Computational Approaches to Dialogue

David Traum

Institute for Creative Technologies, University of Southern California
12015 Waterfront Drive, Playa Vista, CA 90292 USA

1 Introduction

Computational approaches to dialogue generally fall into two categories of computational task: *Dialogue Modeling* and *Dialogue Management*. *Dialogue modeling* refers to developing and updating an explicit representational model of the dialogue, with a purpose of inferring other information about the participants, the content or the process of the dialogue. Dialogue modeling takes the perspective of an observer of the dialogue, but not necessarily a participant. *Dialogue management* refers to the process of deciding what to do next in dialogue, and thus takes the perspective of a participant in dialogue, not just an observer. Generally a *dialogue system* will have both dialogue modeling and dialogue management components.

There are some computational systems that include only dialogue modeling, and either analyze a dialogue ex post facto, or assist one or more dialogue participants, without the system making any decisions about dialogue participation. Some examples of roles for computational systems that do dialogue modeling only are:

speech to speech translation: systems such as Verbmobil (Kay et al. 1994) do not act as an autonomous participant in dialogue, but serve as an intermediary between two people who speak different languages. The system translates the utterances of one to the language of the other, and in order to accomplish this task, maintains a model of the ongoing dialogue (Alexandersson et al. 1998), so that this dialogue model can assist in reducing language understanding perplexity, as well as use of references to prior linguistic context, such as inter-sentential anaphora or ellipsis.

meeting summarization: people have tried to model multi-party dialogues such as meeting transcripts or chat logs (Tur et al. 2010, Elsner & Charniak 2008) in order to detect things such as action items (Murray & Renals 2008) and decisions in meetings (Fernández et al. 2008)

medical corpus analysis: Dialogue modeling is also done by systems that analyze corpora of dialoge interaction in order to detect medical or psychological conditions, such as mild cognitive impairment (Lehr et al. 2012), Schizophrenia (Howes et al. 2012), depression and PTSD (DeVault et al. 2013).

In the rest of this chapter we will focus on dialogue systems that include a dialogue management process, and can engage in dialogue. These systems may require profoundly different dialogue models from the above examples, since they have to be adequate not just for understanding (some aspect of) dialogue, but also for providing information sufficient to decide what to say. Another big difference is that because the system takes a role in the dialogue, it can shape the contributions of other participants as well, explicitly signaling when it has problems of understanding, or taking initiative to direct the kinds of contributions of others. Thus human-computer dialogue can often be much easier for a system to model than human-human dialogue.

In the next section, we introduce human computer dialogue, look at several orientations towards systems that can engage in dialogue, and several roles for computer dialogue Systems. In Section 3 we review several *organizing principles* for dialogue systems, including what sort of *information state* they represent, and how they decide on next moves to make in dialogue. We follow this in section 4 with a discussion of some current challenge issues for dialogue systems, and conclude in Section 5.

2 Human Computer Dialogue

Currently, computer systems are not able to engage in fully human dialogue, communicating using the full range of human expressions and capable of comprehending and intending the full range of meanings and purposes that such communication is put to. Indeed, some doubt that computers could ever engage in "real dialogue". On the other hand, there are certainly many similarities between human dialogue and some forms of human-machine interaction, such as purposeful use and understanding of natural language expressions and interactive patterns. Examining the types of formal and computational models that are used to model and manage computer dialogue systems can be instructive as to both how much can be captured with simplifications as well as what is missing. A general strategy for most of these systems is to focus on only one or a small set of specific genres, roles and activities for dialogue interaction. The goals of may of these systems are thus much more modest than trying to handle all types of dialogue interaction, but rather capturing *some* of them well enough to be useful or interesting.

There are two broad types of systems that participate in dialogue, differentiated on how their designers see their function. Larsson calls these two types *simulations* and *interfaces* (Larsson 2005). *Interfaces* function fairly similarly to the translation systems described above, the difference being that they are not translating between two humans using two human languages but rather between one human speaking a natural language, and a *back-end* system that communicates according to a computer language or *Application Program interface* (API). In this case, the back end system will determine the content that should be expressed to the human, and the dialogue system will take care of the translation to human language, but also, importantly, including dialogue conventions of length of turns, and will also make some autonomous decisions about how best to convey the information to the user, and engage in various feedback (Allwood et al.

1992), repair (Schegloff et al. 1977), and grounding (Clark & Wilkes-Gibbs 1986) processes to establish reliable communication. Interfaces, or more generally, the whole system, including the back-end, are generally designed to perform some sort of service for a human user, such as providing information (Raux et al. 2005, Hajdinjak & Mihelic 2004, Nakano et al. 2011, Sadek et al. 1994) or completing a simple transaction (Walker et al. 1998, 2001), but allowing the user to speak in natural language rather than use a different kind of interface. These systems are seen as tools of a user, and fully subordinate to the user's goals (within the confines of the capabilities of the system). The goals of these systems are to accomplish the user's goal as efficiently and effectively as possible. It is seen as perfectly acceptable if these systems diverge from actual human-human dialogue patterns, as long as they perform these tasks effectively and efficiently, and satisfy their users.

On the other hand, the *simulation* type of systems, attempt to model human dialogue capability. They do not necessarily submit completely to the user's goals and desires (especially when it would be unnatural or unproductive to do so), and are judged not just on task efficiency, but on how human-like their behavior is. The ultimate expression of this kind of evaluation metric is the *Turing Test* (Turing 1950). This test posits that computers have reached human-level intelligence if an interlocutor, communicating through a text-based interface can not tell whether their partner is a human or computer with any more accuracy than they could tell whether it is a man or a woman. More modest evaluation schemes for this kind of system do not try to assess whether a computer dialogue participant is indistinguishable from a human, but rather whether their contributions are coherent and motivated (Traum et al. 2004) or whether it has certain qualitative abilities that are present in human communication (Bos et al. 1999). Simulation type systems include:

- those primarily aimed at passing the turing test (or the slightly more modest Loebner competition[1])
- systems to support role-play training of conversational and other human-interaction skills, such as mission rehearsal (Hill et al. 2003) , negotiation (Traum et al. 2005*a*), tactical questioning (Traum et al. 2007), leadership training (Hays et al. 2012), patients for medical (Stevens et al. 2006) and psychological (Kenny et al. 2007) patient interaction, and language learning (Johnson et al. 2004)
- systems to educate a user, acting as tutors (Graesser et al. 2005, Litman & Silliman 2004, Zinn et al. 2002) or experts (Swartout et al. 2013, Traum et al. 2015)
- systems designed to act as confederates in a psychology experiment designed to measure behavior in a specific kind of interactive context (Gratch et al. 2013).

One can also generalize from systems primarily intended to simulate human dialogue behavior to *intelligent agents* that intend to model human-like cognitive,

[1] http://www.loebner.net/Prizef/loebner-prize.html

intentional, and communicating entities, without specifically aiming at human emulation. These systems, like the *interface* systems, may prioritize effectiveness over human-likeness, but, more like the *simulation* systems, may put other goals ahead of user-intention satisfaction. For any given system, it can often be difficult to tell whether the goals of the designers are for it to be a tool/interface, or a simulation/agent, but these differences are often implicit in the design and evaluation methods for such systems.

There are a large number of types of applications for computer systems that can act as dialogue participants. Some of these line up well with the previous dichotomy of interface vs agent, while others may cut across either design. Some examples of applications for dialogue systems include:

information provider - the system can answer user queries about some domain, e.g. information in a database. (Meng et al. 1996, Raux et al. 2005)

instruction giver - the system can provide step-by-step instructions for a user to follow (Cassell et al. 2002)

service provider - the system can carry out a sequence of instructions to provide a service for the user, such as checking email (Walker et al. 1998) or booking an air flight (Walker et al. 2001)

advisor - the system can provide advice on courses of action that a user should take (Carberry et al. 1999)

collaborative partner - the system can work with one or more human participants to plan and/or complete a task. These situations can be described as either *symmetric* if the humans and computers can perform the same actions (Rich et al. 2001) or *asymmetric* if there are complementary actions that can be performed by human and system, for example human sensing the world, and system providing expert knowledge (Smith & Hipp 1994).

tutor - the system doesn't instruct the user how to do things, but rather helps the user learn some material and/or how to think about and solve a class of problems.

competitor - the system acts as an opponent in a game, or competitor for resources in a negotiation.

These different points of view of dialogue systems' main function: interface or agent, as well as the different types of role for the system, as exemplified above, do much to characterize the types of interactions that the systems can engage in, as well as evaluation metrics for determining whether they are successful.

3 Dialogue Manager Organizing Principles

In this section, we consider the decision-making process for a dialogue manager, and how it decides what kinds of moves to make in a dialogue. Key for this process is also the dialogue modeling component used to inform these decisions. We consider several approaches, including those based on modeling the structure of a dialogue, and characterizing legal or desirable end-states, as well as those based more on principles to apply, seeing the structure of dialogue as a collective achievement, rather than a goal in its own right.

3.1 Structure-based Approaches

Structure-based approaches to dialogue management explicitly model some notion of dialogue structure and use that model to decide on specific moves that will create valid or desirable structures. Perhaps the simplest example is a "ritual" or "script", specifying the dialogue contributions as an ordered list of actions by actors playing different roles. Dialogue systems can use scripts to decide what to do next based on observation of what has just happened and the script specifying the next action in the sequence. This kind of formulation may be adequate for very simple kinds of interactions, like "knock knock jokes". However, these systems are very limited, because they have no way to respond intelligently if their interlocutors go "off-script".

Other structural accounts specify constraints on future actions but provide some flexibility as to the specific contributions. We can further classify structure-based approaches into two groups - those that focus on local structure and local coherence between utterances, and those that focus on the structure of a whole dialogue and whether it includes all of the required elements in a proper order, essentially forming a grammar of a dialogue. We examine each of these in turn.

3.1.1 Local Structure Most local coherence accounts are based on *adjacency pairs* (Schegloff & Sacks 1973), in which dialogue units are composed of two parts, one for each speaker, and in which expectations for the second part are set up by recognition of the first part. Examples are *Question-Answer*, or *Offer-Acceptance*. Often there is a choice of type of second part, including preferred or dispreferred responses, such as rejecting an offer or (the legitimacy of) a question. Other names for this kind of local structure unit include Exchanges (Sinclair & Coulthard 1975) and Initiative-Response (IR) units (Dahlbäck & Jönsson 1998). User-initiative systems often use this kind of structure, in which the system responds to a user initiative following a policy to generate an appropriate response. Information-retrieval based approaches (Chu-Carroll & Carpenter 1999, Leuski et al. 2006) use information from the initiative to select from a set of possible responses, using relationships between the words in the initiative and the possible responses.

IR units can also be used by system-initiative systems, in which a series of questions or other prompts are given to a user to respond to. These systems simplify the Natural Language Understanding (NLU) task, since the system is concerned with understanding only those answers that make sense given the context of the system initiative, rather than the full set of possibilities.

There are also some extensions to the dyadic IR structure. First, some systems allow *mixed initiative* in which either system or user can start an IR unit. Secondly, recursive structures can be allowed, in which one IR unit can embed another, for example a question is initially responded to by a clarification question and the answer to the original question occurring only after receiving the clarification response. Turns can also consist of more than one move, so that an "RI" turn can be introduced that both responds to one initiative and also starts a new one for the original speaker to respond to. Eliza one of the first computer dialogue systems (Weizenbaum 1966), used pattern-matching to construct responses from input initiatives, using some syntactic replacements from the initiative in some of the responses. Many system turns were RI units that generated follow-up questions to part of the user's utterance. If there is no matching pattern, the system can also go back to earlier utterances. Eliza, and similar *chatbot* systems, such as ALICE (Wallace 2009) allow conversational interaction that follows many of the rules of local sequential dialogue structure, but, despite minimal use of context to refer back to previous topics, are generally not able to carry on more globally coherent interactions.

Some have also looked at local exchange structures that can include more than two moves. (Sinclair & Coulthard 1975) introduce an IRF unit, including also a feedback move, which is particularly common for instructional dialogue, in which this move can contain an evaluation of the student's response, but more generally can also include feedback on contact, perception, understanding, and attitudinal reaction (Allwood et al. 1992). The key distinction between serial IR structures and an IRF structure is that the feedback move in an IRF structure does not require a response or feedback (while the initial response does require further feedback. A similar structure was proposed by (Carletta et al. 1996), who describe *dialogue games* as a local structure involving a variable amount of individual dialogue moves to establish the intention behind an initiating act.

These local structure-based accounts can capture some of the reactive aspects of dialogue, but don't have much to say about why particular local structures were chosen or how to transition from one to another.

3.1.2 Global and grammar-based structural approaches In addition to local dialogue structure, an important set of concerns involve the high-level structure of dialogue: how it begins and ends, why it is undertaken, how do topics flow logically from one to another to accomplish these purposes? Many system designers approach this problem using *dialogue grammars* or equivalent *finite-state networks* that define legal dialogues (e.g., Sutton et al. (1996)). These formalisms allow system designers to specify the set of legal and preferred moves at any point in the dialogue, and insure that all necessary actions occur and that actions occur in a prescribed order. These kinds of systems are good for system-initiative dialogue systems, where a fairly simple conversation structure is allowed. They are less well suited when there is a large number of options as to what can be said or the order in which parts can fit into the whole, because these options generate much larger sets of possible dialogue states, and also, the states provide less context to guide the NLU component or the decision-making component of the dialogue manager.

The structural approaches mentioned above are able to represent some aspects of dialogue context that can be used for further interpretation of utterances, dialogue modeling, and dialogue management decisions. One problem is that while these approaches can capture prescriptive notions of local and global dialogue coherence, they don't express or allow a system to reason about *why* these guidelines exist, the consequences of not following the guidelines, or how severe different kinds of divergence might be (e.g. uncooperative vs incoherent dialogue). Moreover, it can be very difficult to efficiently capture regularities in behavioral patterns that lead to similar, but not identical structures.

3.2 Principle-based approaches

Principle-based approaches to dialogue modeling focus not so much on the sequential relationship of dialogue contributions, but on the impact that contributions have on the interactive dialogue context and how aspects of the context make different kinds of contributions conditionally relevant. Some of these approaches allow the system to reason about both local and global coherence of a dialogue, but also often give more indications about the nature and relative cost of violations. In this section, we review a few popular methods of framing dialogue context for dialogue management decisions.

3.2.1 Frame Tree and finite-state models are suitable for modeling the logical progression of a dialogue when each piece of information provided leads to a logical next initiative, however these models are less felicitous when what is important is just the set of information received, rather than the order in which it has been received. For example, a dialogue system that looks up information from a database may require a set of parameters in order to formulate an appropriate query. For example, in trying to locate an appropriate air flight, a system might need to know the user's departure and arrival cities and date of travel. However it does not matter which order these are presented in, or whether they all are provided as responses to individual queries or offered unsolicited in a single sentence, or any combination of methods. A *frame* or *form* (Goddeau et al. 1996) model of context keeps track of the relevant set of "slots" for a query, and whether they have already been "filled" by user responses or are still pending in dialogue. This structure can be used by the NLU component to recognize potential fillers in user utterances, and can also be used by the dialogue manager to constrain the set of relevant next utterances. To contrast with the finite-state approach, if the finite-state model dictates first asking about the departure and then the destination, it might still ask the latter question even after a response of "from Los Angeles to Miami". This could be avoided by explicitly keeping track of separate state for all possible sets of information provided at each state, but this quickly gets unwieldy. On the other hand, the frame-approach simply updates two rather than one slot after the response, and both queries are removed from the set of pending slots that generate queries.

3.2.2 Logic and Plan-based Approaches Another kind of principle used to drive dialogue management decisions is logic and practical reasoning, to either deduce new propositional attitudes (Hintikka 1971) or actions for the system to do. The circuit-fixit shop (Smith & Hipp 1994, Smith et al. 1995) can instruct a user in how to assemble circuits from a circuit-board kit. It can tutor the user about how to perform actions or names of pieces, as well as given instructions to complete a circuit. The main method for engaging in dialogue is the *missing axiom theory*, in which the system attempts to prove that the circuit is working – a successful proof leads to dialogue termination, but inability to complete the proof because of "missing" axioms leads to dialogue contributions to establish support for the proof. These contributions can be questions about unknown state of the circuit, or directions to do physical actions on circuit components.

The Artimis system (Bretier & Sadek 1996) is based on a theory of rational balance, including basic attitudes of belief, desire, and intention, and rationality principles, in which communication is seen as a special case of rational interaction to align the attitudes (e.g. adopting intentions to perform actions that will satisfy desires).

Plan-based approaches are similar to logic-based approaches, however the focus is more on means-end reasoning than general inference. The roots of plan-based approaches are in Speech Act theory (Austin 1962, Searle 1969), formalized using AI planning systems (Allen et al. 1990), which are able to reason back from goal states to the initial state, using planning operators. Planning operators establish effects and can be applied if their pre-conditions hold. If not, other actions can be selected to be performed first, whose effects establish the needed pre-conditions. Speech acts are defined as actions, using planning operators (Cohen 1978, Cohen & Perrault 1979), with pre-conditions and effects related to beliefs, and desires of participants. Using this framework, systems can form a complex plan, involving sequences of speech acts to achieve dialogue goals. Unlike the structural approaches and even the frame-based approaches, plan-based approaches can be more dynamic and responsive to changes in conditions, so that a new plan can be formulated whenever the system notices other changes, such as information being provided or superfluous to the main goal. Moreover, plan-based accounts of speech acts can also be used in reverse to understand user intentions behind utterances (Allen & Perrault 1980, Allen 1983), and facilitate cooperative behavior that addresses the user's presumed deeper goal rather than surface obligations.

There have also been several attempts to model joint attitudes between teams of agents working on the same task, or conversing, or both. Cohen and Levesque (Cohen & Levesque 1990, 1991) develop an account of Joint action based on primitives of mutual belief and intention, that leads to individual intentions to keep coordinated. Grosz and Sidner (Grosz & Sidner 1990) develop a similar model of *Shared Plans*, that has been the foundation of a series of dialogue systems using the *Collagen* architecture (Rich et al. 2001).

3.2.3 Information State The *Information State Approach* (Traum & Larsson 2003) is an attempt to abstract away from the specific structures and principles used for guiding dialogue management decisions, and cast them in a way where they can be more easily combined and compared. For example, many dialogue systems need to deal with the coherence between question and answer, and why a system should answer a user question. Structural approaches focus on the sequential connection between question and answer, while principle-based approaches focus on how the occurrence of the question creates a motivation for the answer, however as described above, there can be many ways of representing the structure or the motivations. The information state approach consists of five abstract components to model developing dialogue context. **Informational components** are a functional specification of the aspects of context (e.g. structures, principles) that are part of the dialogue model. **Formal representations** are data structures, including accessibility relations for each of the informational components. For example, a history of previous utterances might be represented as a simple list, while a set of topics might be represented as a stack (Grosz & Sidner 1986). Informational components and formal representations together model the statics of dialogue context at any moment in the dialogue. The dynamics are modelled by three additional components. First, **Dialogue Moves** are abstractions of the kinds of utterance information that is seen as relevant for updates. This could be the same sort of speech acts or dialogue acts as used by many approaches (e.g., (Cooper & Larsson 1999, Lewin 1998), but also more nuanced information, including the contributions of individual words to an incremental discourse representation (Poesio & Traum 1995, Poesio 1995, Poesio & Traum 1997, 1998). **Update rules** specify how the informational components change as a result of observation and processing of dialogue moves. For example, once a question has been answered, the question is no longer a motivating factor requiring an answer. Finally, an **Update Strategy** is a method for deciding which rules to apply when. In addition to these dialogue modeling components, there are parellel methods for selecting system utterances for dialogue management. **Selection Rules** specify when a dialogue move could be performed, given the current information state, while a *selection algorithm* decides how to prioritize the selection rules.

3.2.4 Statistical Reward Models With the increasing amount of linguistic dialogue corpus data available, as well as increasing pools of users for systems, statistical approaches toward dialogue modeling and management have become more popular. These models generally represent dialogue as a Markov decision Process (MDP) (Levin et al. 2000), where the resulting state after an action follows a probability distribution, or partially observable Markov decision process (POMDP) (Williams & Young 2007), where the system's belief about the current state is also a probability distribution, rather than being known. Reinforcement Learning can be used to optimize a policy for action given the current state or belief state, by using a reward function. In this way, the "rules" used for dialogue model (or belief state) updating and dialogue policy selection can be learned from dialogue data so as to maximize expected reward. Often there is not enough existing data, so systems learn from interactions with users or simulated users (Georgila et al. 2006). Initially, these systems adapted an extension of the structural approach, with transitions between absolute dialogue states. More recently, however, these systems have used an information state, involving a set of features that are either present or absent in a state, rather than a full model of all possible states.

4 Issues for Dialogue Systems

In this section, we briefly summarize some important issues for dialogue systems, particularly those attempting to model dialogue that is similar to human-human dialogue. Many systems attempt to simplify or eliminate these issues by constraining the types of interactions, number of users, and dialogue activities that the systems can participate in, however this is usually at the cost of naturalness and flexibility of normal human-human dialogue.

4.1 Grounding, Feedback and Repair

Grounding (Clark & Wilkes-Gibbs 1986, Clark & Schaefer 1989) is the process by which participants in a conversation establish new common ground. Many early computational accounts and systems assume that mutual belief results just from performing an utterance in a common, mutually perceptible space. By contrast, studies of human dialogue show that there is often explicit feedback (Allwood et al. 1992) given, indicating the degree and sometimes content of what has been understood. When a (perceived or assumed) lack of common ground exists, then participants often take the trouble to *repair* the (potential) misunderstanding (Schegloff et al. 1977). Grounding is thus a *Collaborative Process* (Clark & Wilkes-Gibbs 1986), in which the participants of a conversation jointly work to establish common ground, rather than separately performing tasks of encoding and decoding meaning to/from language.

Grounding can be even more of a difficult issue for human-computer dialogue than for human-human dialogue, because of difficulties in speech recognition, language interpretation, common ontologies, and common-sense reasoning for computer dialogue systems.

Many systems include explicit dialogue acts and sequences or principles to deal with grounding phenomena. For example, consider the system-initiated question-answer exchange in (1), utterances 1 and 2. Each of the possible system follow up moves in 3 is plausible, but have a different role in terms of grounding.

(1) 1 **System:** Where do you want to go?
2 **User:** Boston.
3 **System:**
  a. Tell me more about your travel plans.
  b. When would you like to go?
  c. When would you like to go to Boston?
  d. ok
  e. Boston.
  f. Boston?
  g. Do you want to go to Boston?

h. Did you say Boston?
i. Boston or Austin?
j. Where?
k. Please repeat that.

Responses 3a-c all indicate a desire to move forward with the dialogue, moving from the filled slot of destination to an open slot (e.g. time of departure). Response 3a gives initiative back to the user to select which slot to talk about next, while 3b and 3c specify the departure slot. Responses 3a and 3b do not give any explicit signal of what was understood, however they implicitly provide evidence of understanding by passing up an opportunity to request repair and continuing with a next relevant contribution (Clark & Schaefer 1989). Response 3c both indicates that “Boston” was understood as the destination, as well as moving along to the departure time slot. Responses 3d and 3e acknowledge understanding of what was said, without either requiring further feedback or confirmation but also without moving on. It is thus unclear whether the next turn should be by system or user. 3d, like 3a,b just claims understanding (in this case more explicitly), while 3e displays that understanding. Utterances 3c,3e-i all give some evidence of having understood “Boston” as the destination. 3c is what is often called (e.g., (Möller 2005)) an *implicit confirmation*, contrasted with 3f-h, which are termed *explicit confirmation*. Even though both kinds of utterance are equally explicit about what was heard, the difference is that “implicit confirmations” have as a main purpose, something other than confirmation, while the main purpose of an “explicit confirmation” is to confirm the system’s expressed understanding. Implicit confirmations can be more efficient, if understanding is correct, however they can be harder to notice or produce a correction if needed. 3f-h are all “explicit confirmations”, however they differ in terms of what aspect they are confirming. 3g focuses on “Boston” as being the destination, while 3h focuses on the surface expression rather than the meaning. 3f is ambiguous between whether it is confirming what was heard or meant, though is equally vague about the slot name. 3i-k indicate less confidence in having understood than the previous alternatives. 3i presents two options, and might be preferred if both are likely, given the overall context and what was heard. This also makes it more likely than previous utterances that an error will not go unnoticed. 3j and 3k are both more appropriate when it is more likely that the system did not understand correctly, and a confirmation would be more likely wrong and confusing than helpful. Like the contrast with 3g and 3h, 3j indicates the slot name that is at issue, while 3k focuses on the surface level of what was said rather than what was meant or the context in which it is to be understood.

Systems may produce all or only a subset of these response types, as well as other variations. Any of the control mechanisms from Section 2 could be used to decide on a policy of when to use each response type. For example, a reinforcement learning model might use features such as the current confidence level of the interpretation, confidence levels for the next highest interpretation, user style preferences, and number of attempts made to elicit this information, and learn the ideal threshold values for deciding on one move or another, in terms of dialogue length and user satisfaction.

Some systems go beyond simple confirmation and repair strategies, and model the state of grounding for different aspects of information. For example, the systems described in (Allen & Schubert 1991, Matheson et al. 2000, Traum & Rickel 2002, Traum et al. 2005*a*), all use versions of the speech acts model of grounding of (Traum & Allen 1992, Traum 1994), which makes use of local structures for each "discourse unit" as well as principles for deciding on next moves. An extended grounding model, that takes into account degrees of grounding and different patterns of grounding was presented by (Roque & Traum 2008), and used in several systems (Roque et al. 2006, Roque & Traum 2009).

4.2 Non-cooperative Dialogue

Many accounts of meaning in dialogue rely on cooperatively principles (Grice 1975). This is not a problem for the majority of dialogue systems that play a role that is either a tool for the human user to accomplish her goals, or at least a collaborative partner who shares goals with the user. However many dialogue roles are not purely cooperative: goals fail to align exactly, or may even be diametrically opposed. In a commercial transaction, both parties may have commonality of goals in coming to a mutually satisfactory deal, but the buyer will want the price to be as low as possible, while the seller would like the price as high as possible. Even in cases in which overall goals converge, such as for a tutoring system, local goals may still be such that ideal systems would not be perfectly cooperative. For example, a student might ask for the answer to a difficult problem, while the tutor might think it would be better for the student to figure it out himself rather than be given the full answer.

Dialogue systems that play a non-cooperative role, such as in a negotiation/commercial transaction, tutoring/coaching session, or zero-sum game, must make some adjustments compared to tool and helper systems. Understanding and complying with user intent is not sufficient, and in some cases not necessary, for deciding on system behaviors. Other mechanisms, such as discourse obligations (Traum & Allen 1994), are needed to motivate coherence, in the absence of adopting the goals of a collaborative partner. A certain amount of deliberation is also needed to optimally balance concerns of the user with the systems own goals, to decide when to choose one over the other.

The PRACMA system (Jameson et al. 1994, Jameson & Weis 1995) was one of the first to explore principle-based non-cooperative agents, in a sales context. (Traum 2012) summaries some of the issues involved, and briefly describes systems for training people in negotiation (Traum et al. 2005*b*, Plüss et al. 2011) and Tactical questioning (Roque & Traum 2007, Gandhe et al. 2008). Other recent work on dialogue models for negotiation and non-cooperative dialogue have been presented by (Guhe & Lascarides 2012, Asher & Lascarides 2013, Efstathiou & Lemon 2014).

4.3 Multi-modal Dialogue

While most dialogue systems focus only on text or speech, natural dialogue has a lot of information coming through other channels as well. Multi-modal dialogue systems allow one to communicate also through other channels. Some systems that are on a computer, or other device with a screen allow the system to display graphical or tabular information. Some of these systems also allow users to use pointer input as well as speech or text. For example (Johnston et al. 2002) allows users to select a region on a map to ask about. Other systems allow more human-like natural gestures in communication. These include complementary deictic gestures that point out an object or location, as well as gestures for emphasis, iconic gestures, and signals of addressee selection, turn-taking, and various kinds of feedback.

(Cassell et al. 2000) present *embodied conversational agents* that give the system an animated body as well as speech to express conversational signals. These systems are also called *Virtual humans* (Gratch et al. 2002). Some researchers also focus on dialogue with robots (Carl Burke & Loehr 2002, Matsusaka 2008), that have physical bodies, and in some cases the ability to move around the environment, change points of view and perspective. There are currently several toolkits that allow people to associate specific physical behaviors with communicative intent (Cassell et al. 2001, Lee & Marsella 2006), and which can automatically apply rules to select these behavior.

There are also active research efforts on understanding communicative behaviors from the visual or multimodal channels, and integrating these channels in dialogue models (Traum & Morency 2010). Often using multiple channels can provide not just more and different kinds of information, but better accuracy or confidence in information that is observable in different modalities. For instance, visual lip reading could enhance speech recognition, or using gaze as well as lexical information could more accurately allow recognition of addressees. In this case, there is a need for *multimodal fusion* of independent input signals. *Late fusion* is more common, in which recognition is made via each modality, and then the separate pieces of information are merged into a unified whole. *Early Fusion* means that the raw signals or intermediate features from different modalities are combined in order to select a joint hypothesis.

4.4 Multi-party Dialogue

Most dialogue systems look only at dyadic interaction, primarily between the system and one human user. However much of human dialogue occurs in a context of small group interaction involving more than two entities. Multiparty conversation brings additional issues that must be modelled. First, there are multiple conversation participant roles, rather than just speaker and hearer (Goffman 1981, Levinson 1987). Addressee identification (Martinovski et al. 2003, Jovanovic & op den Akker 2004, op den Akker & Traum 2009) is an important problem, with more than one non-speaker present, so the system can decide whether or not it must respond. (Traum 2004) discusses how other issues such as turn-taking, grounding, interaction management and obligations are impacted by multiple participants.

Several systems simulate multiparty interaction (Padilha & Carletta 2002, Patel et al. 2004) including variations for different cultures (Jan et al. 2007). Other systems allow interactions with multiple people and an artificial system (Matsusaka 2008) or one user and multiple agents (Traum et al. 2008, Plüss et al. 2011). (Bohus & Horvitz 2009) also consider the case of multiple users who may have side conversations, and how to detect engagement in open-world dialogue, not constrained to a laboratory experiment or single task.

5 Conclusions

In this paper we have briefly surveyed some of the diversity of approaches to computational modeling dialogue and constructing systems that can engage in dialogue with users. While they all have in common a certain degree of formalization, in order to allow computers to process dialogue, they all equally fall short of full expression of the range of mechanisms and uses to which human dialogue is put. Some approaches tend more toward engineering solutions to engage in dialogue for a specific task, while other approaches try to capture more of the essence of human-dialogue skills, including some of the ways that dialogue is inextricably connected with cognitive and social skills. What is common in dialogue systems is a focus on a model of dialogue that is not strictly language understanding coupled with language production, but each in support of units containing input from the self and other entities, that must be linked for coherent dialogue interaction.

The formalization creates a kind of linguistic competence that, when coupled with the interactions with real people, creates a kind of "competence-in-performance" (Weigand (2006), in which the systems must cope with novel interactive behaviors. At present, these skills are extremely limited, much of it occurring through human analysis of data and updating of models, rather than the system developing new understanding and capabilities through dialogue use. However, we see rapid growth in the area, not just in fundamental algorithms and vastly increased amounts of data, but also in the more widespread use of such systems and diversity and complexity of the interactions that are being achieved.

Acknowledgements

This work was supported by the U.S. Army. Any opinion, content or information presented does not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

Bibliography

Alexandersson, J., Buschbeck-Wolf, B., Fujinami, T., Kipp, M., Koch, S., Maier, E., Reithinger, N., Schmitz, B. & Siegel, M. (1998), Dialogue Acts in VERBMOBIL-2 – Second Edition, Verbmobil-Report 226, DFKI Saarbrücken, Universität Stuttgart, Technische Universität Berlin, Universität des Saarlandes.

Allen, J. F. (1983), Recognizing intentions from natural language utterances, *in* M. Brady & R. C. Berwick, eds, 'Computational Models of Discourse', MIT Press.

Allen, J. F., Hendler, J. & Tate, A., eds (1990), *Readings In Planning*, Morgan Kaufmann.

Allen, J. F. & Perrault, C. R. (1980), 'Analyzing intention in utterances', *Artificial Intelligence* **15**(3), 143–178.

Allen, J. F. & Schubert, L. K. (1991), The TRAINS project, TRAINS Technical Note 91-1, Computer Science Dept. University of Rochester.

Allwood, J., Nivre, J. & Ahlsen, E. (1992), 'On the semantics and pragmatics of linguistic feedback', *Journal of Semantics* **9**.

Asher, N. & Lascarides, A. (2013), 'Strategic conversation', *Semantics and Pragmatics* **6**, 2:1–62.

Austin, J. A. (1962), *How to Do Things with Words*, Harvard University Press.

Bohus, D. & Horvitz, E. (2009), Models for multiparty engagement in open-world dialog, *in* 'Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue', Association for Computational Linguistics, pp. 225–234.

Bos, J., Bohlin, P., Larsson, S., Lewin, I. & Matheson, C. (1999), Evaluation of the model with respect to restricted dialogue systems, Technical Report Deliverable D3.2, Trindi.

Bretier, P. & Sadek, M. D. (1996), A rational agent as the kernel of a cooperative spoken dialogue system: Implementing a logical theory of interaction, *in* J. P. Müller, M. J. Wooldridge & N. R. Jennings, eds, 'Intelligent Agents III — Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages (ATAL-96)', Lecture Notes in Artificial Intelligence, Springer-Verlag, Heidelberg.

Carberry, S., Chu-Carroll, J. & Elzer, S. (1999), 'Constructing and utilizing a model of user preferences in collaborative consultation dialogues', *Computational Intelligence* **15**, 185–217.

Carl Burke, L. H. & Loehr, D. (2002), A dialogue architecture for multimodal control of robots, *in* 'Proceedings of the International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems'.

Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G. & Anderson, A. (1996), HCRC dialogue structure coding manual., Technical Report 82, HCRC.

Cassell, J., Stocky, T., Bickmore, T., Gao, Y., Nakano, Y., Ryokai, K., Tversky, D., Vaucelle, C. & Vilhjálmsson, H. (2002), Mack: Media lab autonomous conversational kiosk, *in* 'Imagina'.

Cassell, J., Sullivan, J., Prevost, S. & Churchill, E., eds (2000), *Embodied Conversational Agents*, MIT Press, Cambridge, MA.

Cassell, J., Vilhjálmsson, H. & Bickmore, T. (2001), Beat: the behavior expression animation toolkit, *in* 'Proceedings of ACM SIGGRAPH', ACM Press, New York, pp. 477–486.

Chu-Carroll, J. & Carpenter, B. (1999), 'Vector-based natural language call routing', *Journal of Computational Linguistics* **25**(30), 361–388.

Clark, H. H. (1992), *Arenas of Language Use*, University of Chicago Press.

Clark, H. H. & Schaefer, E. F. (1989), 'Contributing to discourse', *Cognitive Science* **13**, 259–294.

Clark, H. H. & Wilkes-Gibbs, D. (1986), 'Referring as a collaborative process', *Cognition* **22**, 1–39. Also appears as Chapter 4 in Clark (1992).

Cohen, P. R. (1978), On Knowing What to Say: Planning Speech Acts, PhD thesis, University of Toronto. Reproduced as TR 118 Department of Computer Science, University of Toronto.

Cohen, P. R. & Levesque, H. J. (1990), Rational interaction as the basis for communication, *in* P. R. Cohen, J. Morgan & M. E. Pollack, eds, 'Intentions in Communication', MIT Press.

Cohen, P. R. & Levesque, H. J. (1991), 'Teamwork', *Nous* **35**.

Cohen, P. R. & Perrault, C. R. (1979), 'Elements of a plan-based theory of speech acts', *Cognitive Science* **3**(3), 177–212.

Cooper, R. & Larsson, S. (1999), Dialogue moves and information states, *in* H. Bunt & E. C. G. Thijsse, eds, 'Proceedings of the Third International Workshop on Computational Semantics'.

Dahlbäck, N. & Jönsson, A. (1998), A coding manual for the linköping dialogue model. unpublished manuscript.

DeVault, D., Georgila, K., Artstein, R., Morbini, F., Traum, D., Scherer, S., Rizzo, A. S. & Morency, L.-P. (2013), Verbal indicators of psychological distress in interactive dialogue with a virtual human, *in* 'Proceedings of SIGdial'.

Efstathiou, I. & Lemon, O. (2014), Learning non-cooperative dialogue behaviours, *in* 'Proc. of SIGDIAL'.

Elsner, M. & Charniak, E. (2008), You talking to me? a corpus and algorithm for conversation disentanglement, *in* 'Proceedings of ACL-08: HLT', Association for Computational Linguistics, Columbus, Ohio, pp. 834–842.
URL: *http://www.aclweb.org/anthology/P/P08/P08-1095*

Fernández, R., Frampton, M., Ehlen, P., Purver, M. & Peters, S. (2008), Modelling and detecting decisions in multi-party dialogue, *in* 'Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue', SIGdial '08, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 156–163.
URL: *http://dl.acm.org/citation.cfm?id=1622064.1622095*

Gandhe, S., DeVault, D., Roque, A., Martinovski, B., Artstein, R., Leuski, A., Gerten, J. & Traum, D. (2008), From domain specification to virtual humans: An integrated approach to authoring tactical questioning characters, *in* 'Proceedings of Interspeech Conference'.

Georgila, K., Henderson, J. & Lemon, O. (2006), User simulation for spoken dialogue systems: Learning and evaluation, *in* 'Interspeech'.

Goddeau, D., Meng, H., Polifroni, J., Seneff, S. & Busayapongchai, S. (1996), 'A form-based dialogue manager for spoken language applications', *Proceedings of Fourth International Conference on Spoken Language Processing, ICSLP 96.* **2**, 701–704.

Goffman, E. (1981), Footing, *in* 'Forms of Talk', Philadelphia: University of Pennsylvania Press, pp. 124–159.

Graesser, A., Chipman, P., Haynes, B. & Olney, A. (2005), 'Autotutor: an intelligent tutoring system with mixed-initiative dialogue', *Education, IEEE Transactions on* **48**(4), 612–618.

Gratch, J., Hartholt, A., Dehghani, M. & Marsella, S. C. (2013), Virtual Humans: A New Toolkit for Cognitive Science Research, *in* 'Cognitive Science', Berlin, Germany.

Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E. & Badler, N. (2002), 'Creating interactive virtual humans: Some assembly required', *IEEE Intelligent Systems* pp. 54–63.

Grice, J. P. (1975), Logic and conversation, *in* P. Cole & J. L. Morgan, eds, 'Syntax and Semantics', Vol. 3: Speech Acts, Academic Press, pp. 41–58.

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

Grosz, B. J. & Sidner, C. L. (1990), Plans for discourse, *in* P. R. Cohen, J. Morgan & M. E. Pollack, eds, 'Intentions in Communication', MIT Press.

Guhe, M. & Lascarides, A. (2012), Trading in a multiplayer board game: Towards an analysis of non-cooperative dialogue, *in* 'Proc. of CogSci'.

Hajdinjak, M. & Mihelic, F. (2004), Information-providing dialogue management, *in* P. Sojka, I. Kopecek & K. Pala, eds, 'TSD', Vol. 3206 of *Lecture Notes in Computer Science*, Springer, pp. 595–602.

Hays, M., Campbell, J., Trimmer, M., Poore, J., Webb, A., Stark, C. & King, T. (2012), Can Role-Play with Virtual Humans Teach Interpersonal Skills?, *in* 'Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)', Orlando, FL.

Hill, Jr., R., Gratch, J., Marsella, S., Rickel, J., Swartout, W. & Traum, D. (2003), 'Virtual humans in the mission rehearsal exercise system', *In KI Embodied Conversational Agents* **17**, 32–38.

Hintikka, J. (1971), Semantics and propositional attitudes, *in* L. Linsky, ed., 'Reference and Modality', Oxford University Press, chapter X, pp. 145–167.

Howes, C., Purver, M., McCabe, R., Healey, P. G. T. & Lavelle, M. (2012), Predicting adherence to treatment for schizophrenia from dialogue transcripts, *in* 'Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue', Association for Computational Linguistics, Seoul, South Korea, pp. 79–83.
URL: *http://www.aclweb.org/anthology/W12-1610*

Jameson, A., Kipper, B., Ndiaye, A., Schäfer, R., Simons, J., Weis, T. & Zimmermann, D. (1994), Cooperating to be Noncooperative: The Dialog System PRACMA, *in* B. Nebel & L. Dreschler-Fischer, eds, 'Proceedings of the

Eighteenth German Conference on Artificial Intelligence (KI-94)', Springer, pp. 106–117.
URL: *citeseer.ist.psu.edu/jameson94cooperating.html*

Jameson, A. & Weis, T. (1995), How to juggle discourse obligations, *in* 'Proceedings of the symposium on Conceptual and Semantic Knowledge in Language Generation'.

Jan, D., Herrera, D., Martinovski, B., Novick, D. G. & Traum, D. R. (2007), A computational model of culture-specific conversational behavior, *in* Pelachaud et al. (2007), pp. 45–56.

Johnson, W. L., Beal, C. R., Fowles-Winkler, A., Lauper, U., Marsella, S., Narayanan, S., Papachristou, D. & Vilhjálmsson, H. H. (2004), Tactical language training system: An interim report, *in* J. C. Lester, R. M. Vicari & F. Paraguaçu, eds, 'Intelligent Tutoring Systems', Vol. 3220 of *Lecture Notes in Computer Science*, Springer, pp. 336–345.

Johnston, M., Bangalore, S., Vasireddy, G., Stent, A., Ehlen, P., Walker, M., Whittaker, S. & Maloor, P. (2002), Match: An architecture for multimodal dialogue systems, *in* 'Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)', pp. 376–383.

Jovanovic, N. & op den Akker, R. (2004), Towards automatic addressee identification in multi-party dialogues, *in* 'Proceedings of 5th SIGdial Workshop on Discourse and Dialogue', Cambridge, MA, USA.

Kay, M., Norvig, P. & Gawron, M. (1994), *Verbmobil: A Translation System for Face-to-Face Dialog*, University of Chicago Press, Chicago, IL, USA.

Kenny, P. G., Parsons, T. D., Gratch, J., Leuski, A. & Rizzo, A. A. (2007), Virtual patients for clinical therapist skills training, *in* Pelachaud et al. (2007), pp. 197–210.

Larsson, S. (2005), Dialogue systems: Simulations or interfaces?, *in* Gardent & Gaiffe, eds, 'Proceedings of the ninth workshop on the semantics and pragmatics of dialogue'.

Lee, J. & Marsella, S. (2006), Nonverbal behavior generator for embodied conversational agents, *in* 'Proc. 6th IVA'.

Lehr, M., Prud'hommeaux, E., Shafran, I. & Roark, B. (2012), Fully automated neuropsychological assessment for detecting mild cognitive impairment, *in* 'Interspeech 2012: 13th Annual Conference of the International Speech Communication Association', Portland, Oregon.

Leuski, A., Patel, R., Traum, D. & Kennedy, B. (2006), Building effective question answering characters, *in* 'Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue', pp. 18–27.

Levin, E., Pieraccini, R. & Eckert, W. (2000), A stochastic model of human-machine interaction for learning dialog strategies, *in* 'Proc. of ICASSP'.

Levinson, S. C. (1987), Putting linguistics on a proper footing: Explorations in goffman's participation framework, *in* P. Drew & A. Wootton, eds, 'Erving Goffman: Exploring the interaction order', Oxford: Polity Press, pp. 161–227.

Lewin, I. (1998), The autoroute dialogue demonstrator, Technical Report CRC-073, SRI Cambridge Computer Science Research Centre.

Litman, D. J. & Silliman, S. (2004), Itspoke: An intelligent tutoring spoken dialogue system, *in* 'Demonstration papers at HLT-NAACL 2004', Association for Computational Linguistics, pp. 5–8.

Martinovski, B., Traum, D., Robinson, S. & Garg, S. (2003), Functions and patterns of speaker and addressee identifications in distributed complex organizational tasks over radio, *in* 'Diabruck: seventh workshop on semantics and pragmatics of dialogue'.

Matheson, C., Poesio, M. & Traum, D. (2000), Modelling grounding and discourse obligations using update rules, *in* 'Proceedings of the First Conference of the North American Chapter of the Association for Computational Linguistics'.

Matsusaka, Y. (2008), History and current researches on building a human interface for humanoid robots, *in* I. Wachsmuth & G. Knoblich, eds, 'Modeling Communication with Robots and Virtual Humans', Springer, pp. 109–124.

Meng, H., Busayapongchai, S., Giass, J., Goddeau, D., Hethetingron, L., Hurley, E., Pao, C., Polifroni, J., Seneff, S. & Zue, V. (1996), Wheels: A conversational system in the automobile classifieds domain, *in* 'Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on', Vol. 1, IEEE, pp. 542–545.

Möller, S. (2005), *Quality of telephone-based spoken dialogue systems*, Springer Science & Business Media.

Murray, G. & Renals, S. (2008), Detecting action items in meetings, *in* A. Popescu-Belis & R. Stiefelhagen, eds, 'Machine Learning for Multimodal Interaction', Vol. 5237 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 208–213.

Nakano, M., Sato, S., Komatani, K., Matsuyama, K., Funakoshi, K. & Okuno, H. G. (2011), A two-stage domain selection framework for extensible multi-domain spoken dialogue systems, *in* 'Proceedings of the SIGDIAL 2011 Conference', Association for Computational Linguistics, Association for Computational Linguistics, Portland, Oregon, p. 18–29.
URL: *http://www.aclweb.org/anthology/W/W11/W11-2004*

op den Akker, R. & Traum, D. (2009), A comparison of addressee detection methods for multiparty conversations, *in* 'Proceedings of DIAHOLMIA'09, the 13th Semdial Workshop on the Semantics and Pragmatics of Dialogue', Stockholm, Sweden, pp. 99–106.

Padilha, E. & Carletta, J. (2002), 'A simulation of small group discussion', *Proceedings of EDILOG 2002: Sixth Workshop on the Semantics and Pragmatics of Dialogue* pp. 117–124.
URL: *citeseer.ist.psu.edu/padilha02simulation.html*

Patel, J., Parker, R. & Traum, D. (2004), Simulation of small group discussions for middle level of detail crowds, *in* 'Proceedings of the Army Science Conference'.

Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K. & Pelé, D., eds (2007), *Intelligent Virtual Agents, 7th International Conference, IVA 2007, Paris, France, September 17-19, 2007, Proceedings*, Vol. 4722 of *Lecture Notes in Computer Science*, Springer.

Plüss, B., DeVault, D. & Traum, D. (2011), Toward rapid development of multi-party virtual human negotiation scenarios, *in* 'Proceedings of SemDial 2011, the 15th Workshop on the Semantics and Pragmatics of Dialogue'.

Poesio, M. (1995), A model of conversation processing based on micro conversational events., *in* 'Proceedings of the Annual Meeting of the Cognitive Science Society'.

Poesio, M. & Traum, D. R. (1995), A multi-purpose model of conversational context, *in* 'IJCAI-95 Workshop on Context in Natural Language Processing'.

Poesio, M. & Traum, D. R. (1997), 'Conversational actions and discourse situations', *Computational Intelligence* **13**(3).

Poesio, M. & Traum, D. R. (1998), Towards an axiomatization of dialogue acts, *in* 'Proceedings of Twendial'98, 13th Twente Workshop on Language Technology: Formal Semantics and Pragmatics of Dialogue'.

Raux, A., Langner, B., Bohus, D., Black, A. W. & Eskenazi, M. (2005), 'Let's go public! taking a spoken dialog system to the real world', *Proceeding of the International Speech Communication Association* .

Rich, C., Sidner, C. L. & Lesh, N. (2001), 'Collagen: Applying collaborative discourse theory to human-computer interaction', *Artificial Intelligence Magazine* **22**(4), 15–25.

Roque, A., Leuski, A., Rangarajan, V., Robinson, S., Vaswani, A., Narayanan, S. & Traum, D. (2006), Radiobot-cff: A spoken dialogue system for military training, *in* 'Proceedings of Interspeech Conference'.

Roque, A. & Traum, D. (2007), A model of compliance and emotion for potentially adversarial dialogue agents, *in* 'In Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue'.

Roque, A. & Traum, D. (2008), Degrees of groundedness based on evidence of understanding, *in* 'The 9th SIGdial Workshop on Discourse and Dialogue'.

Roque, A. & Traum, D. (2009), Improving a Virtual Human Using a Model of Degrees of Grounding, *in* 'Proceedings of International Joint Conerence on Artificial Intelligence IJCAI-09', Pasadena, CA.

Sadek, M. D., Ferrieux, A. & Cozannet, A. (1994), Towards an artificial agent as the kernel of a spoken dialogue system: A progress report, *in* 'Proceedings of the AAAI Workshop on Integration of Natural Language and Speech Processing'.

Schegloff, E. A., Jefferson, G. & Sacks, H. (1977), 'The preference for self correction in the organization of repair in conversation', *Language* **53**, 361–382.

Schegloff, E. A. & Sacks, H. (1973), 'Opening up closings', *Semiotica* **7**, 289–327.

Searle, J. R. (1969), *Speech Acts*, Cambridge University Press, New York.

Sinclair, J. M. & Coulthard, R. M. (1975), *Towards an analysis of Discourse: The English used by teachers and pupils.*, Oxford University Press.

Smith, R. W. & Hipp, D. R. (1994), *Spoken Natural Language Dialog Systems: A Practical Approach*, Oxford University Press.

Smith, R. W., Hipp, D. R. & Biermann, A. W. (1995), 'An architecture for voice dialogue systems based on prolog-style theorem proving', *Computational Linguistics* **21**(3), 281–320.

Stevens, A., Hernandez, J., Johnsen, K., Dickerson, R., Raij, A., Harrison, C., DiPietro, M., Allen, B., Ferdig, R., Foti, S., Jackson, J., Shin, M., Cendan, J., Watson, R., Duerson, M., Lok, B., Cohen, M., Wagner, P. & Lind, D. (2006), 'The use of virtual patients to teach medical students history taking and communication skills', *American Journal of Surgery* **191**(6), 806–811.

Sutton, S., Novick, D. G., Cole, R. A. & Fanty., M. (1996), Building 10,000 spoken-dialogue systems., *in* '*Proceedings 4th International Conference on Spoken Language Processing (ICSLP-96)*'.

Swartout, W., Artstein, R., Forbell, E., Foutz, S., Lane, H. C., Lange, B., Morie, J., Noren, D., Rizzo, A. & Traum, D. (2013), 'Virtual Humans for Learning', *AI magazine; Special issue on Intelligent Learning Technologies* **34**(4), 13–30.

Traum, D. (2004), Issues in multi-party dialogues, *in* F. Dignum, ed., 'Advances in Agent Communication', LNAI 2922, Springer Verlag, pp. 201–211.

Traum, D. (2012), 'Non-cooperative and Deceptive Virtual Agents', *IEEE Intelligent Systems: Trends and Controversies: Computational Deception and Noncooperation* **27**(6), 66–69.

Traum, D., Jones, A., Hays, K., Maio, H., Alexander, O., Artstein, R., Debevec, P., Gainer, A., Georgila, K., Haase, K., Jungblut, K., Leuski, A., Smith, S. & Swartout, W. (2015), New dimensions in testimony: Digitally preserving a holocaust survivors interactive storytelling, *in* H. Schoenau-Fog, L. E. Bruni, S. Louchart & S. Baceviciute, eds, 'Interactive Storytelling', Vol. 9445 of *Lecture Notes in Computer Science*, Springer International Publishing, pp. 269–281.

Traum, D. & Larsson, S. (2003), The information state approach to dialogue management, *in* J. van Kuppevelt & R. Smith, eds, 'Current and New Directions in Discourse and Dialogue', Kluwer, pp. 325–353.

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. & Allen, J. F. (1992), A speech acts approach to grounding in conversation, *in* 'Proceedings 2nd International Conference on Spoken Language Processing (ICSLP-92)', pp. 137–40.

Traum, D. R. & Allen, J. F. (1994), Discourse obligations in dialogue processing, *in* 'Proceedings of the 32^{nd} Annual Meeting of the Association for Computational Linguistics', pp. 1–8.

Traum, D. R., Marsella, S., Gratch, J., Lee, J. & Hartholt, A. (2008), Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents, *in* H. Prendinger, J. C. Lester & M. Ishizuka, eds, 'IVA', Vol. 5208 of *Lecture Notes in Computer Science*, Springer, pp. 117–130.

Traum, D. R. & Morency, L. P. (2010), Integration of visual perception in dialogue understanding for virtual humans in multi-party interaction., *in* 'AAMAS International Workshop on Interacting with ECAs as Virtual Characters'.

Traum, D. R. & Rickel, J. (2002), Embodied agents for multi-party dialogue in immersive virtual worlds, *in* ‘Proceedings of the first International Joint conference on Autonomous Agents and Multiagent systems’, pp. 766–773.

Traum, D. R., Robinson, S. & Stephan, J. (2004), Evaluation of multi-party virtual reality dialogue interaction, *in* ‘Proceedings of Fourth International Conference on Language Resources and Evaluation (LREC 2004)’, pp. 1699–1702.

Traum, D., Roque, A., Leuski, A., Georgiou, P., Gerten, J., Martinovski, B., Narayanan, S., Robinson, S. & Vaswani, A. (2007), Hassan: A virtual human for tactical questioning, *in* ‘The 8th SIGdial Workshop on Discourse and Dialogue’.

Traum, D., Swartout, W., Marsella, S. & Gratch, J. (2005*a*), Fight, flight, or negotiate: Believable strategies for conversing under crisis, *in* ‘proceedings of the Intelligent Virtual Agents Conference (IVA)’, Springer-Verlag Lecture Notes in Computer Science, pp. 52–64.

Traum, D., Swartout, W., Marsella, S. & Gratch, J. (2005*b*), Virtual humans for non-team interaction training, *in* ‘In proceedings of the AAMAS Workshop on Creating Bonds with Embodied Conversational Agents’.

Tur, G., Stolcke, A., Voss, L., Peters, S., Hakkani-Tur, D., Dowding, J., Favre, B., Fernandez, R., Frampton, M., Frandsen, M., Frederickson, C., Graciarena, M., Kintzing, D., Leveque, K., Mason, S., Niekrasz, J., Purver, M., Riedhammer, K., Shriberg, E., Tien, J., Vergyri, D. & Yang, F. (2010), ‘The calo meeting assistant system’, *Audio, Speech, and Language Processing, IEEE Transactions on* **18**(6), 1601–1611.

Turing, A. M. (1950), ‘Computing machinery and intelligence’, *Mind* **59**, 433–460.
URL: *http://cogprints.org/499/*

Walker, M. A., Fromer, J. C. & Narayanan, S. (1998), Learning optimal dialogue strategies: A case study of a spoken dialogue agent for email, *in* ‘Proceedings COLING-ACL-98’.

Walker, M. A., Passonneau, R. & Boland, J. E. (2001), Quantitative and qualitative evaluation of darpa communicator spoken dialogue systems, *in* ‘ACL ’01: Proceedings of the 39th Annual Meeting on Association for Computational Linguistics’, Association for Computational Linguistics, Morristown, NJ, USA, pp. 515–522.

Wallace, R. S. (2009), The anatomy of a.l.i.c.e., *in* R. Epstein, G. Roberts & G. Beber, eds, ‘Parsing the Turing Test’, Springer Netherlands, pp. 181–210.

Weigand, E. (2006), ‘Argumentation: The mixed game’, *Argumentation* **20**(1), 59–87.

Weizenbaum, J. (1966), ‘ELIZA – a computer program for the study of natural langage communication between man and machine’, *Communications of the Association for Computing Machinery* .

Williams, J. D. & Young, S. (2007), ‘Partially observable Markov decision processes for spoken dialog systems’, *Computer Speech and Language* **21**(2), 393–422.

Zinn, C., Moore, J. D. & Core, M. G. (2002), A 3-tier planning architecture for managing tutorial dialogue, *in* S. A. Cerri, G. Gouardères & F. Paraguaçu, eds, 'Intelligent Tutoring Systems', Vol. 2363 of *Lecture Notes in Computer Science*, Springer, pp. 574–584.