interface for a tablet display. User utterances can combine speech and pen; output combines text (and synthesized speech) as well as graphics such as maps and diagrams. Particular strengths of the work include robust multimodal understanding to arrive at integrated interpretation for user input and customizable dialogue management based on a detailed decision-theoretic model of user goals and preferences.

Embodied conversational agents

- [Nagao 1994] Describes a prototype system with an animated face that could use facial displays for feedback about the state of the conversation. Implemented a variety of patterns observed by Chovil [Chovil 1991a] (see behavior—facial expressions). While listening, the system used the face to indicate attention, understanding and agreement. During its own utterances, the system's face marked emphasis and portrayed

aspects of the interactive process. An experimental evaluation suggests that nonverbal cues allow the system to enable smoother interaction than a speech-only version of the system (though users quickly compensate for the speech-only system by learning to use it effectively).

- [Cassell 1994a] Describes a system for ANIMATED CONVERSATION in which contributions to conversation, including nonverbal and verbal cues, are planned and synthesized automatically. This work focused on the representation of characters' behaviors for conversation, on the design of animation processes that could realize the animation with appropriate synchrony and control, and on orchestrating the use of gesture and intonation to reflect an agent's communicative plans and context. See also [Cassell 1994b].

- [Thórisson 1997] Describes GANDALF, a graphical character that could give virtual tours of the solar system. Turn-taking is one of GANDALF's most relevant and robust features. He addresses the problem of real-time turn-taking by integrating turn-taking with a model of interlocutors' activity. Interlocutors act either as speaker or as hearer; each role is recognizably associated with different classes of behaviors, including perceptual, decision, and motor tasks.

- [Rickel 1999] Describes STEVE, a virtual human for procedural training. Among STEVE's strengths was its effective use of space to communicate—STEVE would move to objects in the virtual world and then generate a deictic gesture at the beginning of an explanation about that object. Another research contribution for STEVE was the ability to demonstrate and to describe action from the same underlying task representation.

- [Lester 1999] Describes COSMO, a lifelike pedagogical agent guiding and critiquing students' problem-solving in a 3D simulation. Here the main challenge was using a character to motivate and engage students; the solution involves coordinating communicative behavior with emotional reactions to students, and effectively animating the character's whole body in virtual space to deliver a convincingly lifelike presentation.

- [Cassell 1999a] Describes REA, a virtual real estate agent; REA was a platform for reconciling floor management models with deep (and improved) models of gesture synthesis [Cassell 2000e]. Later, REA served as a testbed for exploring posture [Cassell 2001b] and rapport [Cassell 2002]. See also [Cassell 2000b].

- [Wahlster 2001] Describes SMARTKOM, a system for conversational interaction, including an interactive character, for a range of domains including home automation. Advocated the goal of platform-independence in language interface technology, and investigated ways of enabling a single architecture to be deployed across very different applications, devices and interface personas.

- [Rich 2001] Describes COLLAGEN, a platform for building interactive agents using the theory of shared plans and collaborative discourse of Grosz and Sidner [Grosz 1990] and Lochbaum [Lochbaum 1998]; see Functions—Negotiation. The focus is on algorithms for plan recognition, and the representation of user tasks and human-computer interactions in collaborative terms. As the project has progressed, the architecture has been applied to handle a diverse range of interactive domains, including embodied conversation with a physical robot [Sidner 2004; Lesh 2004].

- [Bickmore 2003] Describes LAURA, an exercise advisor designed to monitor and encourage users' efforts to get in shape over a period of a month. Focuses on the long-term relationship the system needed to develop with its users, including the problems of drawing on previous interactions and establishing expectations for future ones.

- [Rickel 2002] Describes MRE, the Mission Rehearsal Exercise, an immersive virtual world for simulation-based training where students learn by role playing and interacting with virtual characters. Central questions for this project include how to model complex conversations across multiple participants in a dynamic environment, and how to reflect the cognitive and emotional states of thoroughgoing intelligent agents in appropriate conversational strategies and linguistic choices. See also [Hill 2003].

- [Kopp 2003] Describes MAX, the Multimodal Assembly Expert, an virtual character in a domain of construction tasks. Focuses on the real-time synthesis of gesture, facial expressions and speech and delivering them as natural animation.

- [Matheson 2003] Describes MAGICSTER, a wide-ranging European Project on conversational agents, including research on specifying, animating and evaluating embodied characters and research on handling a broader range of conversations, including settings in which the user observes or participates in an interaction among several characters.

- [Theune 2005] Describes ANGELICA, an embodied character for generating route descriptions in a 3D virtual building. Works with a rich markup language for describing the coordination of nonverbal behavior with simultaneous speech, and emphasizes that such representations in fact allow standard techniques and architectures from natural language generation to be retarged to output utterances for embodied conversational agents.

References

ALLEN, J. F., SCHUBERT, L. K., FERGUSON, G., HEEMAN, P., HWANG, C. H., KATO, T., LIGHT, M., MARTIN, N. G., MILLER, B. W., POESIO, M., AND TRAUM, D. R. 1995. The TRAINS Project: A case study in building a conversational planning agent. *Journal of Experimental and Theoretical AI 7*, 7–48.

ALLEN, J., FERGUSON, G., AND STENT, A. 2001. An architecture for more realistic conversational systems. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI 2001)*, 1–8.

ALLWOOD, J., NIVRE, J., AND AHLSEN, E. 1992. On the Semantics and Pragmatics of Linguistic Feedback. *Journal of Semantics 9*, 1–26.

ANDRÉ, E., AND RIST, T. 2000. Presenting through performing: On the use of multiple animated characters in knowledge-based presentation systems. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI 2000)*, 1–8.

ARGYLE, M., AND COOK, M. 1976. *Gaze and Mutual Gaze*. Cambridge University Press, Cambridge.

BADLER, N., ALLBECK, J., ZHAO, L., AND BYUN, M. 2002. Representing and parameterizing agent behaviors. In *Proceedings of Computer Animation*, 133–143.

BATES, J. 1992. Virtual reality, art and entertainment. *PRESENCE: Teleoperators and Virtual Environments 1*, 133–138.

BAVELAS, J. B., CHOVIL, N., COATES, L., AND ROE, L. 1995. Gestures specialized for dialogue. *Personality and Social Psychology 21*, 394–405.

BAVELAS, J. B., AND CHOVIL, N. 2000. Visible acts of meaning: An integrated message model of language in face-to-face dialogue. *Journal of Language and Social Psychology 19*, 163–194.

BESKOW, J., EDLUND, J., AND NORDSTRAND, M. 2002. Specification and Realisation of Multimodal Output in Dialogue Systems. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 2002)*, vol. 1, 181–184.

BICKMORE, T. W. 2003. *Relational Agents: Effecting Change through Human–Computer Relationships*. PhD thesis, MIT.

BLAYLOCK, N., ALLEN, J., AND FERGUSON, G. 2003. Managing communicative intentions with collaborative problem solving. In *Current and New Directions in Dialogue*, R. Smith and J. van Kuppevelt, Eds. Kluwer, Dordrecht, 63–84.

BRACHMAN, R., MCGUINNESS, D., SCHNEIDER, P. P., RESNICK, L. A., AND BORGIDA, A. 1990. Living with CLASSIC: when and how to use a KL-ONE-like language. In *Principles of Semantic Networks*, J. Sowa, Ed. Morgan Kaufmann, 401–456.

BREAZEAL, C. 2002. Regulation and entrainment for human–robot interaction. *International Journal of Experimental Robotics 21*, 883–902.

BRENNAN, S. E., AND CLARK, H. H. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition 22*, 1482–1493.

BROWN-SCHMIDT, S., CAMPANA, E., AND TANENHAUS, M. K. 2002. Reference resolution in the wild: On-line circumscription of referential domains in a natural interactive problem-solving task. In *Proceedings of the Cognitive Science Society*, 148–153.

BROWN, P., AND LEVINSON, S. C. 1987. *Politeness: Some Universals in Language Use*. Cambridge University Press, Cambridge.

BUCK, R. 1991. Social Factors in Facial Display and Communication: A Reply to Chovil and Others. *Journal of Nonverbal Behavior 15*, 155–161.

BYUN, M., AND BADLER, N. 2002. FacEMOTE: qualitative parametric modifiers for facial animations. In *ACM SIGGRAPH Symposium on Computer Animation*, 65–71.

CARBERRY, S., AND LAMBERT, L. 1999. A process model for recognizing communicative acts and modeling negotiation subdialogues. *Computational Linguistics 25*, 1–53.

CARLETTA, J. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics 22*, 249–254.

CARLETTA, J., ISARD, A., ISARD, S., KOWTKO, J. C., DOHERTY-SNEDDON, G., AND ANDERSON, A. H. 1997. The reliability of a dialogue structure coding scheme. *Computational Linguistics 23*, 13–31.

CASSELL, J., PELACHAUD, C., BADLER, N., STEEDMAN, M., ACHORN, B., BECKET, T., DOUVILLE, B., PREVOST, S., AND STONE, M. 1994. Animated Conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1994)*, 413–420.

CASSELL, J., STONE, M., DOUVILLE, B., PREVOST, S., ACHORN, B., STEEDMAN, M., BADLER, N., AND PELACHAUD, C. 1994. Modeling the interaction between speech and gesture. In *Proceedings of the Cognitive Science Society*, 153–158.

CASSELL, J., BICKMORE, T., BILLINGHURST, M., CAMPBELL, L., CHANG, K., VILHJÁLMSSON, H., AND YAN, H. 1999. Embodiment in Conversational Characters: Rea. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1999)*, 520–527.

CASSELL, J., AND THÓRISSON, K. 1999. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence 13*, 519–538.

CASSELL, J., ANANNY, M., BASU, A., BICKMORE, T., CHONG, P., AND MELLIS, D. 2000. Shared reality: Physical collaboration with a virtual peer. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, 259–260.

CASSELL, J. 2000. Embodied conversational interface agents. *Communications of the ACM 43*, 70–78.

CASSELL, J. 2000. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. In *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, Eds. MIT Press, Cambridge, MA, 1–28.

CASSELL, J., BICKMORE, T., CAMPBELL, L., VILHJALMSSON, H., AND YAN, H. 2000. Human conversation as a system framework. In *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, Eds. MIT Press, Cambridge, MA, 29–63.

CASSELL, J., STONE, M., AND YAN, H. 2000. Coordination and context-dependence in the generation of embodied conversation. In *First International Conference on Natural Language Generation*, 171–178.

CASSELL, J., SULLIVAN, J., PREVOST, S., AND CHURCHILL, E., Eds. 2000. *Embodied Conversational Agents*. MIT Press, Cambridge.

CASSELL, J., VILHJÁLMSSON, H., AND BICKMORE, T. 2001. BEAT: the behavioral expression animation toolkit. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001)*, 477–486.

CASSELL, J., NAKANO, Y., BICKMORE, T. W., SIDNER, C. L., AND RICH, C. 2001. Non-verbal cues for discourse structure. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2001)*, 106–115.

CASSELL, J., AND BICKMORE, T. 2002. Negotiated Collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and Adaptive Interfaces 12*, 1–44.

CAVE, C., GUAITELLA, I., BERTRAND, R., SANTI, S., HARLAY, F., AND ESPESSER, R. 1996. About the relationship between eyebrow movements and F0 variations. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 1996)*, 2175–2179.

CHI, D., COSTA, M., ZHAO, L., AND BADLER, N. 2000. The EMOTE model for effort and shape. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2000)*, 173–182.

CHOVIL, N. 1991. Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction 25*, 163–194.

CHOVIL, N. 1991. Social determinants of facial displays. *Journal of Nonverbal Behavior 15*, 141–154.

CHOVIL, N., AND FRIDLUND, A. J. 1991. Why emotionality cannot equal sociality: Reply to Buck. *Journal of Nonverbal Behavior 15*, 163–167.

CLARK, H. H., AND WILKES-GIBBS, D. 1986. Referring as a collaborative process. *Cognition 22*, 1–39.

CLARK, H. H., AND SCHAEFER, E. F. 1989. Contributing to discourse. *Cognitive Science 13*, 259–294.

CLARK, H. H., AND BRENNAN, S. E. 1991. Grounding in communication. In *Perspectives on Socially-Shared Cognition*. American Psychological Association, Washington, DC, 127–149.

CLARK, H. H. 1996. *Using Language*. Cambridge University Press, Cambridge.

COHEN, P. R., AND LEVESQUE, H. J. 1991. Teamwork. *Noûs 24*, 487–512.

CONDON, W. S., AND OSGTON, W. D. 1971. Speech and body motion synchrony of the speaker–hearer. In *The Perception of Language*, D. Horton and J. Jenkins, Eds. Academic Press, 150–184.

DAHLBÄCK, N., JÖNSSON, A., AND AHRENBERG, L. 1993. Wizard of Oz studies—Why and how. In *Proceedings of the International Workshop on Intelligent User Interfaces (IUI 1993)*, 193–200.

DALE, R. 1992. *Generating Referring Expressions: Constructing Descriptions in a Domain of Objects and Processes*. MIT Press, Cambridge MA.

DALE, R., AND REITER, E. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science 18*, 233–263.

DECARLO, D., REVILLA, C., STONE, M., AND VENDITTI, J. 2002. Making discourse visible: coding and animating conversational facial displays. In *Proceedings of Computer Animation*, 11–16.

DECARLO, D., REVILLA, C., STONE, M., AND VENDITTI, J. 2004. Specifying and animating facial signals for discourse in embodied conversational agents. *Computer Animation and Virtual Worlds 15*, 27–39.

DI EUGENIO, B., AND WEBBER, B. 1996. Pragmatic overloading in natural language instructions. *International Journal of Expert Systems 9*, 53–84.

DILLENBOURG, P., TRAUM, D., AND SCHNEIDER, D. 1996. Grounding in Multi-modal Task-Oriented Collaboration. In *Proceedings of the European Conference on Artificial Intelligence in Education*, 415–425.

DUNCAN, STARKEY, J. 1974. Some signals and rules for taking speaking turns in conversation. In *Nonverbal Communication*, S. Weitz, Ed. Oxford University Press, Oxford, 2998–311.

DUNCAN, JR., S., AND FISKE, D. W. 1977. *Face-to-face Interaction: Research, Methods, and Theory*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

EKMAN, P., AND FRIESEN, W. V. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica 1*, 49–98.

EKMAN, P., AND FRIESEN, W. V. 1975. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Prentice-Hall, Englewood Cliffs, New Jersey.

EKMAN, P., AND FRIESEN, W. V. 1978. *Facial Action Coding System*. Consulting Psychologists Press, Palo Alto, CA.

EKMAN, P. 1979. About brows: Emotional and conversational signals. In *Human Ethology: Claims and Limits of a New Discipline: Contributions to the Colloquium*, M. von Cranach, K. Foppa, W. Lepenies, and D. Ploog, Eds. Cambridge University Press, Cambridge, 169–202.

EMMOREY, K., TVERSKY, B., AND TAYLOR, H. 2000. Using space to describe space: Perspective in speech, sign and gesture. *Spatial Cognition and Computation 2*, 157–180.

ENGLE, R. A. 2000. *Toward a Theory of Multimodal Communication: Combining Speech, Gestures, Diagrams and Demonstrations in Instructional Explanations*. PhD thesis, Stanford University.

FAIGIN, G. 1990. *The Artist's Complete Guide to Facial Expressions*. Watson-Guptill, New York.

FERGUSON, G., AND ALLEN, J. F. 1998. TRIPS: An Intelligent Integrated Problem-Solving Assistant. In *Proceedings of the National Conference on Artificial Intelligence (AAAI 1998)*, 567–573.

GARROD, S., AND ANDERSON, A. 1987. Saying what you mean in dialog: a study in conceptual and semantic co-ordination. *Cognition 27*, 181–218.

GOODWIN, C. 1981. *Conversational Organization: Interaction between Speakers and Hearers*. Academic Press, New York.

GROSZ, B., AND SIDNER, C. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics 12*, 175–204.

GROSZ, B. J., AND SIDNER, C. L. 1990. Plans for discourse. In *Intentions in Communication*, P. Cohen, J. Morgan, and M. Pollack, Eds. MIT Press, Cambridge MA, 417–444.

GUNDEL, J. K., HEDBERG, N., AND ZACHARSKI, R. 1993. Cognitive status and the form of referring expressions in discourse. *Language 69*, 274–307.

HALLIDAY, M. A. K., AND HASAN, R. 1976. *Cohesion in English*. Longman, London.

HEEMAN, P., AND HIRST, G. 1995. Collaborating on referring expressions. *Computational Linguistics 21*, 351–382.

HILL, RANDALL W., J., GRATCH, J., MARSELLA, S., RICKEL, J., SWARTOUT, W., AND TRAUM, D. 2003. Virtual humans in the mission rehearsal exercise system. *Künstliche Intelligenz 17*, 5–10.

HIRSCHBERG, J. 1993. Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence 63*, 305–340.

HIRSCHBERG, J., AND NAKATANI, C. 1996. A prosodic analysis of discourse segments in direction-giving monologues. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 1996)*, 286–293.

HOBBS, J. R. 1990. The Pierrehumbert-Hirschberg theory of intonational meaning made simple: Comments on Pierrehumbert and Hirschberg. In *Intentions in Communication*, P. R. Cohen, J. Morgan, and M. E. Pollack, Eds. MIT Press, Cambridge, 313–323.

JARMON, L. 1996. *An Ecology of Embodied Interaction: Turn-Taking and Interactional Syntax in Face-to-Face Encounters*. PhD thesis, University of Texas at Austin.

JOHNSTON, M., BANGALORE, S., VASIREDDY, G., STENT, A., EHLEN, P., WALKER, M., WHITTAKER, S., AND MALOOR, P. 2002. MATCH: An architecture for multimodal dialogue systems. In *ACL*, 376–383.

KEHLER, A. 2000. Discourse. In *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, D. Jurafsky and J. H. Martin, Eds. Prentice-Hall, New York.

KEHLER, A. 2001. *Coherence, Reference and the Theory of Grammar*. CSLI Publications, Stanford.

KENDON, A. 1972. Some relationships between body motion and speech. In *Studies in Dyadic Communication*, A. W. Siegman and B. Pope, Eds. Pergamon, New York, 177–210.

KENDON, A. 1990. *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge University Press, New York.

KOPP, S., AND WACHSMUTH, I. 2002. Model-based animation of coverbal gesture. In *Proceedings of Computer Animation*, 252–257.

KOPP, S., JUNG, B., LESSMANN, N., AND WACHSMUTH, I. 2003. Max—A multimodal assistant in virtual reality construction. *Künstliche Intelligenz 17*, 11–17.

KOPP, S., TEPPER, P., AND CASSELL, J. 2004. Towards integrated microplanning of language and iconic gesture for multimodal output. In *Proceedings of the International Conference on Multimodal Interfaces (ICMI 2004)*, 97–104.

KOPP, S., AND WACHSMUTH, I. 2004. Synthesizing multimodal utterances for conversational agents. *Computer Animation and Virtual Worlds 15*, 39–52.

KRAHMER, E., RUTTKAY, Z., SWERTS, M., AND WESSELINK, W. 2002. Preceptual evaluation of audiovisual cues to prominence. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 2002)*, vol. 3, 1933–1936.

KRONFELD, A. 1986. Donellan's distinction and a computational model of reference. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 1986)*, 186–191.

LANDAU, B., AND JACKENDOFF, R. 1993. 'What' and 'where' in spatial language and spatial cognition. *Behavioral and Brain Science 16*, 217–265.

LARSSON, S., AND TRAUM, D. 2000. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering 6*, 323–340.

LEE, S. P., BADLER, J. B., AND BADLER, N. I. 2002. Eyes alive. *ACM Transactions on Graphics (Special issue for SIGGRAPH 2002) 21*, 637–644.

LESH, N., MARKS, J., RICH, C., AND SIDNER, C. L. 2004. 'Man-Computer Symbiosis' revisited: Achieving natural communication and collaboration with computers. *IEICE Transactions on Information and Systems E87-D*, 1290–1298.

LESTER, J. C., TOWNS, S. G., AND FITZGERALD, P. J. 1999. Achieving affective impact: Visual emotive communication in lifelike pedagogical agents. *International Journal of Artificial Intelligence in Education 10*, 278–291.

LOCHBAUM, K. E. 1998. A collaborative planning model of intentional structure. *Computational Linguistics 24*, 525–572.

MATHESON, C., POESIO, M., AND TRAUM, D. 2000. Modelling grounding and discourse obligations using update rules. In *Proceedings of the Annual Meeting of the North American Association of Computational Linguistics (NAACL 2000)*, 1–8.

MATHESON, C., PELACHAUD, C., DE ROSIS, F., AND RIST, T. 2003. MagiCster: Believable Agents and Dialogue. *Künstliche Intelligenz 17*, 24–29.

MCNEILL, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press, Chicago.

MCNEILL, D., QUEK, F., MCCULLOUGH, K.-E., DUNCAN, S., FURUYAMA, N., BRYLL, R., MA, X.-F., AND ANSARI, R. 2001. Catchments, Prosody and Discourse. *Gesture 1*, 9–33.

NAGAO, K., AND TAKEUCHI, A. 1994. Speech dialogue with facial displays: Multimodal human–computer conversation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 1994)*, 102–109.

NAKANO, Y. I., REINSTEIN, G., STOCKY, T., AND CASSELL, J. 2003. Towards a model of face-to-face grounding. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2003)*, 553–561.

NOVICK, D. 1988. *Control of Mixed-Initiative Discourse Through Meta-Locutionary Acts: A Computational Model*. PhD thesis, University of Oregon. also available as U. Oregon Computer and Information Science Tech Report CIS-TR-88-18.

NOVICK, D. G., HANSEN, B., AND WARD, K. 1996. Coordinating turn-taking with gaze. In *Proceedings of the International Conference on Spoken Language Processing ICSLP*, vol. 3, 1888–1891.

PELACHAUD, C., BADLER, N., AND STEEDMAN, M. 1996. Generating facial expressions for speech. *Cognitive Science 20*, 1–46.

PELACHAUD, C., AND POGGI, I. 2002. Subtleties of facial expressions in embodied agents. *Journal of Visualization and Computer Animation 13*, 301–312.

PERLIN, K., AND GOLDBERG, A. 1996. Improv: a system for interactive actors in virtual worlds. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1996)*, 205–216.

PIERREHUMBERT, J., AND HIRSCHBERG, J. 1990. The Meaning of Intonational Contours in the Interpretation of Discourse. In *Intentions in Communication*, P. Cohen, J. Morgan, and M. Pollack, Eds. MIT Press, Cambridge MA, 271–311.

PIWEK, P., KRENN, B., SCHRÖDER, M., GRICE, M., BAUMANN, S., AND PIRKER, H. 2002. RRL: A Rich Representation Language for the Description of Agent Behavior in NECA. In *AAMAS Workshop: Embodied Conversational Agents—Let's specify and evaluate them!*

POESIO, M., AND TRAUM, D. R. 1997. Conversational actions and discourse situations. *Computational Intelligence 13*, 309–347.

POGGI, I., AND PELACHAUD, C. 2000. Eye communication in a conversational 3D synthetic agent. *AI Communications 13*, 169–181.

POLLACK, M. E. 1990. Plans as complex mental attitudes. In *Intentions in Communication*, P. Cohen, J. Morgan, and M. Pollack, Eds. MIT Press, Cambridge MA, 77–103.

POWER, R. 1977. The Organisation of Purposeful Dialogues. *Linguistics 17*, 107–152.

PREECE, A. 1992. Collaborators and critics: The nature and effects of peer interaction on children's conversational narratives. *Journal of Narrative and Life History 2*, 277–292.

PRINCE, E. F. 1992. The ZPG Letter: Subjects, Definiteness and Information Status. In *Discourse Description: Diverse Analyses of a Fund-raising Text*, W. C. Mann and S. A. Thompson, Eds. John Benjamins, Philadelphia, 295–325.

PURVER, M. 2004. *The Theory and Use of Clarification Requests in Dialogue*. PhD thesis, Univ. of London.

RICH, C., SIDNER, C. L., AND LESH, N. 2001. COLLAGEN: Applying collaborative discourse theory to human–computer interaction. *AI Magazine 22*, 15–25.

RICKEL, J., AND JOHNSON, W. L. 1999. Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition and Motor Control. *Applied Artificial Intelligence 13*, 343–382.

RICKEL, J., MARSELLA, S., GRATCH, J., HILL, R., TRAUM, D., AND SWARTOUT, W. 2002. Toward a new generation of Virtual Humans for Interactive Experiences. *IEEE Intelligent Systems 17*, 32–38.

SACKS, H., SCHEGLOFF, E. A., AND JEFFERSON, G. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language 50*, 696–735.

SCHEGLOFF, E. A., AND SACKS, H. 1973. Opening up closings. *Semiotica 8*, 289–327.

SENGERS, P. 1999. Designing comprehensible agents. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 1999)*, 1227–1232.

SIDNER, C. L., KIDD, C. D., LEE, C., AND LESH, N. 2004. Where to look: a study of human-robot interaction. In *Proceedings of the International Conference on Intelligent User Interfaces*, 78–84.

SILVERMAN, K. E. A., BECKMAN, M., PITRELLI, J. F., OSTENDORF, M., WIGHTMAN, C., PRICE, P., AND PIERREHUMBERT, J. 1992. ToBI: A standard for labeling English prosody. In *Proceedings of ICSLP*, 867–870.

STEEDMAN, M. 2000. Information structure and the syntax-phonology interface. *Linguistic Inquiry 31*, 649–689.

STENT, A. J. 2002. A conversation acts model for generating spoken dialogue contributions. *Computer Speech and Language 16*, 313–352.

STONE, M., DORAN, C., WEBBER, B., BLEAM, T., AND PALMER, M. 2003. Microplanning with communicative intentions: The SPUD system. *Computational Intelligence 19*, 311–381.

STONE, M. 2004. Communicative Intentions and Conversational Processes. In *Approaches to Studying World-Situated Language Use*, J. Trueswell and M. K. Tanenhaus, Eds. MIT, 39–70.

STONE, M. 2004. Intention, interpretation and the computational structure of language. *Cognitive Science 28*, 781–809.

SVENNEVIG, J. 1999. *Getting Acquainted in Conversation*. John Benjamins, Philadelphia.

TALMY, L. 2000. *Toward a Cognitive Semantics*, vol. I and II. MIT Press, Cambridge.

THEUNE, M., KLABBERS, E., ODIJK, J., DE PIJPER, J. R., AND KRAHMER, E. 2001. From Data to Speech: A General Approach. *Natural Language Engineering 7*, 47–86.

THEUNE, M., HEYLEN, D., AND NIJHOLT, A. 2005. Generating embodied information presentations. In *Multimodal Intelligent Information Presentation*, O. Stock and M. Zancanaro, Eds. Kluwer, Dordrecht, 47–70.

THÓRISSON, K. R. 1997. Gandalf: An embodied humanoid capable of real-time multimodal dialogue with people. In *Proceedings of the International Conference on Autonomous Agents (Agents 1997)*, 536–537.

TICKLE-DEGNEN, L., AND ROSENTHAL, R. 1990. The nature of rapport and its nonverbal correlates. *Psychological Inquiry 1*, 285–293.

TRAUM, D. R. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, Department of Computer Science, University of Rochester. Also available as TR 545, Department of Computer Science, University of Rochester.

TRAUM, D. R., AND ALLEN, J. F. 1994. Discourse Obligations in Dialogue Processing. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 1994)*, 1–8.

TRAUM, D. R., AND RICKEL, J. 2002. Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the International Joint conference on Autonomous Agents and Multiagent systems (AAMAS 2002)*, 766–773.

TRAUM, D., RICKEL, J., GRATCH, J., AND MARSELLA, S. 2003. Negotiation over tasks in hybrid human-agent teams for simulation-based training. In *Proceedings of the International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2003)*, 441–448.

TRINDI CONSORTIUM. 2001. The Trindi Book. Tech. rep., Goteborg University. http://www.ling.gu.se/projekt/trindi/book.ps.

VILHJÁLMSSON, H., AND CASSELL, J. 1998. BodyChat: Autonomous communicative behaviors in avatars. In *Proceedings of the International Conference on Autonomous Agents (Agents 1998)*, 269–276.

WAHLSTER, W., Ed. 2000. *Verbmobil: Foundations of Speech-to-speech Translation*. Springer, Berlin.

WAHLSTER, W., REITHINGER, N., AND BLOCHER, A. 2001. SmartKom: Multimodal Communication with a Life-Like Character. In *Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH 2001)*, vol. 3, 1547–1550.

WALKER, M. A., AND WHITTAKER, S. 1990. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 1990)*, 70–78.

WALKER, M. A., LITMAN, D. J., KAMM, C. A., AND ABELLA, A. 1997. Paradise: a Framework for Evaluating Spoken Dialogue Agents. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics and Conference of the European Chapter of the Association for Computational Linguistics (ACL 1997)*, 271–280.

WALKER, M. A., LITMAN, D. J., KAMM, C. A., AND ABELLA, A. 1998. Evaluating Spoken Dialogue Agents with PARADISE: Two case studies. *Computer Speech and Language 12*, 317–347.