# Communicative Intentions and Conversational Processes in Human-Human and Human-Computer Dialogue

Matthew Stone[1]
Department of Computer Science and Center for Cognitive Science
Rutgers, the State University of New Jersey
110 Frelinghuysen Road, Piscataway NJ 08854-8019
mdstone@cs.rutgers.edu

## Abstract

This chapter investigates the computational consequences of a broadly Gricean view of language use as intentional activity. In this view, dialogue rests on coordinated reasoning about communicative intentions. The speaker produces each utterance by formulating a suitable communicative intention. The hearer understands it by recognizing the communicative intention behind it. When this coordination is successful, interlocutors succeed in considering the same intentions—that is, the same representations of utterance meaning—as the dialogue proceeds. In this paper, I emphasize that these intentions can be formalized; we can provide abstract but systematic representations that spell out what a speaker is trying to do with an utterance. Such representations describe utterances simultaneously as the product of our knowledge of grammar and as actions chosen for a reason. In particular, they must characterize the speaker's utterance in grammatical terms, provide the links to the context that the grammar requires, and so arrive at a contribution that the speaker aims to achieve. Because I have implemented this formalism, we can regard it as a possible analysis of conversational processes at the level of computational theory. Nevertheless, this analysis leaves open what the nature of the biological computation involved in inference to intentions is, and what regularities in language use support this computation.

# 1 Introduction

The topic of this chapter is *pragmatic interpretation*, that is, our understanding of an utterance as an action chosen by its speaker to contribute to a conversation. In (1a), for example, we understand that *A* utters *hello* in order to greet *B* and get the conversation started. This is our pragmatic interpretation of (1a).

(1) a *A*: Hello.
    b *B*: Hello.

(2) a *A*: Where are the measuring cups?
    b *B*: In the middle drawer on the far right.

(3) a *A*: Pass the cake mix.
    b *B*: Here you go [handing over the package].

(2) and (3) are parallel, though more involved. In (2a), when *A* utters *where are the measuring cups*, we understand that *A*'s strategy is for *B* to reply by identifying a place where the cups *A* has in mind can be found, so that *A* will have this information. In (3a), when *A* utters *pass the cake mix*, we understand that *A*'s strategy is for *B* to perform the intended action, so that *A* will then obtain the mix *A* has in mind. Each of these attributed plans counts as an interpretation; it gives a rationale that explains why *A* used the utterance *A* did. Our ability to signal and recognize these interpretations holds our conversations together. In (1b), (2b) and (3b), for example, *B* simply recognizes the strategy behind *A*'s utterance, adopts that strategy, and follows through on it.

This chapter offers a formal, computational perspective on the role of representations of pragmatic interpretation in explaining our competence in contributing to conversation. In considering conversational competence, I will adopt the perspective of the knowledge level (Newell, 1982) or the level of computational theory (Marr, 1982) in cognitive science, and attempt to characterize our general ability, idealizing away from incidental errors and failings, to successfully formulate, use and understand utterances in conversation, as in (1), (2) and (3).[2] By formalizing representations of pragmatic interpretation, I hope to show in a precise way how these representations might serve as a bridge between the *language-as-product* tradition in the cognitive science of language use, which characterizes language processing in terms of the construction of symbolic grammatical representations, and the *language-as-action* tradition, which characterizes language processing in terms of the actions and interactions of collaborating interlocutors.

In particular, I will show how results from computational logic allow us to formalize a pragmatic representation as an abstract but systematic explanation of what a speaker is trying to do with an utterance. The formalization captures important insights from both traditions because its representations of interpretation enjoy these three properties simultaneously:

- They are recursive, symbolic structures. Thus, they are characteristically linguistic in being constituted by formal rules; they do not attempt, as our more general world knowledge might, to encode empirical regularities in a general way.

- They are sufficiently detailed to encompass all steps of disambiguating a linguistic structure. Thus, in important respects, we can characterize linguistic processes in terms of these representations; understanding is the hearer's inference to the representation behind a speaker's utterance, while production is the speaker's inference to a new representation with the potential to mediate a desired contribution to a conversation.

- They represent utterances as actions, and are in fact structured as reasons to act. Thus, deliberation in conversation can also be characterized in terms of these representations; doing so connects with a broader literature on intentions in communication and collaboration.

To achieve this, the formalism itself draws closely and evenly from both traditions. For example, in order to model linguistic problems such as disambiguation in terms of pragmatic representations, we must connect the rules that structure pragmatic representations directly with our knowledge of language—indeed, more specifically, with the derivations licensed by the mental grammar (Larson and Segal, 1995). At the same time, in order to connect such symbolic structures to choices in an uncertain and open-ended world, we must understand them as records of agents' commitments in linguistic action (Pollack, 1992), and recognize how keeping track of these commitments supports the diverse deliberative and collaborative processes we need for conversation (Clark, 1996). This balanced synthesis offers a number of advantages. This proposal is readily implemented (see (Stone et al., 2001; Stone, 2001)). It is simpler than than formalizations of speech act theory in the tradition of (Cohen and Perrault, 1979; Allen and Perrault, 1980), and more perspicuous than previous formal attempts to use action theory to link grammatical knowledge to participation in conversation (Appelt, 1985; Heeman and Hirst, 1995). In addition, this proposal strengthens and extends the genuine points of overlap between the language-as-action tradition and dynamic formalisms for meaning in dialogue, such as (Kamp and Reyle, 1993; Stokhof and Groenendijk, 1999; Ginzburg and Cooper, 2001; Asher and Lascarides, 2003), which already represent change (and by extension, action) as fundamental to interpretation.

In this chapter, I hope to suggest how an account of such representations can help to explain our impressive abilities in language use, and may provide working hypotheses for the qualitative and quantitative characterization of those abilities. At the same time, I hope to provide an introduction to research on intentions in cooperative dialogue from artificial intelligence (AI) and computational linguistics. Despite its engineering focus, this research increasingly holds peoples' utterances and meanings in conversation up to empirical scrutiny, and so increasingly converges with psycholinguistics both in its methods and in its results.

The structure of this chapter is as follows. In Section 2, I describe task-oriented dialogue, a prototypical setting in which pragmatic competence reveals itself. This description helps to motivate representations of speakers' intentions as essential for language users. In Section 3, I review philosophical and formal accounts of intentions in deliberation and agency, and suggest an understanding of intentions as complex mental representations structured to support decision-making and collaboration. This understanding, together with some assumptions about the syntactic structure and semantic representations of utterances, suffices to flesh out communicative intentions in particular (though of course many challenges remain). The payoff comes in Sections 4 and 5, where we look at understanding and production as operations on pragmatic representations. We can sketch how to implement understanding as a constraint-satisfaction process in which a language user reconstructs the interpretation of an utterance. When we take these interpretations to describe speakers' intentions systematically and abstractly, we help explain how language users infer consistent interpretations that are faithful to the grammar, faithful to the context and goals of the conversation, and also faithful to a wide range of probabilistic regularities in language use. Conversely, we can sketch how to implement production as a process of deliberation, in which a speaker formulates a suitable communicative intention. In production, the account helps explain how a speaker might exploit that same grammar, the same understanding of context, and perhaps even those same statistical regularities, to plan concise, grammatical and easily-understood contributions to a conversation.

## 2 Motivating Pragmatic Representations

One important source of evidence about the processes behind ordinary language use comes from conversation known as task-oriented dialogue, in which interlocutors aim to accomplish some practical real-world purpose, and language use serves this collaboration. The real-world focus of task-oriented dialogue offers analysts independent evidence from the world about the uses people make of language in cooperative interaction. Moreover, task-oriented dialogue represents a constrained form of language use; it offers analysts an idealized setting to investigate cooperative conversation which abstracts away from important elements of conversation, such as politeness, humor, or small talk, which reflect other functions of dialogue, such as supporting interlocutors' social relationships.[3] In this respect, task-oriented dialogue attracts enduring interest as the form of linguistic interaction that might most usefully be recreated in machines.[4]

Task-oriented dialogue exhibits a rich and detailed functional organization, which we can use to characterize the specific goals of individual utterances. This organization is best described by example; (4) offers a fragment (constructed to illustrate a range of typical phenomena in a short space), of a hypothetical dialogue in which interlocutors *A* and *B* prepare dinner together.

(4)  a  *A*: So are we all set?

b   *B*: The vegetables (pointing) are still too crunchy.

c   *A*: The zucchini there?

d   *B*: Yeah, the zucchini...

e   *A*: OK, I'll take care of it.

The dialogue suggests the effort that people make when they collaborate to maintain a detailed shared understanding of the status and direction of their joint activity. This effort goes well beyond simply keeping track of the real-world tasks that have been accomplished and the real-world tasks that remain (though such research as (Power, 1977) shows what a substantial endeavor this alone can be). This additional effort involves interlocutors' *attention* in dialogue and their *intentions* for dialogue.

As an activity progresses, new objects, actions and relationships in the world may come into play. Collaborators must redirect their attention accordingly. Interlocutors can draw on this coordinated attention to talk about their task more concisely and more coherently (Grosz and Sidner, 1986). More generally, in using utterances that describe particular objects, actions and relationships, interlocutors can also set up strong expectations about where to center attention for subsequent utterances (Grosz et al., 1995).

Example (4) illustrates both aspects of coordinated attention. *A* and *B* are able to use *the vegetables* and *the zucchini* to refer specifically to the zucchini that they have planned to cook for dinner, because their attention to the task distinguishes this zucchini from other things which they might have cause to talk about more generally—tomorrow's zucchini, still in the fridge, perhaps. Subsequent utterances about the zucchini reflect this attention and cement it linguistically. More spectacularly, with *I'll take care of it*, *A* is able to identify a specific task and commit to do it—putting the zucchini back in the microwave and heating them some more, let us suppose—without this task having being described explicitly in the recent conversation. *A* and *B* can be presumed to be attending to the task *A* identifies because of its relevance to their ongoing discussion and collaboration.

Meanwhile, as an activity progresses, further problem-solving and negotiation may be required to address outstanding goals. Collaborators' intentions in dialogue must address these meta-level tasks in addition to real-world tasks. 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. Modeling this problem-solving means recognizing the indirect role utterances play in achieving real-world goals (Litman and Allen, 1990; Lambert and Carberry, 1991; Carberry and Lambert, 1999) and the explicitly collaborative stake participants have in problem-solving discourse as well as real-world action (Grosz and Sidner, 1990; Lochbaum, 1998; Blaylock et al., in press).

Example (4) gets its coherence in part from the collaborative problem-solving strategy *A* and *B*

exhibit. For example, *B* takes *A*'s opening question (4a) as advancing a specific problem-solving activity: identifying any further task that remains to be done for dinner. *B*'s response in (4b) furthers this same problem-solving. *B* offers an indirect answer to *A*'s literal question; the two are not all set. But *B* also proposes, again indirectly, that finishing the zucchini is an outstanding subtask of preparing dinner that *A* and *B* should pursue next.

By characterizing the functional organization of task-oriented dialogue along these different dimensions, we are able to identify specific functional roles for individual utterances in collaboration. Consider (4a). By (4a), *A* makes the specific proposal in (5):
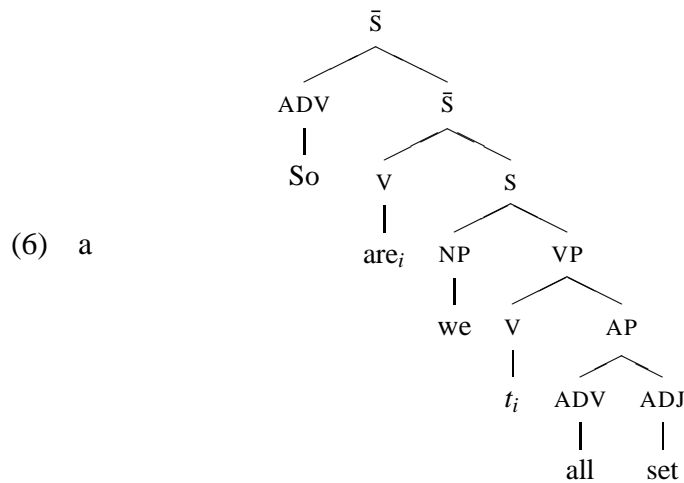
(5)     The dialogue should continue with an answer as to whether *A* and *B* are done cooking
        as of that time.

In so doing, *A* draws on the joint attention *A* and *B* maintain to the overall task they are engaged in, and suggests a way of identifying further subtasks to be done and thereby contributing to their collaboration.

Spelling out these functions explicitly, as in (5), enables a more precise characterization of interlocutors' collaboration in conversation. Clearly, people do not have to agree with one another completely to have a cooperative conversation. Our actions reflect our personal preferences, even in task-oriented dialogue. In (4b), for example, perhaps *B* proceeds this way in part because *B* doesn't really like zucchini, and privately hopes that *A* might now overcook tonight's batch into an inedible mush. Even if *B* has this nefarious ulterior motive, *B*'s response is still collaborative in that *B* uses it to acknowledge the evident meaning in what *A* has said, and to build on *A*'s contribution to develop the conversation further. In other words, what makes this a collaborative conversation is not that *A* and *B* have all the same goals, but simply that *A* and *B* are jointly attempting to integrate one another's utterances into a conversational record that gives a shared interpretation to what they are doing together (Thomason, 1990). This kind of collaboration seems indispensable; after all, as (4c) (*the zucchini there?*) and (4d) (*yeah, the zucchini*) shows, achieving such a shared interpretation can be problematic. Imagine if *B* were to answer uncooperatively with *I'm not telling* instead of (4d). It wouldn't be nefarious: it would end the conversation, or plunge it into absurdity. Accordingly, it makes sense to circumscribe the analysis of the collaborative intentions and deliberation behind utterances and to consider only functions like (5) which address the agreed content and direction of the conversation.

Even when we consider these circumscribed functions, interlocutors' conversational abilities are remarkable. Particularly astounding is the generality of the linguistic knowledge that interlocutors rely on to signal and recognize these functions. The grammar associates an utterance such as (4a) with a complex and abstract syntactic structure, and an equally complex and abstract meaning. For (4a), the syntax represented in (6a) and the semantics represented in (6b)—both undoubtedly

oversimplified—are indicative of the gap that language users must bridge to apply their linguistic knowledge in collaboration.

(6)  a

```
                    S̄
              ┌─────┴─────┐
            ADV           S̄
             │        ┌────┴────┐
            So        V         S
                      │      ┌───┴───┐
                    areᵢ    NP      VP
                             │    ┌───┴───┐
                            we    V       AP
                                  │    ┌───┴───┐
                                  tᵢ  ADV    ADJ
                                       │       │
                                      all     set
```

b   At the current time does the group containing the speaker have the property they require to be ready for the upcoming event?

It is difficult to characterize the relationship between representations like (6) and the specific function of utterances in collaboration like (5), even theoretically. We know that inference is required from philosophical models of discourse interpretation, such as implicature (Grice, 1975) and relevance (Sperber and Wilson, 1986) and in more explicitly computational frameworks for discourse interpretation, such as abductive interpretation (Hobbs et al., 1993) and commonsense entailment (Lascarides and Asher, 1991; Asher and Lascarides, 2003). But in conversation, inference must look beyond discourse, to embrace the collaborative setting and collaborative functions of language. This inference must lay out the resolution of ambiguity in linguistic terms while simultaneously describing utterances as actions that contribute to joint projects. This is a tall order—requiring representations of interpretation, for example, that respect both language-as-product and language-as-action traditions!

Despite the apparent gap between function and grammar apparent in (5) and (6), language users make connections to their ongoing collaboration quickly and easily in their word-by-word understanding of one anothers' utterances. Hanna and Tanenhaus (this volume), offer one clear demonstration. In their experiments, subjects were able to use knowledge of the goals and requirements of an ongoing collaboration to disambiguate referring expressions in instructions such as (7).

(7)      Now pass me the cake mix.

(7) was uttered by a confederate cook in a situation with two packages of cake mix. At stages of the collaboration where the cook needed help to reach only distant objects, subjects took *the cake*

*mix* in (7) to refer to the distant package. At other stages of collaboration, when the cook needed help with all objects, subjects regarded *the cake mix* in (7) as ambiguous. Amazingly, subjects' eye-movements showed that by the time they recognized the word *pass*, they had already arrived at strong expectations for the real-world location where the referent of the object NP could be found.

Such inferences seem just as essential for interlocutors' spontaneous language use. Brown-Schmidt and colleagues (Brown-Schmidt et al., 2002) collected task-oriented dialogue from naive pairs of subjects, and found that speakers systematically produce abbreviated referring expressions, as in (7), to exploit pragmatic constraints on reference. The analysis of hearers' eye-tracking data in these dialogues attests that the abbreviated references pose no difficulty for hearers either.

Such experiments provide strong evidence that peoples' representation and reasoning for interpretation can assess different ways of resolving linguistic ambiguities in light of the consequences for ongoing collaborations. This chapter draws on formal and computational research in pragmatics to explore one scheme by which this might be realized. I have implemented this scheme in a basic dialogue agent, so the agent can achieve the pragmatic disambiguation human subjects exhibit in (7). A preliminary description of this implementation from a computational perspective appears as (Stone, 2001); (Stone et al., 2001) fully describes a more substantial but less general implementation of a related framework.

The starting point for this exploration is Grice's proposal that interpretation is a species of intention (Grice, 1957; Grice, 1969). For Grice, a pragmatic interpretation simply represents what the speaker was trying to do with an utterance; to understand an utterance, language users must simply construct an appropriate such representation.

This idea might be taken as an almost tautological restatement of our problem—to understand an utterance, language users must recognize what the speaker intended. Indeed, when we speak of the intended analysis of an ambiguous expression, or the intended referent of a pronoun, or other intended aspects of utterance interpretation, we rarely stop to consider our implicit appeal to Grice's theory. However, I will show here how Grice's proposal places strong constraints on pragmatic formalization. Grice's proposal suggests that we must develop an account of interpretation by drawing on independent accounts of intention. This means that we must use the same kind of formal structures to record what a speaker was trying to do with an utterance as we use to record agents' commitments in taking other actions in support of collaboration. Likewise, Grice's proposal suggests that we must frame an account of processes in conversation in terms of independent accounts of deliberation and collaboration. This means that we must explain the work interlocutors do to understand one another with the same constructs we use to account for other interactions among people working together in dynamic and unpredictable environments. Grice's proposal is thus a deep and provocative one, whose consequences are yet to be fully worked out.

## 3   Intention, Collaboration and Communication

Our first step in describing pragmatic representations is to develop a more precise account of intention. Such an account must involve at least two ingredients: an account of individual intentions and rational deliberation; and an account of joint intentions and collaboration. In this section, I review one formal approach to these problems, and apply it to the case of communication. The approach is based on an understanding of intentions not simply as goals, propositions, or commitments to specific actions, but rather as rich, complex and symbolic representations of reasons to act.

### 3.1   Individual Intentions and Deliberation

Let us first understand a plan as a mental representation with a complex structure, as in AI (Pollack, 1990; Pollack, 1992). An agent's plan must set out, specifically or abstractly, what the agent is to do, when and in what circumstances the agent is to act, and what outcome the agent will thereby achieve. For the purposes of this paper, I understand an intention as a plan that an agent is committed to.

The simplest case involves plans and intentions that concern physical action in the world. Imagine that agent $A$ plans to turn on a light, for example. The corresponding plan-representation might set out that the agent is to flip the switch, in a situation where the light is off (but functional), and thereby yield the result that the light is on.

Laying out the content of a real-world plan this way recalls formal reasoning about action from AI, a tradition that begins with work on the situation calculus (McCarthy and Hayes, 1969; Green, 1969) and continues with work on more sophisticated models and ontologies today, including (Shanahan, 1997; Thielscher, 1999). Indeed, one specific representation for the content of a plan is an argument or inference in a formal theory of actions and their effects. A planning inference sets out an array of hypotheses—that the world starts out in a specified condition, that the agent performs a selected action, and that the world obeys specified causal principles. The inference then links these assumptions together to characterize the events that must ensue if these assumptions hold. For example, to record $A$'s plan to turn on the light, we might use the inference in (8). (Here $\wedge$ represents logical conjunction, $\supset$ represents logical implication, and $[\mathrm{N}]p$ represents change over time; $[\mathrm{N}]p$ means that $p$ holds in the agent's *next* cycle of deliberation, after one step of action.)

(8)  a  *off*                                                          Hypothesized situation.

    b  *flip*                                                         Hypothesized action.

    c  *off* $\wedge$ *flip* $\supset$ $[\mathrm{N}]on$                       Cause and effect.

    d  $[\mathrm{N}]on$                                          Modus ponens, (8a)–(8c)

(8a) specifies the condition of the world by hypothesizing that the light is *off*; (8b) specifies $A$'s action, that $A$ is going to *flip* the switch. (8c) makes a general hypothesis about the domain, that

if you flip the switch and light is off, it then goes on. (8d) is the consequence that follows under these hypotheses: *A* is going to turn on the light.

Such inferences encapsulate the information intentions must make explicit if they are to guide agents' deliberation. These representations thereby connect with the systematic accounts of rational deliberation proposed by researchers such as Bratman (Bratman, 1987). Inferences such as (8) map out actions for the agent to take (8b), they draw attention to circumstances (8a) and causal connections in the world (8c) that the agent must rely on to take that action, and they record the effects for which the agent might select the action (8d). In committing to this plan, an agent must take all these considerations into account. The agent must believe that the circumstances laid out in the plan will obtain (8a); the agent must expect to decide on the actions in the plan and carry them out in those circumstances (8b); the agent must believe that the outcome spelled out by the plan will occur (8c-d), and must on the whole, regard that outcome as favorable. If the agent reconsiders any of these conditions, the agent has reason to abandon the intention as unworkable or undesirable. But as long as the agent persists in its commitment to the plan, the agent can refer to the plan to determine how to act. In this way, the intention can play a causal role in the agent's pursuit and realization of its desires.

Further evidence for this understanding of intentions comes from the criteria people use to attribute intentions to one another. (See (Malle and Knobe, 1997), or in AI (Pollack, 1990).) Suppose we have observed an agent *A* flip a light switch, and consider what we must implicitly accept about *A* to attribute to *A* the intention of turning on the light. At some point, *A* must have been committed to flipping the switch. *A* must have understood that the light was then off but would go on once the switch was flipped. At that time *A* must, on the whole, have desired that outcome.

In contrast, if any of these conditions fails, we are more reluctant to attribute this intention to *A*. If *A* set out to bump the switch but not flip it, *A* had a different intention and failed to carry it off. Likewise if *A* thought the light was already on, or thought that flipping the switch would do something else instead, *A* had a different intention and failed to carry it off. It is more problematic if *A* was causally motivated to act by the plan to flip the switch and turn on the light, but without in any sense thinking that this course of action had anything to recommend it (even subconsciously). In this implausible situation, perhaps we must give up our idealization that *A* is engaged in rational deliberation at all.

Thus, to attribute an intention to an agent through a representation such as (8) is to say that the agent was guided in acting by that inference, and therefore to take on assumptions about the agent's beliefs about the current circumstances, the agent's causal knowledge, and the agent's desires for the future. This understanding may seem to diverge from ordinary ascriptions of intention. We normally say agents intend to do something, or to bring about some result, as though the content of the intention was simply an action or a goal. In fact, the theory assigns intentions a more complex

structure, linking actions to effects in context, in order to interpret intentions as mental represen-
tations that guide agent's deliberation and action. We should not regard our ordinary language
as an objection to the theory. Any English report of mental state describes both the content of
an individual's attitude and the cognitive representations behind it. However, it is impossible to
describe the objective meaning of an individual's mental state, as natural languages appear to do,
while still reporting individuals' representations exactly; representations with equivalent objective
content can have important differences, for example in the form in which they are represented. Be-
cause of this, semantic accounts of mental-state sentences require substantial flexibility in linking
content and representations. See for example (Crimmins, 1992). Our theory of intention, which
links action-content to inference-representations, is therefore no exception. Nevertheless, it will be
important to remember that when I describe an intention, I refer not just to an action or goal, but
to the complete ensemble of considerations that guide an agent's choice.

The planning inference in (8) describes the results of specific actions that the agent can already
identify and commit to. Of course, intentions must also allow for agents to postpone planning and
decision-making until subsequent steps of deliberation. For example, the agent may anticipate that
future information will affect its upcoming decisions.

The simplest way to accommodate future decisions is to model individual intentions as infer-
ences not about what an agent *will do*, as in (8), but about what an agent *would be able to do*.
This suggestion can be cashed out formally using theories of knowledge and action (Moore, 1985;
Davis, 1994). To carry out a program of action, an agent must meet two conditions at each step
where action is required. First, the agent will have to know what to do next: the agent must an-
ticipate that there will be a specific suitable real-world action to do. Second, the agent will have
to be able to construct a further intention that it can carry successfully through the remaining cy-
cles of deliberation and action. The intention representations we arrive at are symbolic, recursive
structures that appeal to logical accounts of knowledge and time to characterize actions and their
effects in context. (Fuller technical details are available in (Stone, 1998; Stone, 2001).)

Proofs such as (8) may look like they express knowledge about the world, but it is better to
think of them as *programs* that are annotated with assumptions that say when they can be exe-
cuted safely in an uncertain world. Seeing plans and intentions simultaneously as proofs and as
programs is a central idea in computation (Green, 1969; Howard, 1980). This idea is independent
of the more contentious view that our knowledge of the world has the content of logical axioms.
The difficulty with logic is that our empirical claims about the world are usually approximate or
statistical in character. Our best-guess predictions about the future, for example, will inevitably
involve uncertainty among different outcomes with different probabilities. A corresponding state-
ment of our knowledge of the world will be either incomplete or false if expressed in first-order
logic. However, intentions make explicit an agent's *commitments* for the future, not an agent's pre-
dictions or guarantees about what will happen. The uncertainty of our predictions has no bearing

on whether logical structures can record these commitments accurately and precisely. In fact, the commitments agents make in executing specific actions in their current circumstances certainly can have the definite content of logical statements.

Although this understanding of the role of logic is intrinsic to representations of programs as proofs, it is rarely noted explicitly in the literature. Nonetheless, in what follows, I will draw repeatedly on it to motivate simplified representations of action while sidestepping well-known difficulties. In (8), (8c) exemplifies this strategy. The axiom, repeated as (9) below, cannot be read as expressing what we know about lights.

(9)      $off \wedge flip \supset [\mathrm{N}]on$

As such, (9) would be quite a poor description of the world. Actually, we know that flipping the switch turns the light on only when the bulb is operational, the power is flowing to the switch, the switch is capable of making a connection, the circuit to the bulb is functioning, and so on. To describe the world in a general way, we would have to supply an indefinite number of further conditions to (9). The impossibility of specifying these conditions completely is known as the *qualification problem*. Meanwhile, we also know that flipping the switch has many other effects, both direct and indirect, that (9) omits. In these circumstances, flipping the switch might contribute to the wear on the bulb, it might heat the room slightly, it might leave a smudge of dirt on the wall or the switchplate, and so on. The impossibility of specifying these completely is known as the *ramification problem*.[5]

All the same, when agent *A* decides to flip the switch and turn on the light—that is, when agent *A* adopts the intention in (8)—agent *A* really is committed to the truth of (9) in this instance. The qualification problem is important for understanding what *A*'s commitment means; AI researchers now understand that in making commitments such as (9), *A* must be understood as reasoning in a certain *context*, which depends in an indefinite way on unspecified further assumptions (McCarthy and Buvač, 1994). But the qualification problem does not stand in the way of representing *A*'s commitment logically.

Conversely, when agent *A* adopts the intention in (8), *A* is not committed to all the ramifications of flipping the switch. For example, if the ramifications do not take place, *A* will not have failed to carry off this intention. And if *A* discovers an obstacle to this intention, *A* will not reason and act to make sure these ramifications take place anyway. Again, the ramification problem may be important for understanding what *A*'s commitment means. Rationally (and ethically), *A* must strive to identify and defuse potential negative consequences of intended actions. But the ramification problem does not stand in the way of representing *A*'s commitment logically.

We can make similar observations about the idealizations involved in reporting inferences about knowledge in logic. I will use [C] to describe the information that interlocutors presuppose in conversation: their mutual knowledge (Stalnaker, 1973) or common ground (Clark and Marshall,

1981); [C]$p$ means that $p$ is shared. I describe [C] (like [N]) using the logical machinery of *modal logic*, so that the common ground is assumed to give a consistent but incomplete picture of the world. (For an introduction to modal logic, see (Fitting and Mendelsohn, 1998).) The inference from (10a) and (10b) to (10c) is a consequence of this assumption.

(10)  a   [C]$p$
      b   [C]$(p \supset q)$
      c   [C]$q$

If we make unrestricted use of these inferences in assessing our real mental states, we find a problem of *logical omniscience*. We predict that we know all the consequences of what we know, all the theorems of mathematics for example. It is a rather poor description of what our knowledge actually is, and as such could certainly be improved (Konolige, 1985).

But again, this is no obstacle to the use of logic to formalize our epistemic commitments in deliberation. Any one plan is a finite structure that involves only a fixed number of inferences about knowledge. Each such inference requires an agent to perform a specified cognitive operation in the course of carrying out the plan. The agent may have to remember some fact, or link two facts together in a predetermined way. When an agent commits to an intention involving inferences about knowledge, then, the agent commits to these operations, and nothing more. And commit the agent must: if the agent fails to remember, or fails to draw an inference, the agent will lose track of the plan's upcoming choices, or the rationale behind them.

By the foregoing considerations, I hope to underscore that representations of intentions as inferences such as (8) are circumscribed and economical. They offer parsimonious specifications of an agent's commitment to act in a certain way, because they presuppose sophisticated deliberative processes that manage these commitments. For example, the agent's planning processes must use inference about action to identify commitments that the agent can make consistently. The agent's updating processes for plans must use reason-maintenance to assess the continued appropriateness of the agent's commitments in a dynamic and unpredictable environment. And the agent's execution mechanisms must ensure that the agent keeps track of the right information, makes the right choices and completes the right actions while pursuing its plans.

Thus far, of course, we have considered these deliberative processes themselves only in the most general way. So it should be clear that intention representations are compatible with quite different characterizations of these processes, and quite different characterizations of the information and representations that these processes may require in addition to intentions. For instance, since decision theory provides a normative characterization of rational action, we might envision processes that commit to specific intentions based on calculations of probabilities and utilities derived from empirical regularities. (See (Pollack and Horty, 1999) for AI research along these lines.) For example, an agent's decision to commit to *off* ∧ *flip* ⊃ [N]*on* might reflect the agent's judgment

of the probability that the light will go on in circumstances where the light is off and the agent flips the switch. That conditional probability might in turn be estimated from empirical observations. Nevertheless, we need not expect any representations of these empirical generalizations to figure explicitly in intentions; indeed, we need not even expect them to be represented in the same kind of way as intentions.[6]

## 3.2 Joint Intentions and Collaboration

When groups of agents collaborate to achieve goals they share—and conversation must be considered such a case—groups of agents must sometimes commit to plans that lay out programs of coordinated action for the group. I will understand these collaborative plans by analogy to individual plans, as complex mental representations. Collaborative plans set out, specifically or abstractly, what each agent is to do, when and in what circumstances each agent is to act, and what outcome the group will thereby achieve. These plans can again be represented as formal inferences, in a logic of knowledge and time, which describe what members of the group would be able to do together.

As a simple example of collaboration on a real-world task, let's consider taking a posed picture, as you and a companion might do on vacation in front of a famous landmark. In this case we have two agents *A* and *B*, and a situation in which agent *B* has a ready camera. *A* first poses, adopting some distinctive expression and attitude towards *B*; this gets *A* set to be photographed. Then *B* snaps the shutter, with the result that *A*'s pose is recorded for posterity. The inference in (11) records the content of this plan in a simple inferential form that parallels (8).

(11)  a  *ready*                                                            Hypothesized situation.

      b  *pose*                                                         Hypothesized action.

      c  $pose \supset [\text{N}]set$                                        Cause and effect.

      d  $ready \wedge pose \supset [\text{N}]ready$                            Persistence.

      e  $[\text{N}]set$                               Modus ponens, (11b), (11c).

      f  $[\text{N}]ready$                        Modus ponens, (11a), (11b), (11d).

      g  $[\text{N}]snap$                              Hypothesized action.

      h  $[\text{N}](ready \wedge set \wedge snap \supset [\text{N}]pic)$           Cause and effect.

      i  $[\text{N}][\text{N}]pic$       Modus ponens, (11f), (11e), (11g), (11h) and temporal logic.

Here (11a) specifies the condition of the world by hypothesizing that the camera is *ready*; (11b) specifies *A*'s action, that *A* will *pose*. Later, (11g) specifies *B*'s action, to *snap*. Causal assumptions include axioms (11c) and (11h) about change, and the axiom of persistence (11d). The result that follows, by a chain of intermediate reasoning from these hypotheses, is that *A* and *B* will record the picture after the two steps of action, $[\text{N}][\text{N}]pic$ in (11i).

In view of the formal structure of such plans, and the function that commitment to them plays in the deliberation of members of the group, it is reasonable to understand them as joint intentions. As before, these inferences encapsulate the considerations that agents must take into account in committing to and pursuing a course of action. For example, in committing to their collaboration, *A* and *B* must jointly believe that the circumstances described in the plan will fit the situation in which they must act. They must agree that each will decide on their own actions and carry them out as specified in the plan. They must expect that the outcome spelled out by the plan will occur, and agree that this outcome is favorable. These attitudes are described more precisely in (Cohen and Levesque, 1991; Grosz and Kraus, 1996). On this understanding, a joint intention simply reflects the coordination of agents' individual commitments and deliberation (rather than an irreducibly joint mental state as in (Searle, 1990)).

The actions that agents may have to take to carry out a joint project are significantly more involved than a single agent's commitment with an individual intention, however. These additions serve to ensure that agents achieve the most successful possible conclusion, in a coordinated way, even in the face of potential obstacles.

Thus, 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. Otherwise collaborators might suspect that something has gone wrong. In our photography example, *A* must not only adopt a pose; *A* must allow *B* to recognize that *A* has done so. Of course, *A* has a variety of devices for this, from stylized flourishes of movement, to simple verbal announcement: "Here I go" (beforehand) or "OK" (afterwards).

Conversely, each agent must work to recognize the actions of the other agents in the context of the ongoing collaboration. Otherwise one agent may remain unaware of another's failures. Before *B* takes the picture, if *B* is serious about having the picture come out, *B* must attend to *A*, recognize whether *A* has attempted to achieve the right look, and judge whether *A* has succeeded. In addition, when an agent detects that the intention has failed, the agent must communicate this to the group as a whole. Likewise, when an agent achieves success, the agent must make sure that the group as a whole is aware of this.

Joint intentions as represented in (11) thus presuppose sophisticated processes of coordination, just as individual intentions as represented in (8) presuppose sophisticated processes of deliberation. In computational agents, such processes have proved essential in allowing groups of agents to work together on tasks from robotic soccer to search-and-rescue (Tambe et al., 1999) and in allowing individual agents to be understood by their human partners (Sengers, 1999). Of course, we also find such processes of coordination in psychological accounts of language use (Clark, 1996) and computational models of dialogue (Cassell et al., 2000). Again, representations of intentions provide only a partial framework for characterizing these processes; these processes must have access to many other kinds of information, and this information may involve very different form and con-

tent from intentions, especially when it is derived from experience in a general way. However, our systematic appeal to these processes of coordination makes it possible for us to understand simple logical structures such as (11) as principled representations. Intention representations formalize commitments, independent of the effort required to manage and pursue those commitments.

### 3.3 Cooperative Conversation

In Section 2, I argued that to account for peoples' language use in task-oriented dialogue, we must view interpretations as pragmatic representations that link together different ways of resolving linguistic ambiguities to their consequences for an ongoing collaboration. Grice theorized that speakers' intentions are what establish such links (Grice, 1957; Grice, 1969). Grice argued that in making a meaningful contribution to conversation with an utterance, a speaker must intend to make this contribution, in a recognizable way, through the ordinary process of linguistic communication. Ordinary linguistic communication depends on the fact that speakers manifest their intentions in utterances and hearers recognize them in this way. This is just what we would expect from analyzing conversation as a deliberative and collaborative process as characterized in Sections 3.1 and 3.2.

Accordingly, we can now draw on Grice's theory to specify the content and representation of utterance interpretations in dialogue more precisely. Interpretations record the speaker's commitments in using an utterance to advance a cooperative conversation. Specifically, they hypothesize an event of utterance, and perhaps further actions as well. They specify the assumed context in which the utterance is to be made, and they draw on an assumed idealization of the way utterances and other actions bring about context change. They link these assumptions together into an argument that shows how the speaker's use of the utterance in context can advance the status of a joint project. When formalized, such arguments serve as the *representations of pragmatic interpretation* introduced in Section 1 and motivated in Section 2—abstract but systematic explanations of what a speaker is trying to do with an utterance.

Let us proceed by formalizing a simple example, and examining the result from the perspective of language-as-product and language-as-action traditions. Consider (12b).

(12)  a  *A*: Did the man stay?
      b  *B*: The man left.

What are the speaker *B*'s commitments in offering this answer? *B* proposes to utter *the man left*, in a context which provides a discourse referent *m* under discussion. *B* assumes that, in virtue of its meaning, this utterance will contribute the information that *m* left. *B* further presumes that if *m* left, he did not stay. In this way *B* has answered *A*'s question and resolved an outstanding shared goal of the conversation.

(13) lays out these commitments as an inference. In tracing the argument, note that premises (13c) and (13e) contain a free variable $M$, which is instantiated to (that is, replaced with) $m$ when drawing conclusions from it.

(13) a $\quad [\text{C}]man(m)$                            Hypothesized contextual situation.

     b $\quad utter(\text{"the man left"})$                      Hypothesized action.

     c $\quad ([\text{C}]man(M)) \wedge utter(\text{"the man left"}) \supset [\text{N}][\text{C}]left(M)$    Cause and effect (grammar).

     d $\quad [\text{N}][\text{C}]left(m)$               Modus ponens and instantiation, (13a)–(13c).

     e $\quad [\text{N}][\text{C}](left(M) \supset \neg stayed(M))$     Hypothesized contextual situation, persistence.

     f $\quad [\text{N}][\text{C}]\neg stayed(m)$   Modus ponens and instantiation, (13d), (13e) and logic of knowledge and time.

By inspecting (13), you can check that it exhibits the characteristics advertised for such representations in Section 1.

- (13) is a recursive, symbolic structure, constructed according to formal rules—the rules of logical deduction. As we have seen, (13) formalizes a commitment rather than expressing knowledge and so abstracts away from the additional processes of coordination that may be required to follow up (12b), and the corresponding uncertainty in whether (12b) will achieve its intended effect.

- (13) details the precise contribution of the grammar in determining interpretation. Premise (13c) is the crucial one. It says that, provided the common ground saliently provides a man $M$, then the action of uttering *the man left* will contribute the fact that $M$ left to the common ground. This statement is a record of the commitments of the language faculty to a theorem about the analysis of a sentence of English, in the spirit of the knowledge of meaning investigated in (Larson and Segal, 1995); in particular, we can easily imagine (and readily implement) a process of inference that would derive this statement from a suitable compositional syntax and semantics.

- (13) describes $B$'s utterance as an action, and is structured as a reason to act. In particular, premise (13c) characterizes the potential that the utterance has to draw on and to update an evolving context of collaboration. Descriptions of context change are also familiar in linguistics from discourse representation theory and dynamic semantics (Kamp and Reyle, 1993; Stokhof and Groenendijk, 1999). What makes this different is that premise (13c) is a hypothesis about cause and effect which the speaker represents and commits to, as part of a communicative intention. The behavior of (13c) in language use follows from its status as a represented commitment. For example, the antecedent condition $[\text{C}]man(M)$ functions as an anaphoric presupposition (Kripke, 1991; van der Sandt, 1992) in the sense that the

16

speaker *B* intends the common ground to supply a specific resolution for this condition, and in particular to supply the salient value *m* for its free variable *M*. The hearer *A* must recognize this link to understand (12b) as (13).

In generalizing from this small example, the formalism allows us to include two further kinds of inferences that link meaning to context. First, where (13) includes premises that specify contextual parameters like *m* for *M* directly, such parameters may also be derived by inference. For example, in the bridging anaphora first described by (Clark, 1975), context supplies an entity that is new to the discourse but is related to one that has been evoked previously.

(14)     Chris peeled the cucumber and removed the seeds.

In the case of (14), *the seeds* are understood to be the seeds of the cucumber Chris has just peeled. Such relationships can be recorded in pragmatic representations as inferences which establish the antecedents of rules such as (13c) by appeal to premises describing salient objects and general world knowledge. This provides a general interface to link utterance interpretation to attentional state in dialogue. (See also (Hobbs et al., 1993; Piwek, 1998; Webber et al., to appear).)

A second set of inferences may spell out the contribution the utterance makes to the ongoing task. (13) includes the simplest case of such inference; (13e) and (13f) describe why *B*'s utterance answers *A*'s question. More generally, as in (11), these inferences may also proceed by hypothesizing further actions that participants will take as part of their collaboration. This provides a general interface to link utterance interpretations to the intentional structure of dialogue.

The utterances in (4) or (7) depend of their analysis on all three kinds of inference—grammatical inferences, attentional inferences and intentional inferences. To describe these utterances, I will emphasize the content that we can now assign to speakers' intentions, and leave the formalism for future presentations. (But for provisional attempts for related examples consult (Stone, 2000; Stone, 2001; Stone et al., 2001).)

Consider the interpretation of (7), *Now pass the cake mix*. In the circumstances of Hanna and Tanenhaus's experiments, the argument behind the speaker's utterance must proceed along the following lines. The speaker assumes that the context provides a package of cake mix *p*, and hypothesizes uttering (7) in this situation. Drawing on a representation of the syntactic structure of (7) and its dynamic meaning, the speaker is thereby committed that the utterance will impose an obligation on the hearer to pass *p*. Now as a further development, the speaker hypothesizes that the hearer does pass *p*. The result of this action, inferred by a logical representation of the speaker's commitments in action, is that the speaker will have *p*, and can therefore use *p* for further steps in the unfolding recipe. This chain of inference, as needed, links the reference resolution required to interpret *the cake mix* with goals and expectations that the speaker has for the collaboration.

And consider (4a), *So are we all set*. We can reconstruct the speaker *A*'s intention with this utterance as the following argument. *A* begins with a suite of assumptions about the context: that

participants *A* and *B* form a group containing the speaker; that it is now 6:30pm; that *dinner* is an upcoming event; and that being *done cooking* is a property that *A* and *B* must have to be ready for dinner. *A* hypothesizes uttering *So are we all set*, as analyzed in (6a), under these circumstances. The meaning of this utterance—in contexts where *G* is a group containing the speaker, *N* is the current time, *E* is an upcoming event, and *P* is a property that *G* must have to be ready for *E*—is to ask the hearer to provide an answer as to whether *G* does have property *P* at time *N*. With this semantics we specify (6b) in terms of an anaphoric presupposition and a dynamic contribution to the conversation, by analogy to (13c). By inference in this case, then, *A* commits to ask whether *A* and *B* are *done cooking* as of 6:30. To account for the function of this question in the collaboration, the interpretation may map out the further course of *A* and *B*'s collaborative conversation. Suppose *B* responds with an answer, *yes* or *no*: in either case, we can envisage *A* and *B* proceeding to conclude the collaboration thereafter. After a *yes*, *A* and *B* achieve common ground that the goal is achieved. After a *no*, *A* and *B* proceed to work out how to achieve the specific further tasks that they can then identify.

In our Gricean framework, intentions such as these are the objects of interlocutors' deliberation and coordination in conversation. For example, in offering an utterance in conversation as part of a specific plan or intention, the speaker is committed that the circumstances and conditions laid out in the plan obtain, and that the outcome envisaged in the plan is advantageous. These commitments are a standard feature of formal analyses in the speech act tradition of (Searle, 1969; Cohen and Perrault, 1979; Allen and Perrault, 1980). Yet even though these commitments can be derived from the role of intentions in deliberation (Cohen and Levesque, 1990), prior analyses have not offered representations of intentions that abstract away from these commitments in a general way. Meanwhile, since the conversation is a collaboration, the speaker must also ensure that the intention behind the utterance will be recognizable to the other interlocutors. At the same time, collaboration allows the speaker to presume that interlocutors maintain coordinated attention and intentions toward the ongoing task, and it allows the speaker to presume that interlocutors will recognize the plan and pursue it, by carrying its actions through and by grounding its success or failure. Again, previous formalizations in the speech-act tradition recognize the speaker's anticipation of such collaborative effort in understanding and grounding (Appelt, 1985; Traum, 1994; Heeman and Hirst, 1995). But this general reasoning has been formalized explicitly as part of the content of speakers' intentions, offering no interpretation of language as action that abstracts away from it. The formalism sketched here is therefore substantially simpler than what has previously been available—simple enough, in fact, to enable a straightforward, efficient first-principles implementation.

## 4 Understanding as Intention Recognition

The rich but parsimonious pragmatic representations introduced in Section 3.3 can help us, in certain circumscribed ways, to characterize conversational processes. For example, I have used these pragmatic representations to implement processing modules for computational agents whose language use exhibits interesting commonalities with our own.

I start with the problem of interpretation. From this perspective, interpretation is what AI researchers call a plan-recognition problem (Thomason, 1990; Carberry, 2001). The hearer perceives some actions (an utterance), and must determine what the speaker was trying to do by reconstructing a representation of the speaker's intention. When the hearer perceives utterance (7), for example, the plan-recognition problem is to reconstruct the argument for it sketched in Section 3.3.

Reconstructing such an argument involves reconciling constraints from grammar and logic with constraints from attention and intention in dialogue. Consider (7) in the unambiguous context where the cook needs help getting objects near the subject (and we find that subjects do go on to pass the mix near them). The grammar of English analyzes (7) so as to assign it the intended semantic form, presupposing some cake mix and introducing an obligation that the hearer pass it. But the grammar doubtless offers other analyses; one and only one must figure in the recognized interpretation. This is one constraint—one that might be modulated further by probabilistic knowledge about the frequency with which such constructions are used in English.

The attentional state of the conversation is another constraint. The speaker's presuppositions must be resolved by supplying salient individuals and facts from the environment; indeed, perhaps the matches must be as salient as possible. For (7) there is the intended package of cake mix $p$ near the hearer; perhaps there are others elsewhere in the environment. Again, exactly one match must figure in the recognized interpretation.

From the intentional state, we get constraints on the overall goals that the utterance can be meant to achieve. For (7), we know the speaker would initiate a joint project to get some objects, including $p$, but not others. Of course, the collaboration might have other outstanding goals, and the hearer must identify which of them figures in the recognized interpretation (if any).

Finally, of course, these different components of interpretation must be fit together compatibly into a single inference about what speaker and hearer would be able to do together. In the case of (7) as represented here, the intended argument is the only one that fits these constraints. That is, only by taking (7) as an instruction to pass $p$ can we reconcile grammatical options with the salient objects, including $p$, that the speaker might refer to, and the outstanding goals, including potentially obtaining $p$, that the collaboration provides.

We can implement this constraint-satisfaction analysis directly in computational logic (Stone, 2001). This is sufficient to create a system that resolves references by linking evoked discourse referents to speakers' high-level goals via linguistic descriptions, as in (7). Indeed, by representing

grammatical knowledge in a suitable form and applying all constraints incrementally in utterance interpretation, as in (Haddock, 1989) or (Schuler, 2001), such an implementation would even be able to replicate the on-line resolution of ambiguity that Hanna and Tanenhaus describe.

Methodologically, such implementations bear on what Newell (Newell, 1982) calls the knowledge level and Marr (Marr, 1982) calls the level of computational theory: they give evidence about the represented regularities in the world that that make it possible for computational devices (ourselves included) to perform a real-world task. Such implementations have rather less to say about what algorithms might be involved in human language understanding. The case of the grammar is a familiar microcosm. Reconstructing an intended grammatical analysis might be a case of strategic exploration of logical possibilities set out by our knowledge of language, as with the different algorithms explored in (Altmann, 1988; Frazier and Clifton, 1996). Alternatively, grammatical reasoning might be informed by probabilistic generalizations on language use over and above whatever knowledge of grammaticality we may have, as in (Tanenhaus and Trueswell, 1995; Seidenberg and MacDonald, 1999).

To consider the additional constituents of pragmatic interpretation is to introduce such possibilities anew. Attentional state may have a purely logical implementation or it may take into account many kinds of probabilistic influence; either case leaves open wide ranges of processing strategy. The same goes for the outstanding goals of a collaboration. We might consider this a negative, were it not for a suspicion that all such cases, and many others, are governed by common cognitive constraints and common principles of biological computation. (And were it not for our appreciation for the complexity and significance of the problem of language understanding itself!)

## 5   Production as Intention Generation

In production, the speaker starts with a contribution that might usefully be made to an ongoing collaboration. What the speaker needs is a specific utterance that will make this contribution. The utterance can achieve this effect only if the speaker's intention in using it is recognizable. So the speaker's production problem is to formulate a suitable pragmatic representation for a potential utterance. The speaker must judge that the hearer can use shared information, the utterance, the grammar, and the attentional and intentional state of the discourse, to reconstruct this interpretation. Thus, pragmatic representations are the *outputs* of both understanding and production! This fact nicely emphasizes the algorithmic flexibility observed in Section 4.

This is the formulation of the language production problem that I and my colleagues arrived at in the generation system SPUD (for *sentence planning using description*) (Stone and Doran, 1997; Stone and Webber, 1998; Stone et al., 2001). SPUD can generate concise, contextually-appropriate utterances, including both speech and concurrent nonverbal behavior, by applying a simple, uniform and efficient decision-making strategy. This strategy gradually constructs an interpretation by refining the generator's existing commitments and considering new actions that are compatible

with those commitments. In this sense, this strategy can be regarded as a special case of more general processes of deliberating with intentions.

Specifically, SPUD's strategy exploits the lexicalized tree-adjoining grammar (LTAG) formalism in which SPUD's grammar is represented (Joshi et al., 1975; Schabes, 1990). LTAG grammars derive sentences by incorporating meaningful elements one-by-one into a provisional syntactic structure. SPUD makes these choices head-first and incrementally, in the order its grammar provides. (Compare also (Ferreira, 2000; Frank and Badecker, 2001).)

At each stage of derivation, SPUD determines both the intended interpretation for a provisional utterance and the interpretation that the hearer would recognize from it. SPUD again implements this interpretation process directly in computational logic. (SPUD's interpretations do not hypothesize actions that follow the utterance, so SPUD does not account for the role conversational goals in disambiguation; in all other respects SPUD implements the formal account sketched in Section 3.3.)

These pragmatic representations guide SPUD's choices of what elements to add to an incomplete sentence. The structure of the utterance suggests ways the sentence may be elaborated with further meaningful elements. The intended pragmatic interpretation of each elaboration makes explicit the specific information that the utterance could contribute, and the specific links with the context that the utterance establishes. Meanwhile the recognized interpretation records SPUD's progress towards unambiguous formulation of referring expressions. These representations allow SPUD to assess by simple heuristics which choice might best suit the ongoing conversation, and to commit to that choice.

## 6 Conclusion

Let us return to (4), to take stock of the principles of the account, and the many further problems that remain.

(15) a    *A*: So are we all set?
     b    *B*: The vegetables (pointing) are still too crunchy.
     c    *A*: The zucchini there?
     d    *B*: Yeah, the zucchini...
     e    *A*: OK, I'll take care of it.

In order to account for such collaborative conversations, I have suggested a broadly Gricean formalization of language use as intentional activity. Each utterance in a dialogue such as (15) manifests its speaker's intention: a complex, symbolic mental representation that characterizes the speaker's utterance in grammatical terms, links the utterance to the context, and describes the contribution to the collaboration that the speaker commits to making with the utterance. The dialogue itself proceeds through interlocutors' reasoning about these intentions: the speaker produces each

utterance by formulating suitable intentions, while the hearer understands each utterance by recognizing the intention behind it. When this coordination is successful, interlocutors succeed in considering the same representations of utterance meaning as the dialogue proceeds.

Of course, representations of pragmatic interpretation, like all intention representations, should provide a resource for action and cooperation beyond just acting and understanding. A recognized interpretation may help shape questions and elaborations, as in (15b) and (15c). In response to *B*'s answer, *A* proposes a possible refinement of *B*'s interpretation. *A* specializes *B*'s vocabulary but preserves much of the structure of *B*'s utterance, its links to context, and its function for the ongoing task. See (Clark and Wilkes-Gibbs, 1986; Heeman and Hirst, 1995). Coordination at the level of pragmatic interpretation may also help to signal satisfactory understanding, as in (15d). Here *B* repeats not only *A*'s words themselves, but also their interpretation, in order to mark this interpretation as recognized and its contribution as shared (Brennan, 1990; Brennan and Clark, 1996).

More generally, of course, a cognitive science of language use is responsible not only to explain adult conversation, but also to elucidate its relationship to other cognitive abilities in ourselves and other species, and to account in particular for infants' ability to learn language. I am intrigued by the synergy with related work that a computational theory of pragmatics along Gricean lines might afford. For example, in characterizing language use in terms of representations of intentions, it plays into a tradition beginning with Aristotle and continuing with such work as (Sperber, 2000) in linking human language to a representational understanding of one's own and others' mental states that is uniquely human. Meanwhile, intention-based pragmatic representations seem necessary to flesh out the rich bootstrapping view of language acquisition that theorists increasingly adopt—see (Gillette et al., 1999; Seidenberg and MacDonald, 1999; Bloom, 2000), or the chapter by Trueswell and colleagues (this volume)—in which acquisition of language depends on integrating multiple sources of evidence, including not only observed utterances and innate constraints of grammar but also learners' understanding of and interaction with the people whose language they learn.

# References

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

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.

Altmann, G. (1988). Ambiguity, parsing strategies and computational models. *Language and Cognitive Processes*, 3:73–98.

Appelt, D. (1985). *Planning English Sentences*. Cambridge.

Asher, N. and Lascarides, A. (2003). *Logics of Conversation*. Cambridge.

Blaylock, N., Allen, J., and Ferguson, G. (in press). Managing communicative intentions with collaborative problem solving. In Smith, R. and van Kuppevelt, J., editors, *Current and New Directions in Dialogue*. Kluwer.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. MIT.

Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*. Harvard.

Brennan, S. E. (1990). *Seeking and Providing Evidence for Mutual Understanding*. PhD thesis, Stanford University.

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

Brown, P. and Levinson, S. C. (1987). *Politeness: Some Universals in Language Use*. Cambridge.

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*, pages 148–153.

Carberry, S. (2001). Techniques for plan recognition. *User Modeling and User-Adapted Interaction*, 11(1):31–48.

Carberry, S. and Lambert, L. (1999). A process model for recognizing communicative acts and modeling negotiation subdialogues. *Computational Linguistics*, 25:1–53.

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):1–44.

Cassell, J., Bickmore, T., Campbell, L., Vilhjalmsson, H., and Yan, H. (2000). Human conversation as a system framework. In Cassell, J., Sullivan, J., Prevost, S., and Churchill, E., editors, *Embodied Conversational Agents*, pages 29–63. MIT.

Clark, H. H. (1975). Bridging. In Schank, R. and Nash-Webber, B., editors, *Theoretical Issues in Natural Language Processing*, pages 169–174. MIT.

Clark, H. H. (1996). *Using Language*. Cambridge.

Clark, H. H. and Marshall, C. R. (1981). Definite reference and mutual knowledge. In Joshi, A. K., Webber, B. L., and Sag, I., editors, *Elements of Discourse Understanding*, pages 10–63. Cambridge.

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

Cohen, P. R. and Levesque, H. J. (1990). Rational interaction as the basis for communication. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*, pages 221–256. MIT.

Cohen, P. R. and Levesque, H. J. (1991). Teamwork. *Nous*, 24(4):487–512.

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

Crimmins, M. (1992). *Talk About Beliefs*. MIT.

Davis, E. (1994). Knowledge preconditions for plans. *Journal of Logic and Computation*, 4(5):721–766.

Ferguson, G. and Allen, J. F. (1998). TRIPS: An intelligent integrated problem-solving assistant. In *Proceedings of AAAI*, pages 567–573.

Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. In Wheeldon, L., editor, *Aspects of Language Production*, pages 291–330. Psychology Press.

Fitting, M. and Mendelsohn, R. L. (1998). *First-order Modal Logic*. Kluwer.

Frank, R. and Badecker, W. (2001). Modeling syntactic encoding with incremental tree-adjoining grammar: How grammar constrains production and how production constrains grammar: Abstract. In *CUNY Conference on Human Sentence Processing*, page 19.

Frazier, L. and Clifton, C. (1996). *Construal*. MIT.

Geis, M. (1995). *Speech acts and conversational interaction*. Cambridge.

Gillette, J., Gleitman, H., Gleitman, L., and Lederer, A. (1999). Human simulations of lexical acquisition. *Cognition*, 73:135–176.

Ginzburg, J. and Cooper, R. (2001). Clarification, ellipsis and the nature of contextual updates in dialogue. Manuscript, King's College London and Göteborg University.

Green, B., Wolf, A., Chomsky, C., and Laugherty, K. (1961). BASEBALL: an automatic question answerer. In *Proceedings of the Western Joint Computer Conference*, pages 219–224. Reprinted in Grosz et al. (1986), *Readings in Natural Language Processing*. Morgan Kaufman.

Green, C. (1969). Theorem-proving by resolution as a basis for question-answering systems. In Meltzer, B. and Michie, D., editors, *Machine Intelligence*, volume 4, pages 183–205. Edinburgh.

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

Grice, H. P. (1969). Utterer's meaning and intention. *Philosophical Review*, 78(2):147–177.

Grice, H. P. (1975). Logic and conversation. In Cole, P. and Morgan, J., editors, *Syntax and Semantics III: Speech Acts*, pages 41–58. Academic Press.

Grosz, B. and Kraus, S. (1996). Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357.

Grosz, B. and Sidner, C. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12:175–204.

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

Grosz, B. J. and Sidner, C. L. (1990). Plans for discourse. In Cohen, P., Morgan, J., and Pollack, M., editors, *Intentions in Communication*, pages 417–444. MIT.

Haddock, N. (1989). *Incremental Semantics and Interactive Syntactic Processing*. PhD thesis, University of Edinburgh.

Heeman, P. and Hirst, G. (1995). Collaborating on referring expressions. *Computational Linguistics*, 21(3):351–382.

Hobbs, J., Stickel, M., Appelt, D., and Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, 63:69–142.

Howard, W. A. (1980). The formulae-as-types notion of construction. In *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus, and Formalism*, pages 479–490. Academic Press.

Joshi, A. K., Levy, L., and Takahashi, M. (1975). Tree adjunct grammars. *Journal of the Computer and System Sciences*, 10:136–163.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer.

Konolige, K. (1985). A computational theory of belief introspection. In *Proceedings of IJCAI*, pages 502–508.

Kripke, S. (1991). Presupposition and anaphora: Remarks on the formulation of the projection problem. Transcript of a lecture given at Princeton.

Lambert, L. and Carberry, S. (1991). A tripartite plan-based model of dialogue. In *Proceedings of ACL*, pages 47–54.

Larson, R. and Segal, G. (1995). *Knowledge of Meaning: An Introduction to Semantic Theory*. MIT.

Lascarides, A. and Asher, N. (1991). Discourse relations and defeasible knowledge. In *Proceedings of ACL*, pages 55–62.

Litman, D. J. and Allen, J. F. (1990). Discourse processing and commonsense plans. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*, pages 365–388. MIT.

Lochbaum, K. E. (1998). A collaborative planning model of intentional structure. *Computational Linguistics*, 24(4):525–572.

Malle, B. F. and Knobe, J. (1997). The folk concept of intentionality. *Journal of Personality and Social Psychology*, 33:101–121.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman.

McCarthy, J. and Buvač, S. (1994). Formalizing context (expanded notes). Technical Report STAN-CS-TN-94-13, Stanford University.

McCarthy, J. and Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In Meltzer, B. and Michie, D., editors, *Machine Intelligence*, volume 4, pages 473–502. Edinburgh.

Moore, R. C. (1985). A formal theory of knowledge and action. In Hobbs, J. R. and Moore, R. C., editors, *Formal Theories of the Commonsense World*, pages 319–358. Ablex.

Newell, A. (1982). The knowledge level. *Artificial Intelligence*, 18:87–127.

Piwek, P. (1998). *Logic, Information and Conversation*. PhD thesis, Eindhoven University of Technology.

Pollack, M. E. (1990). Plans as complex mental attitudes. In Cohen, P., Morgan, J., and Pollack, M., editors, *Intentions in Communication*, pages 77–103. MIT.

Pollack, M. E. (1992). The uses of plans. *Artificial Intelligence*, 57:43–68.

Pollack, M. E. and Horty, J. F. (1999). There's more to life than making plans: Plan management in dynamic, multi-agent environments. *AI Magazine*, 20(4):71–84.

Power, R. (1977). The organisation of purposeful dialogues. *Linguistics*, 17:107–152.

Schabes, Y. (1990). *Mathematical and Computational Aspects of Lexicalized Grammars*. PhD thesis, University of Pennsylvania.

Schuler, W. (2001). Computational properties of environment-based disambiguation. In *Proceedings of ACL*, pages 466–473.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge.

Searle, J. R. (1990). Collective intentions and actions. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*, pages 401–416. MIT.

Seidenberg, M. S. and MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23(4):569–588.

Sengers, P. (1999). Designing comprehensible agents. In *Proceedings of IJCAI*, pages 1227–1232.

Shanahan, M. (1997). *Solving the Frame Problem*. MIT.

Shoham, Y. (1988). *Reasoning about Change*. MIT.

Sperber, D., editor (2000). *Metarepresentations: an interdisciplinary perspective*. Oxford.

Sperber, D. and Wilson, D. (1986). *Relevance: Communication and Congition*. Harvard.

Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic*, 2(4):447–457.

Stokhof, M. and Groenendijk, J. (1999). Dynamic semantics. In Wilson, R. and Keil, F., editors, *MIT Encyclopedia of Cognitive Science*. MIT.

Stone, M. (1998). Abductive planning with sensing. In *Proceedings of AAAI*, pages 631–636.

Stone, M. (2000). Towards a computational account of knowledge, action and inference in instructions. *Journal of Language and Computation*, 1:231–246.

Stone, M. (2001). Representing communicative intentions in collaborative conversational agents. In Bell, B., editor, *Working Notes of the AAAI Fall Symposium on Intent Inference For Collaborative Tasks*.

Stone, M. and Doran, C. (1997). Sentence planning as description using tree-adjoining grammar. In *Proceedings of ACL*, pages 198–205.

Stone, M., Doran, C., Webber, B., Bleam, T., and Palmer, M. (2001). Microplanning with communicative intentions: The SPUD system. Manuscript, Rutgers University.

Stone, M. and Webber, B. (1998). Textual economy through close coupling of syntax and semantics. In *Proceedings of International Natural Language Generation Workshop*, pages 178–187.

Tambe, M., Adibi, J., Al-Onaizan, Y., Erdem, A., Kaminka, G. A., Marsella, S. C., and Muslea, I. (1999). Building agent teams using an explicit teamwork model and learning. *Artificial Intelligence*, 110:215–240.

Tanenhaus, M. and Trueswell, J. (1995). Sentence processing. In Eimas, P. and Miller, J. L., editors, *Handbook in perception and cognition: Speech Language and Communication*, pages 217–262. Academic Press.

Thielscher, M. (1999). From Situation Calculus to Fluent Calculus: State update axioms as a solution to the inferential frame problem. *Artificial Intelligence*, 111(1–2):277–299.

Thomason, R. H. (1990). Accommodation, meaning and implicature. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*, pages 325–363. MIT.

Traum, D. R. (1994). *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, University of Rochester.

van der Sandt, R. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics*, 9(2):333–377.

Wahlster, W., editor (2000). *Verbmobil: Foundations of Speech-to-speech Translation*. Springer.

Webber, B., Stone, M., Joshi, A., and Knott, A. (to appear). Anaphora and discourse semantics. *Computational Linguistics*.

Winograd, T. (1973). A procedural model of language understanding. In Schank, R. and Colby, K., editors, *Computer Models of Thought and Language*, pages 152–186. W. H. Freeman. Reprinted in Grosz et al. (1986), *Readings in Natural Language Processing*. Morgan Kaufman.

Woods, W. (1968). Procedural semantics for a question-answering machine. In *Proceedings of the AFIPS National Computer Conference*, pages 457–471.

**Notes**

1.  This paper benefits from feedback from audiences at the CUNY conference in Philadelphia, at the Rutgers Cognitive Science Center, at Rutgers Semantics Workshops, and at Johns Hopkins, Harvard, UCLA, USC, MIT and King's College London. Thanks to John Trueswell, Doug DeCarlo and Louis ten Bosch for detailed comments on the written version. The work described here was supported in part by NSF research instrumentation 9818322, by a sabbatical leave from Rutgers, and of course the organizers of CUNY 2001.

2.  As observed from quite different perspectives by (Geis, 1995; Seidenberg and MacDonald, 1999), pragmatics seems to invite this broadening of the traditional understanding of competence. Barry Schein, John Hale and Bob Frank independently observed the relevance of this issue to my research.

3.  Of course this idealization should not obscure the potential importance of such social language not only for cognitive science (Brown and Levinson, 1987) but even for human-computer dialogue (Cassell and Bickmore, 2002).

4.  Computational researchers' interest in task-oriented dialogue has persisted from early work on domain-specific question-answering systems (Green et al., 1961; Woods, 1968; Winograd, 1973) to such current efforts in spoken dialogue agents as (Allen et al., 1995; Ferguson and Allen, 1998; Wahlster, 2000).

5.  In formal psychology, these problems are regarded as part of a common *frame problem*. However, in the AI literature, the frame problem refers to a technical problem of axiomatizing temporal change compactly, for which numerous solutions are now available. So I stick to the more specific terms here. For the AI terminology, see particularly (Shoham, 1988).

6.  In this connection, it is instructive to regard intentions as meta-representations which duplicate or re-represent information available to one cognitive system so as to expose that information to other kinds of processing. See the papers in (Sperber, 2000).