Cognitive Science 28 (2004) 781–809

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

# Intention, interpretation and the computational structure of language

Matthew Stone

*Department of Computer Science and Center for Cognitive Science, Rutgers, the State University of New Jersey, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA*

Received 15 May 2003; received in revised form 24 November 2003; accepted 7 May 2004

Available online 22 July 2004

**Abstract**

I show how a conversational process that takes simple, intuitively meaningful steps may be understood as a sophisticated computation that derives the richly detailed, complex representations implicit in our knowledge of language. To develop the account, I argue that natural language is structured in a way that lets us formalize grammatical knowledge precisely in terms of rich primitives of interpretation. Primitives of interpretation can be correctly viewed intentionally, as explanations of our choices of linguistic actions; the model therefore fits our intuitions about meaning in conversation. Nevertheless, interpretations for complex utterances can be built from these primitives by simple operations of grammatical derivation. In bridging analyses of meaning at semantic and symbol-processing levels, this account underscores the fundamental place for computation in the cognitive science of language use.


*Keywords:* Dialogue; Pragmatics; Tree adjoining grammar

## 1. Introduction

Commonsense descriptions of conversation exhibit a continuity with our broader social understanding in emphasizing the intentional agency that underlies and explains our utterances. Imagine giving a commonsense account of what happens in dialogue (1), for example.

*E-mail address:* mdstone@cs.rutgers.edu (M. Stone).

0364-0213/$ – see front matter © 2004 Cognitive Science Society, Inc. All rights reserved.
doi:10.1016/j.cogsci.2004.05.001

(1) a *Bob:* What should I do next?
b *Alice:* Pass me the cake mix.
c *Bob:* Here you go [handing over the package].
d *Alice:* Thanks.

Such an account will portray (1) as a sequence of actions, each purposefully selected by its speaker to contribute to the ongoing joint activity. First, with utterance (1a), *Bob* poses a question, setting out the information that he intends to obtain in subsequent conversation. Then, with (1b), *Alice* offers an answer, and in so doing makes a request of *Bob*. *Bob* goes on to comply with the request, and offers utterance (1c) in tandem with the action to indicate what he is doing and why. Finally, with (1d), *Alice* accepts *Bob's* action, and acknowledges *Bob's* autonomy and generosity in choosing to carry it out.

This commonsense view of conversation forms the basis for a diverse array of work in cognitive science—what Clark calls the *action* tradition (1996, p. 56). The action tradition systematizes our intuitions about collaborative agency in language use, and documents the ways in which our utterances can serve to signal our intentions, to advance our common projects, and to cement our relationships with one another. These investigations offer elegant analyses of the meaning of utterances in conversation, such as (Grice, 1957, 1975; Lewis, 1979; Searle, 1969, 1975), and wide-ranging characterizations of the organization and functions of natural conversation, such as (Brown & Levinson, 1987; Sacks, Schegloff, & Jefferson, 1974). Indeed, they show that speakers seem to rely on their commonsense intuitions about utterances not only to work together in the world, as illustrated in (1), but also to negotiate a shared understanding of utterances, resolve miscommunication, and keep conversation itself on track (Clark, 1996). Moreover, we can use a computational framework to formalize these intuitions precisely and operationalize them consistently—such results not only corroborate collaborative analyses of language use but inspire researchers to take face-to-face conversation as a model for new kinds of human–computer interaction (Allen et al., 2001; Cassell, 2000; Rich, Sidner, & Lesh, 2001). These converging lines of research substantiate the power and the soundness of our understanding of utterances in conversation as intentional, collaborative actions.

If systematic study of language has substantiated our intuitive understanding of conversation, it has also challenged it. Starting with Chomsky (1957), generative approaches to grammar have accumulated evidence for a prodigious range of precise linguistic conventions that interlocutors appeal to in dialogue. We use this knowledge effortlessly, but have no direct conscious access to it. We see this already in (1). None of its utterances exhibits the canonical subject-verb-complement constituency of English sentences; instead, we find fragments—as in *Pass me the cake mix* or *Thanks*—and sentences with dislocated elements—as in *What should I do next?* or *Here you go*. These syntactic constructions vary in generality, from full productivity as in (1a) through the intermediate cases such as (1b) and (1c) to frozen expressions such as (1d), but all inherit aspects of their form and meaning through broader generalizations of English grammar (Goldberg, 1995; Ginzburg & Sag, 2002). At the same time, these constructions exhibit semantic and pragmatic specificity that motivates their use for particular messages in particular contexts (Birner & Ward, 1998; Levin, 1993; Prince, 1986). *Here you go*, for example, instantiates a distinctive construction that calls attention to an event of motion or discovery as meeting an established expectation.

The range and complexity of interlocutors' linguistic knowledge in turn challenges our understanding of the cognitive processes that support conversation. Experimental investigations of dialogue confirm our subjective impression that we recover one anothers' meanings in conversation quickly, automatically and incrementally. For example, Hanna and Tanenhaus (2004) used eye-tracking to investigate peoples' understanding of utterances like (1b) when uttered by a confederate experimenter in a collaborative interaction. Before they heard the noun phrase direct object of *pass*, subjects were already looking at the regions in space from which they would be expected to pass something. This shows that subjects were immediately able to link up the action described by *pass* with the requirements of the ongoing task. Brown-Schmidt, Campana, and Tanenhaus (2004) found evidence that participants in spontaneous dialogue rely on the same incremental coordination of linguistic and task constraints in understanding one anothers' references.

To explain our linguistic knowledge and abilities, it seems necessary to postulate specialized mechanisms that construct and manipulate distinctively linguistic representations. The study of such mechanisms and representations constitutes what Clark calls the *product* tradition in the cognitive science of language (1996, p. 56). The particularity of linguistic knowledge and processing, as emphasized by the product tradition, belies our apparent ability, emphasized by the action tradition, to understand our utterances in conversation in commonsense intentional terms. A central question in cognitive science is thus to reconcile the product and action traditions—to understand how our intuitions about what we say relate to our actual knowledge of language and the actual cognitive processes that underlie our linguistic behavior.

In this paper, I argue that the two traditions really do offer compatible perspectives on our participation in conversation. The resolution of the two perspectives depends on the distinctive *computational* structure of natural language. In particular, I invoke two distinctive computational principles—both of which are in fact epitomized in the pioneering research of Aravind Joshi. First, in the spirit of Joshi, Levy, and Takahashi (1975), and subsequent work including Schabes, 1990; Steedman, 2000b) and many others, we must describe natural grammar by atomic, meaningful elements with richly detailed internal structure but simple and sharply constrained combinatorics. Second, in the spirit of Joshi (1987) and related work including (Schabes & Shieber, 1994; Steedman, 2000b; Vijay-Shanker, 1987) and others, we must recognize the general duality between constructed linguistic representations—the *products* of computation—and the trace of processes that construct them—the *actions* of computation. These principles show how a conversational process that takes simple, intuitively meaningful steps may simultaneously be understood as a sophisticated computation that derives the richly detailed, complex representations implicit in our knowledge of language.

The distinctive contribution of this paper is to extend these computational principles so that they explicitly characterize *language use*. In contrast to models of grammar which pair each utterance with a representation of *logical form* which specifies its *semantic* content, to characterize language use we must link each utterance with a representation of how its speaker intends to *use* it. I have called such representations *pragmatic interpretations* (Stone, 2004). Here, I explore how these interpretations can be factored into rich atomic units that encapsulate syntax, semantics and pragmatics and that therefore characterize the full range of our linguistic knowledge as applied in specific conversational settings. Units of interpretation can still be combined by steps of grammatical derivation, and so can still enable simple, automatic, incremental interpretive processing. At the same time, constructed interpretations take the form of intention representations, and can be correctly understood as explanations of choices of words. So interpretive procedures can correctly be characterized intentionally, as *choosing* and *recognizing* abstract linguistic elements for their potential to contribute to the processes and goals of an ongoing conversation. Thus, there is no reason why we should not expect to find fast, special-purpose operations underlying our abilities to participate in conversation. Nevertheless, there is no reason why we should not expect our intentional intuitions about conversation to be reliable.

This paper can be understood as part of a broader theoretical program in cognitive science laying the groundwork for "a scientific psychology that validates belief/desire explanation" (Fodor, 1987, p. 16) via computational and representational mechanisms. All natural computation can be characterized abstractly, in terms of the problems it solves and the information it uses, or concretely, in terms of the representations and algorithms it implements (Marr, 1982; Newell, 1982; Pylyshyn, 1986). Formal, computational analysis plays a central role across cognitive science in demonstrating that specific explanatory mechanisms simultaneously respect these different characterizations. This paper epitomizes this framework. It argues the action and product perspectives on language seem to compete with one another, not because they make incompatible predictions, but because they use different vocabulary and different levels of abstraction to explain the same patterns of intelligent behavior.

The argument of this paper, however, relies not just on general representational approaches to psychological explanation, but on specific points of contact between specific analyses of collaborative rationality (Grosz & Sidner, 1990; Pollack, 1992) and specific theories of formal pragmatics (Beaver, 2001; Kamp & Rossdeutscher, 1994; van der Sandt, 1992). Bridging these disparate approaches exploits the availability—both for contributing to dialogue (Larsson & Traum, 2000) and for language processing (Stone, Doran, Webber, Bleam, & Palmer, 2003)—of simple, integrated mechanisms for conversational processing that support both semantic and operational characterizations. Some of these developments are not well-known outside computational approaches to linguistics. Certainly, they have never previously been considered as an integrated whole, as required for my argument here.

The argument proceeds in two steps. In Section 2, I use research on collaborative rationality to motivate symbolic structures that can serve as representations of communicative intentions. This account links our semantic intuitions about conversation with specific representations and processing that validates them. Then, in Section 3, I argue that these representations of intentions can be treated as linguistic representations. They can capture sophisticated syntactic, semantic and pragmatic constraints, and be manipulated through processes of grammatical derivation. Together, this construction shows how specialized linguistic processes can be analyzed semantically in commonsense intentional terms.

## 2. Meaning, intention and agency

The inspiration for this paper is research on agency at the intersection of philosophy and artificial intelligence. This tradition aims to develop *agent theories*, mathematical characterizations of collaborative rationality (Breiter & Sadek, 1996; Cohen & Levesque, 1990b; Grosz

& Sidner, 1990, 1996; Lochbaum, 1998). By combining the methods of philosophical analysis (Bratman, 1987; Bratman, Israel, & Pollack, 1988) with formal logic (Fagin, Halpern, Moses, & Vardi, 1995; Halpern & Moses, 1985), agent theories flesh out our intuitions of collaborative agency as consistent and explanatory descriptions of computational processes.

The scientific framework behind all agent theories is the representational theory of mind (RTM). As Fodor is quick to remind us, RTM is “the only game in town” (Fodor, 2000). RTM seeks to naturalize commonsense intentional explanations of human action by postulating symbolic representations and algorithms as a bridge between commonsense characterizations in semantic terms and scientific characterizations in physical terms (Fodor, 1987; Newell, 1982; Pylyshyn, 1986). In commonsense explanations, we explain our actions semantically, by appeal to the information that motivates them: a package of beliefs, commitments and desires that we take into account whenever we make a choice. This semantics explains our choices in terms of who we are as rational agents, as our identity inheres, in part, in the information we have about the world, the commitments we make to ourselves and one another, and the outcomes we value. By appealing to computation, we can also view this body of information operationally. Any of our beliefs, commitments and desires can be represented physically as symbolic structures, and we can give precise algorithms for mechanically deriving symbolic structures that motivate our actions from such representations. Logic brings an exact correspondence between semantic entailment and computational inference. Therefore, we can view our actions simultaneously as the exact consequences of a physical mechanism and as the exact manifestation of our identity.

While RTM provides the general theoretical background for agent theories, RTM leaves open the particular attitudes and causal connections that commonsense explanations actually attribute to agents that act collaboratively. Our intuitions still need analysis. Agent theories have made progress towards filling in this detail, by highlighting two important dimensions of our common-sense understanding of collaboration.[1] The first is the recognition that *intending* is a complex mental attitude in which an agent commits not just to planned action but also to a specific set of objectives and contingencies that motivate the action in context (Cohen & Levesque, 1990a; Pollack, 1990). The second is the recognition that *collaboration* requires agents to undertake a complex network of *mutual* commitments. In particular, teamwork depends on agreed instrumental intentions, but embraces a range of supporting communicative actions and intentions, which are required if agents are to perform their individual actions in concert. It also embraces mutual commitments that constrain agents’ behavior in unsuccessful attempts as well as when things go according to plan; this ensures that agents continue to work in concert as they repair or abort their ongoing efforts (Cohen & Levesque, 1991; Grosz & Kraus, 1996; Grosz & Sidner, 1990). Together, these insights give specific, computational content to the Grice’s (1957, 1969) proposal that a speaker’s meaning in using an utterance in conversation is on a par with the intentions that motivate other collaborative actions (Stone, 2004).

At the same time as agent theories have articulated more sophisticated understandings of rationality, they have developed more perspicuous ways to connect their analyses to possible symbol-processing realizations. Of course, these developments cannot improve on the expressive power of classic symbol-processing architectures such as production systems, which can already represent any possible computation (Newell, 1980). Rather, these developments offer abstractions in which agent theories are more readily understood and realized. A hallmark of these approaches is a general cycle of *perception* and *action* (Russell & Norvig, 1995; Wooldridge, 2000). The advantage of this abstraction is to isolate the computation involved in choosing actions as a separate process of *deliberation*, which reasons from the agent's current information state to initiate and pursue intentions as analyzed by agent theories. This information state offers a single structured representation that aggregates together all the factors that an agent must keep track of to act in the world, such as the agent's beliefs and future-oriented intentions. This makes it easy to describe the decision-making of the agent *declaratively*, simultaneously in terms of the information available to it in its environment and in terms of specific steps of processing.

Applying such specifications to dialogue leads to a general framework known as the information-state approach (Larsson & Traum, 2000; Matheson, Poesio, & Traum, 2000), which I adopt in this paper. These specifications focus on collaboration in shared environments, where utterances constitute the evidence and actions available to collaborating agents. Here, the information state describes public information, and so offers a formal counterpart to theoretical concepts of the discourse context or the conversational score (Clark & Marshall, 1981; Lewis, 1979; Stalnaker, 1998). Participants' preferences, meanwhile, are characterized by a private resource of goals and beliefs. The information-state approach focuses analysts' efforts on identifying the qualitative distinctions that distinguish the content of different dialogues, and on characterizing the course of dialogue through qualitative descriptions of the changing content available to interlocutors. This leads to information states that spell out the content that has been agreed so far over the course of the dialogue; the pending information that has been proposed by some participants in the dialogue, but not yet agreed by all; and the outstanding goals that remain still to be addressed.

The information-state approach comes with strong theoretical foundations (Poesio & Traum, 1997, 1998) that pave the way for principled and perspicuous analyses of the linguistic structures of natural dialogue in the context of models of agency. In particular, as I explain in Section 2.1, the framework allows us to justify specific representations that carry the content of intentions. Using intention representations and specifications of deliberative processes that use them, we can relate our collaborative intuitions about conversation to the kinds of mechanisms that might realize them. I provide such a specification in Section 2.2 and illustrate it in Section 2.3. The discussion thus situates and constrains possible linguistic processes in the context of a general cognitive architecture for collaborative conversation.

### *2.1. Specifying communicative intentions*

In the information-state approach, the effect of events in dialogue is to change, or *update*, the information state. Typically, we hypothesize that utterances instantiate a constrained inventory of possible actions, called *dialogue moves*, which can be parameterized by tokens of an underlying system of conceptual representation. With each kind of dialogue move comes a general specification describing how to update the information state in response to it. A specific move instantiated with specific concepts is called a *message*.[2] A message represents a specific contribution to conversation; however, to communicate a message requires formulating and manifesting a suitable communicative intention.

The experience of formal research in cognitive science suggests that messages must almost always exhibit a different level of granularity than appears in the typical utterances of conversation. In some cases, natural language utterances depend on an understanding of the world that goes beyond what the message itself contains—when language users formulate utterances, they can draw freely on a broad range of contextual information. In other cases, reasoning correctly about the world requires detail that natural language suppresses. In such cases, language users seem to exploit a recognition that their audience can work the details out for themselves.

To see this, let's give a simple, concrete example. Consider an interaction between agents *Alice* and *Bob*. *Bob* has just described a disappointing visit to the farmer's market, where he has purchased a number of prize apples (of unusual varieties) only to discover, upon arriving home, that one of them has been damaged. The conversation proceeds as in (2).

(2) a *Alice:* Which apple is bruised?
b *Bob:* The red apple is bruised.
c *Alice:* Too bad.

Consider the message behind (2b), as formalized, from *Bob's* perspective, by (3).

(3) *assert (Bob, marred-by-contusion (bobs-object-4))*

The frame *assert* indicates the type of dialogue move that is involved. What is needed is for *Bob* to add to and extend the common ground; the content to convey is the proposition that a specified object *bobs-object-4* (e.g., the apple) has a specified property *marred-by-contusion* (e.g., being bruised).

This message exhibits both kinds of mismatch. In (3), the apple is identified not by a description, but by an atomic symbol. To emphasize the commonality between representations like (3) and more general cognitive models of mental representation, it is probably best to think of such symbols as *deictic* or *indexical* representations, in the sense of Agre (1997) or better Pylyshyn (2000). These representations can refer to real-world objects in the agent's environment, but do so in virtue of the real relationship and interaction that the agent has with them. For example, such representations may be constructed and updated through preconceptual mechanisms that give them a robust perceptual connection with objects the agent sees. The name *bobs-object-4*, though technically arbitrary, suggests this relational, grounded content.

At the same time, the name *bobs-object-4* underscores that mental representations cannot be translated for other agents—agents cannot share perceptual relations to real-world objects. In our conversation, *Alice* will have her own mental representation, *alices-object-15* let us say, which, like *Bob's* representation *bobs-object-4*, refers to the actual apple in the context. The representations are private, and so are the indexical connections that link them to their common referent. To get other interlocutors to lock on to the same object in the world, language users must draw on additional communication knowledge, and construct a description that characterizes the intended referent through its distinctive and public attributes. Here, for example, *Bob* has to formulate a description corresponding to *bobs-object-4* that will lead *Alice* to recover *alices-object-15* as its interpretation. *Bob's* choice is *the red apple*.

In (3), the property is also represented as an atomic symbol: *marred-by-contusion*. Indeed, I assume that mental representations of properties ultimately have the same kind of relational, grounded content that mental representations of objects do; concept symbols get their meaning because through them we are locked on to their referents, namely kinds of things in the real world. See Fodor (1998). To suggest this content, we might characterize the symbol *marred-by-contusion* as picking out the property that *Bob* has seen in apples when their flesh is marred by a superficial region of damaged texture and color (as is characteristically produced by a blow or similar physical mistreatment). Note that this concept encodes an indirect, explanatory judgment rather than underlying perceptual features. This is characteristic of human concepts (Fodor, 1998).[3] Philosophy of language increasingly emphasizes that our understanding of the world naturally supports such distinctions, and that our pragmatic interpretations of utterances are correspondingly fine-grained; see e.g., Travis (1997).[4] Indeed, such conceptual abstraction and specificity is also a hallmark of formal knowledge representation, because consistency and precision are always crucial to the systematic development of useful and accurate knowledge; see Brachman et al. (1990).

Speakers need to identify properties, just as they do objects. *Alice* will have her own private representation, *spoiled-from-hitting* say, which refers to the same property as *Bob's* conceptual representation *marred-by-contusion*. *Alice's* representation will differ in indexing into and generalizing from *Alice's* real-world experience. *Bob's* utterance exploits knowledge of language—in this case, knowledge of the way English speakers use *is bruised*—to coordinate his conceptual representation with *Alice's*.

*If Bob's* utterance gets the message across, *Alice* and *Bob* will come to represent its content *isomorphically*. They will construct representations whose atoms have the same real-world reference, and are organized by corresponding compositional relationships. These isomorphic representations can be understood to specify the same objective propositions. Example (4), for instance, contrasts *Bob's* (4a) with *Alice's* isomorphic (4b).

(4) a *marred-by-contusion (bobs-object-4)*
    b *spoiled-from-hitting (alices-object-15)*

Both representations represent the proposition that *that* object, *Bob's* red apple, has *that* property, of being damaged as if by a blow. Observe, then, how language users, in coordinating such isomorphic representations, subject entities to much more precise predications than is evident in their linguistic descriptions. Thus, whereas the gap between *bobs-object-4* and *the red apple* suggests the ways in which messages may need to be elaborated in utterances, the gap between *marred-by-contusion* and *is bruised* suggests the ways in which conceptual content is a refinement of linguistic underspecification. It is to bridge these gaps that language users need complex communicative intentions.

As we shall see in the following, the communicative intentions underlying utterances such as (2b) instantiate the general form of action representations from agent theories. However, communicative intentions describe linguistic actions that typically would not be necessary for a single agent acting individually in its physical environment. In artificial intelligence, for example, the knowledge needed to describe linguistic action typically is not already implicit in the system's domain representations, and so has to be developed to extend the system to support dialogue interaction. These new descriptions of linguistic actions involve general communicative conventions that establish flexible links between a fixed, domain-independent vocabulary and the subject-matter of a particular conversation. It is convenient to organize these descriptions into two kinds of *resources*. Speakers' *linguistic* resources describe the form and meanings of utterances, and abstract away from specific connections to the subject-matter of a conversation. Speakers' *communicative* resources capture how to connect these utterances appropriately to that subject-matter. At the center of the distinction between the two are mismatches of granularity which place an item *W* from our general vocabulary (e.g., *is bruised*) in flexible correspondence with the various real-world concepts (e.g., *marred-by-contusion*) which speakers might be able to express by using *W*. Linguistic resources abstract away from speakers' concepts using a variable *C* linked with *W* through a constraint *W (C)*; communicative resources establish links between a specific concept *c* and *W* by supplying the instance *W (C)*. Building an utterance interpretation requires the processor to match the constraint with the corresponding instance.

Specifically, for *the red apple is bruised*, we require communicative resources to link the adjectival predication *is bruised* to the corresponding specific real-world concept here, *marred-by-contusion*. Analogously, we require communicative resources to link the words *apple* and *red* to corresponding concepts, too. *Bob's* representations for these resources might be presented as spelled out in (5).

(5) a *red (has-crimson-peel)*
b *apple (mcintosh-variety)*
c *bruised (marred-by-contusion)*

For communication, we assume that *Alice* has isomorphic representations, induced from her experience of language use in the same community.

For *the red apple is bruised*, a broad statement of its form and meaning is set out in (6), again from *Bob's* perspective. Note that the statement takes the form of a rule that applies to a range of objects, properties and other generalized individuals (Hobbs, 1985). These individuals are abstracted using variables (including higher order variables, so *A(X)* indicates that the property denoted by *A* holds of the entity denoted by *X*). The variables cannot take arbitrary values, however; they are restricted based on constraints that should be matched against communication knowledge and the dialogue context.

(6) a *The red apple is bruised* constitutes a possible utterance (under a suitable linguistic analysis).
b A speaker *U* can use this utterance with reference to any properties *R*, *A* and *B* and an individual *X*, constrained by communication knowledge so that *red (R)* and *apple (A)* and *bruised (B)*, and constrained by the context so that *A(X)* and *R(X)*.
c In this case, this utterance has the effect of contributing *B(X)* to the context, and so effects the move *assert (U, B(X))*.

Overall then, this description defines the utterance in (6a); it abstracts over the possible ways the speaker could use the utterance to portray relationships among objects and concepts in (6b) so as to pick out a *bruised* property *B* and a red apple *X*; and it characterizes the possible effects of the utterance in (6c) in accomplishing the speaker *U*'s move to assert that *X* has property

*B*. As we shall see further in Section 3, utterance descriptions such as (6) cannot be modeled holistically, but must be derived from a collection of atomic linguistic resources, stated in the principled terms of general knowledge of language.

Language users also need to appeal to the dialogue context. A general theory of dialogue context has to describe both the shared information and the coordinated attention that interlocutors maintain in conversation and rely on to understand one another. See Stone and Thomason (2002, 2003) for one such general approach. For simplicity in this paper, I emphasize the information in the context. In this case, the context has to characterize the apple using the same properties with which the utterance identifies it. That requires the presence of the facts in (7) in *Bob's* representation of the context.

(7) a *mcintosh-variety (bobs-object-4)*
b *has-crimson-peel (bobs-object-4)*

In general, accounting for utterances as collaborative actions means formalizing the *intentions* behind them (Allen & Perrault, 1980; Cohen & Perrault, 1979; Grice, 1957, 1969). Intentions are complex mental representations that explain why an agent should expect a planned action to bring about desired effects in the circumstances in which it is to be carried out (Pollack, 1990, 1992). In dialogue, in particular, speakers intend utterances to instantiate recognizable patterns of meaningful action: to link up with the context in specific ways, and to bring about specific, desired contributions to the dialogue. (8) spells out the content of such an intention, as it applies to *Bob's* use of (2b) to achieve message (3).

(8) a Since the context now supports
*red (has-crimson-peel)* $\wedge$ *apple (mcintosh-variety)*
$\wedge$ *bruised (marred-by-contusion)*

(as communication knowledge) and supports
*mcintosh-variety (bobs-object-4)* $\wedge$ *has-crimson-peel (bobs-object-4)*

(in the current information state),
b *Bob's* action of uttering *the red apple is bruised*, as analyzed in (17)
c must, in view of (6)
(with $U = Bob$, $R =$ *has-crimson-peel*, $A =$ *mcintosh-variety*,
$B =$ *marred-by-contusion* and $X =$ *bobs-object-4*)
d effect *assert (Bob, marred-by-contusion (bobs-object-4))*.

(8) is an argument or plan for *Bob* that *the red apple is bruised* is the right utterance. (8) starts from the assumed context and the assumed action. From these assumptions, it uses general knowledge of cause and effect to predict a desired result. For further discussion of such intention representations, including an analysis of the role of these representations in managing agents' commitments, and an argument that this role mitigates traditional challenges to logical approaches to mental representation such as the qualification and ramification problems, see Stone (2003, 2004). Note, of course, that such representations still allow us to recover a specific utterance that has been planned from an intention. I'll write this *a(i)*, the *action* envisaged in intention *i*. But such representations also allow us to recover, for example, the message that a specific utterance has been planned to convey: I'll write this *e(i)*, the *effect* envisaged in intention *i*.

```
loop {
    Perception
        e ← SENSE()
        i ← UNDERSTAND(r, c, e)
        c ← DIALOGUE-UPDATE(c, i)
    Deliberation
        m ← DIALOGUE-SELECT(p, c)
        i ← GENERATE(r, c, m)
    Action
        SAY(a(i))
        c ← DIALOGUE-UPDATE(c, i)
}
```

Fig. 1. A general, computational specification for a dialogue agent.

### 2.2. Specifying conversational agency

To specify conversational agency, we describe agents' coordinated reasoning about these 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, isomorphic representations of utterance meaning—as the dialogue proceeds. Even when communication is problematic, speakers can seek and provide evidence for mutual understanding by attempting to recognize and follow through on communicative intentions (Brennan, 1990; Brennan & Clark, 1996). Thus, this account makes a bridge with a broader collaborative perspective.

A representative specification of agency in dialogue is presented in Fig. 1. It specializes the general loop of agent theories to collaboration by *splitting* the notional processes of perception and deliberation. By introducing separate steps of understanding and update in perception, the new specification guarantees the agent's own actions and its partners' actions the same contextual effects. By splitting deliberation into separate steps of selection and generation, the new specification defines matched operations of understanding and generation that can coordinate their interpretive reasoning. Together these moves achieve *symmetry* between an agent's own actions, which it performs, and its collaborators' actions, which it now senses.

Because of the symmetry of collaborative agency, we have to present this cycle starting in the middle, with the deliberation that motivates all the actions of agents. In deliberation, the *dialogue selection process* draws on the information state $c$ together with the agent's private information and strategies $p$, and computes the agent's next move in the dialogue. In specifying the computational role of this process, we use an assignment operator ← to store the result of an invocation of DIALOGUE-SELECT in a variable $m$, as in (9):

(9) $m \leftarrow$ DIALOGUE-SELECT $(p, c)$

We assume that this next move corresponds to a specific message. See particularly Stent (2002).

The problem of communicating this message falls on the *generation process*. As we have seen, the generator depends on the *resource base* as well as the information state and the message. The generator draws on the specified resource base to construct an intention which represents a recognizable use of a natural language utterance to convey the specified message in the specified context. A generation process thus acts as a function GENERATE taking arguments *r* (for resources), *c* (for context) and *m* (for message) and returning the interpretation, giving us the specification:

(10) $i \leftarrow \text{GENERATE}\ (r, c, m)$

According to collaborative perspective on dialogue, the role of understanding mirrors generation; its task is to reason from a speaker's utterance to the intention behind it. Fig. 1 characterizes this process formally in the same way we characterized generation. In fact, the only difference is that where generation starts from an input message, understanding starts from an input utterance. Formally, the understanding process thus takes arguments *r*, *c* and *u*, and we specify an invocation of UNDERSTAND whose result is assigned to variable *i* as in (11):

(11) $i \leftarrow \text{UNDERSTAND}\ (r, c, u)$

Coordination in dialogue depends on whether a hearer, working in a specific context and with specific resources, can recognize the intention behind an utterance that the speaker has planned (ideally assuming the hearer shares the speaker's context and the speakers' linguistic and communicative resources). We have seen that a speaker's intention *i* in some utterance is a representation, like that in (8), obtained by GENERATE (*r*, *c*, *m*). Of this intention, the hearer is supplied only with the utterance *a(i)*. Successful recognition requires that the computation UNDERSTAND (*r, c, a(i)*) also results in *i*. Formally, then, the symmetry between generation and understanding required for coordination in dialogue is embodied (12).

(12) For all resources *r* and contexts *c*,
for any *i* obtained by GENERATE (*r*, *c*, *m*),
we also have $i = \text{UNDERSTAND}\ (r, c, a(i))$.

(12) offers a semantic constraint on the linguistic coordination between speaker and hearer; and as we shall see in Section 3, it is a constraint that we can precisely respect in operational characterizations of processes of generation and understanding.

The dialogue update operation of Fig. 1 is now solely responsible for keeping the information state current as events unfold in dialogue. Computationally, we model DIALOGUE-UPDATE as a function that takes as arguments the information state as a context *c* and the communicative intention *i* associated with an utterance, and computes a new context.

(13) $c \leftarrow \text{DIALOGUE-UPDATE}\ (c, i)$

The speaker uses DIALOGUE-UPDATE to anticipate the effects that he envisages for his utterance, and the hearer uses the very same procedure—ideally, with the very same arguments—to calculate the effects of her partner's utterance. Thus, it is the dialogue update process that gives content to the role of utterance interpretations as the kinds of collaborative intentions characterized by agent theories. In particular, it is the dialogue update operation that tracks the evolving common ground and the evolving mutual commitments of the interlocutors (Ginzburg, 1995; Larsson & Traum, 2000; Traum, 1994).

### *2.3. An example*

A systematic formal description of joint activity in conversation would involve a rich, multidimensional account of context in dialogue and how utterances change it (Bunt, 2000). In pursuit of such a general analysis, researchers have developed theories of aspects of the context that can account for much of the complexity of natural dialogue. For example, we can link the context to ongoing activity by systematically characterizing the background plans and domain coordination that interlocutors pursue, perhaps as Rich et al. (2001) propose. We can describe the evolving status of contributions to the dialogue itself through systematic models of dialogue obligations and grounding, perhaps as Larsson and Traum (2000) propose. And we can characterize the shared information and coordinated attention in the linguistic context systematically las a function of the linguistic elements and the linguistic relationships that make up dialogue (Asher & Lascardies, 2003; Ginzburg & Cooper, 2004; Stone & Thomason, 2002, 2003). Of course, substantial research remains in giving unified analyses of dialogue by exploring the connections among these compatible models. Presenting such analyses is beyond the scope of this paper, but see DeVault and Stone (2004) for one example.

I limit myself here to a suggestive sketch emphasizing the ability of this model to support intentional explanations of participation in dialogue. Consider interaction (2) between *Alice* and *Bob*. Let us assume the two interlocutors are similar, as in Cassell et al. (1994), Power (1977), and Sidner (1994). That is, while each agent brings a different set of private beliefs and goals to the interaction, the agents share common operations and a common context. Each agent follows the cycle in Fig. 1, but what one agent says is what the other agent hears and thus the two agents carry out these processes in complementary phases.

As background to the formal treatment, we assume update rules where *ask (Alice, q)* changes the context to introduce *q* as the principal question under discussion for *Alice*. Whatever the principal question under discussion is, the move *assert (Bob, r)* changes the context so that *r* is a pending proposed answer to it from *Bob*. Whatever proposal is pending (from the interlocutor other than *Alice*), the move *accept (Alice)* changes the context so this proposal is no longer pending but accepted into the common ground; if the proposal is an answer, the corresponding question is no longer pending, but becomes resolved. A detailed formal specification of these updates is beyond the scope of this paper—for more specifics about representing questions under discussion, proposals, and grounding in related formal analyses of dialogue, see (Ginzburg, 1995; Larsson & Traum, 2000; Traum, 1994; Traum & Allen, 1994). In this case, the successive moves, from *Bob's* perspective, are:

(14) a *ask (Alice, [which x* $\in$ *{bobs-object-3, bobs-object-4, ...}](marred-by-contusion (x)))*
 b *assert (Bob, marred-by-contusion (bobs-object-4))*
 c *accept (Alice)*

The representation of (14a) encodes an assumption that there is some salient set of apples that *Alice* is asking about; I give a partial enumeration of the elements of this set, including *bobs-object-3* and *bobs-object-4*, to facilitate further discussion. In what follows I will abbreviate the conceptual content of the question *Alice* asks simply as $q$.

From *Bob's* perspective, the dialogue should take *Alice* and *Bob* from an initial context $c_1$, through a context $c_2$ where the question $q$ is under discussion, through a context $c_3$ where the question $q$ is under discussion and *marred-by-contusion (bobs-object-4)* is a proposed answer to it, to a context $c_4$ where this question is resolved and *marred-by-contusion (bobs-object-4)* is common ground. To motivate these changes, *Alice* and *Bob* must also have suitable policies of selection. In $c_1$, *Alice* should select a move corresponding to (14a); perhaps *Alice* knows she needs to know about $q$ and will ask when the opportunity arises. In $c_2$, *Bob* should select (14b); perhaps *Bob* instantiates a general policy that retrieves information from his private knowledge $p_B$ by querying pending questions. In $a$, *Alice* should select *accept (Alice)*; perhaps *Alice* accepts any pending proposal unless her private information $p_A$ gives her a reason not to.

Based on this background, we can align the transcript of the dialogue between *Alice* and *Bob* with a trace of their steps of processing as in Fig. 2. Side-by-side comparison of the two traces shows the parallel operations that *Alice* and *Bob* undertake to coordinate the conversation. They synchronize on the same dialogue moves and respond to them by updating the context in the same way, according to their commitments in collaboration and the meanings they contribute to the conversation. The gaps in the table highlight the only asymmetries: these are the steps where only one agent is active, as that agent deliberates to choose the next move. The alternating pattern of gaps shows the complementarity of the agents' cycles of processing.

We can illustrate how the general coordination of the information-state approach carries over to linguistic utterances by looking at *Bob's* response in detail. As *Bob* plans the response, *Alice* and *Bob* are both in the context $c_2$ where the question $q$ of which apple is bruised is under discussion. *Bob* invokes DIALOGUE-SELECT ($c_2$, $p_B$) to identify the needed message *answer (Bob, marred-by-contusion (bobs-object-4))* as in (3). *Bob* invokes GENERATE, which realizes this message through a particular interpretation $i_2$, formalizing the content outlined in (8).

| *Alice* | *Bob* | result of computation |
|---|---|---|
| DIALOGUE-SELECT | | $ask(Alice, [which\ x \in \{\ldots\}](marred\text{-}by\text{-}contusion(x)))$ |
| GENERATE | | $i_1$ |
| SAY | SENSE | Which apple is bruised? |
| | UNDERSTAND | $i_1$ |
| DIALOGUE-UPDATE | DIALOGUE-UPDATE | $c_2$ |
| | DIALOGUE-SELECT | $assert(Bob, marred\text{-}by\text{-}contusion(bobs\text{-}object\text{-}4))$ |
| | GENERATE | $i_2$ as outlined in (8) |
| SENSE | SAY | The red apple is bruised. |
| UNDERSTAND | | $i_2$ |
| DIALOGUE-UPDATE | DIALOGUE-UPDATE | $c_3$ |
| DIALOGUE-SELECT | | $accept(Alice)$ |
| GENERATE | | $i_3$ |
| SAY | SENSE | Too bad. |
| | UNDERSTAND | $i_3$ |
| DIALOGUE-UPDATE | DIALOGUE-UPDATE | $c_4$ |

Fig. 2. Agents *Alice* and *Bob* are interlocutors meeting the specification in Fig. 1. The figure shows their steps of computation and the transcript as *Alice* asks a question and *Bob* answers it.

Associated with $i_2$ is the utterance *a*($i_2$) *the red apple is bruised*. *Bob* exchanges this utterance with *Alice*. Symmetrically, *Alice* invokes UNDERSTAND to infer $i_2$ from *a*($i_2$); now both *Alice* and *Bob* will invoke DIALOGUE-UPDATE ($c_2$, $i_2$) before proceeding with the dialogue. Fig. 2 hypotheses similar intentions $i_1$ and $i_3$ behind the other utterances in the interaction.

## 3. Generation and grammar

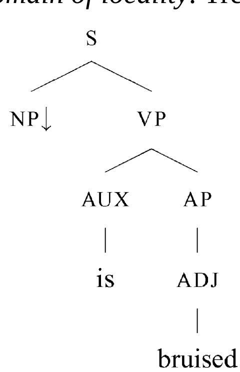
In the last section I showed how agent theories allow us to give a symbol-level account of our commonsense understanding of communicative intentions. Representations of interpretation, such as (8), are central to this account. In this section, we revisit these representations from a new perspective—the perspective of knowledge of language. In particular, I will use these intention representations to link our commonsense understanding of communication with processes that construct detailed and specialized linguistic representations. This provides the second half of the argument of the paper.

Central to this perspective is to analyze utterances as simple combinations of rich primitive elements. We can continue to regard each of these primitive elements as an action, as required by the general theory of dialogue. Nevertheless, as we see in Section 3.1, the description of each element is continuous with models of knowledge of language already in play in formal linguistics. Likewise, as we see in Section 3.2, the assembly of interpretation as a whole is continuous with computational processes of grammatical derivation. In this sense, algorithms for constructing an interpretation, such as that presented in Section 3.3, simultaneously reason about action and carry out grammatical derivation.

### *3.1. Deriving surface structure from complex primitives*

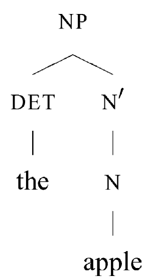
I follow tree-adjoining grammar, or TAG (Joshi et al., 1975; Schabes, 1990), and adopt a grammar formalism whose basic elements are not words or phrases but *tree fragments*. Each fragment includes a lexical item, together with a specification of all its syntactic arguments. Another way to say this is that the tree fragments are *lexicalized* and that they exhibit an *extended domain of locality*. Tree (15) illustrates both features.

(15)

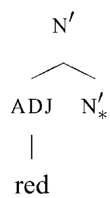



This tree is lexicalized in that it expands out to the content word *bruised* that determines the information conveyed by the construction as a whole. The tree has an extended domain of locality in that it encompasses not only the adjective but also the auxiliary verb *is* and a subject NP which remains to be specified. (The ↓ diacritic indicates a node where additional material must be supplied, called a *substitution site*.) This extended structure facilitates the statement of syntactic, semantic and pragmatic constraints by presenting a subject in the same atomic structure as the predicate that applies to it. The other trees I assume to analyze (2b) are given in (16); they exhibit the same principles.

(16) a

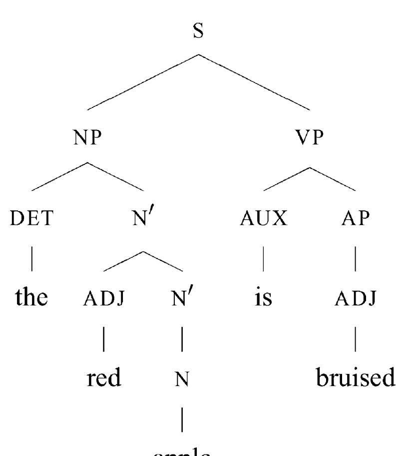



b




Such trees are combined together directly by grammatical operations in TAG. *Substitution* replaces a substitution site in one tree by a full tree fragment, and corresponds to syntactic complementation. *Adjunction* splices a tree fragment and its associated structure into an internal node in one tree, to create a larger constituent; the original subtree appears as a subtree of the new tree fragment at a distinguished node called the *foot node*. (The foot node is labeled (*), like the N′ node in the *red* tree.) This corresponds to syntactic modification. See Vijay-Shanker (1987) for more detail on derivation and structure in TAG. We can derive *the red apple is bruised* from the trees of (15) and (16) by first substituting tree (16a) at the NP node in (15), and then adjoining tree (16b) at the N′ node in (16a). That gives the overall structure in (17).

(17)



Extensive research has documented the suitability of this formalism for characterizing speakers' knowledge of linguistic structure. More detailed linguistic analyses in TAG exhibit the same kinds of structure and locality as the trees in (15) and (16), but typically involve more finely articulated syntactic constituency (Abeille & Rambow, 2000; Frank, 2002). For example, trees might include explicit projections for functional and morphological elements, and use syntactic features as part of the labels of nodes. Such ingredients figure prominently in the XTAG wide-coverage English grammar (Doran, Egedi, Hockey, Srinivas, & Zaidel, 1994; Doran, Hockey, Sarkar, Srinivas, & Xia, 2000).

What is significant here is that this formalism offers an explicit representation of the steps of assembly involved in deriving (17). This representation is called a *derivation tree*. The derivation tree shows the elements that are combined and the operations used to combine them, as in (18).

(18)

(15) *is bruised*
|
(16a) *the apple*, substitution at NP
|
(16b) *red*, adjunction at N′

### 3.2. Deriving interpretations from complex primitives

A derivation tree such as (18) is a dependency structure, tracing the action of a computational mechanism to add grammatical elements step by step to a growing syntactic structure. We can use derivation trees to account for the composition of an utterance as an array of choices of meaningful actions. In particular, we can characterize derivations that trace out *the construction of plans*, like (8), not just the composition of surface structures like (17). One way to do that is to *enrich our primitive objects* so that they become units of interpretation. Concretely, this means that the elements should now specify potential connections with conceptual representations; they should specify the assumptions about the context that they carry; and they should specify the effects that they can contribute to dialogue.

Consider then the revised elements in Fig. 3. Examination of the entries of Fig. 3 brings out both a close connection with interpretation and a continuity with broader accounts of knowledge of language.

The elements of Fig. 3 indicate connections between language and conceptual representations through the standard linguistic convention of indexing nodes to mark reference. A node P:*I* labels a phrase of syntactic category P as describing conceptual entity *I*. For example, the subject NP of *is bruised* gets the index *X* for the entity the NP describes (e.g., the apple). But in that tree the ADJ node is also indexed by *B* for the specific property of *being bruised* at issue. We often see such indices only on noun phrases, but the same descriptive correspondences are characteristic of all linguistic constituents; thinking of description crosscategorially becomes indispensable once one adopts the aim of linking language to other cognitive systems (Jackendoff, 1983, 1990). Extending indexing is also a natural computational step, as soon as we have a sufficiently sophisticated conceptual ontology (Hobbs, 1985). It is now a common idea in specifying

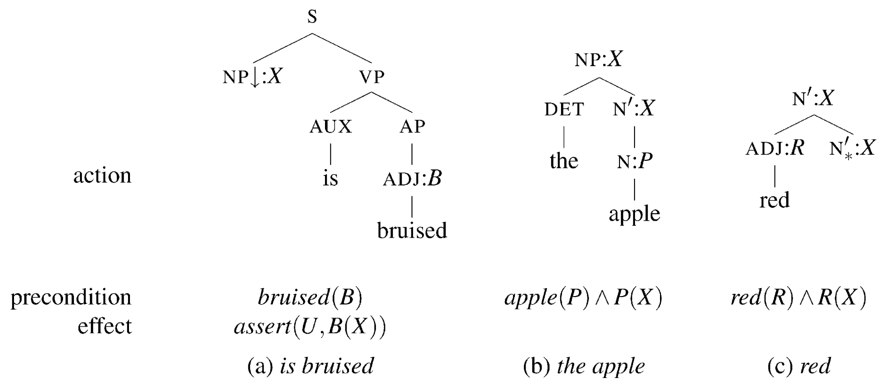



Fig. 3. Grammar as units of interpretation.

meaning in tree-adjoining-grammars (Joshi & Vijay-Shanker, 1999), and other computational grammar formalisms (Copestake et al., 2001). Gardent and Kallmeyer (2003) offer a particularly clear a demonstration of the comparable expressive power of unification-based TAG semantics and traditional approaches to compositionality based on the λ-calculus. Gardent and Kallmeyer's analyses span a range of core phenomena in semantics, including quantifiers and scope, intentionality, and control. The entries in Fig. 3 suppress out much of the detail of these semantic representations, just as they suppress much of the detail required for realistic syntactic representations. However, the formalism and mechanism remains general.

The elements of Fig. 3 come with a precondition—a link with the context that must be established whenever the element is used. Such links are known as *presuppositions* in semantic theory, which has been giving them increasingly prominent place in analyzing the meaning of discourse; see e.g. (Beaver, 2001; Kamp & Rossdeutscher, 1994; van der Sandt, 1992). Presuppositions capture constraints on felicitous use of grammatical forms. This makes presuppositions especially useful for pragmatic analyses, where constructions can mark presupposed open propositions (Prince, 1986) and the status of entities and information in the discourse (Birner & Ward, 1998; Steedman, 2000a).

Finally, elements of Fig. 3 come with a specification of how they change the context. In fact, an analysis of meaning in terms of context-change potential essentially recapitulates an analysis of meaning in terms of truth, while offering an attractive model of discourse interpretation; see (Dekker, 2002; Groenendijk & Stokhof, 1990; Muskens, 1996). By these assumptions, the preconditions and effects associated with actions in Fig. 3 faithfully realize broader semantic proposals.

In choosing to use any of the elements of Fig. 3, a speaker commits to realize a particular linguistic structure with specific intended reference. The speaker commits that the context supports the preconditions of the structure on its intended reference, and commits to bring about the structure's specified effects. In other words, the use of a particular structure with a particular intended reference commits the speaker to a particular intention—as schematized as in (19).

(19) a Since we now have contextual information supporting the preconditions of the element on its intended reference
b the speaker's action in uttering the element
c must, in view of the corresponding entry in the grammar,
d bring about the effects of the element on its intended reference.

Concretely, for example, consider an utterance of *is bruised* on the specific grammatical analysis given in Fig. 3a and with a specific intended pattern of reference indicated by $U = Bob$, $R = $ *marred-by-contusion*, $X = $ *bobs-object-4*. Instantiating the schema (19) determines an intention for the utterance of *is bruised*, as in (20).

(20) a Since we now have contextual information supporting *bruised (marred-by-contusion)*
b *Bob's* action in uttering *is bruised*, analyzed under the structure of Fig. 3a,
c must, in view of the grammatical entry Fig. 3a,
d bring about the effect *assert* (*Bob, marred-by-contusion* (*bobs-object-4*)).

This outlines the content of a commitment that a speaker would have to make to use *is bruised* with this pattern of reference. The intention in (20) would commit the speaker to performing the action, an utterance of *is bruised* as analyzed in Fig. 3a. The intention would commit the speaker to specific expectations about the context: it should support the conventional constraint that *is bruised* can express the property represented by *marred-by-contusion*. And the intention would commit the speaker is committing to making an assertion that *bobs-object-4* has this property.

An utterance of multiple elements in syntactic combination simply commits the speaker simultaneously to each of the corresponding component intentions. A corresponding derivation structure therefore shows how the speaker's intention in using a complete utterance is directly composed of elementary intentions in linguistic combination. Consider such a derivation structure, as in (21):

(21)

Figure 3a, *is bruised*, $U = Bob, B = $ *marred-by-contusion*, $X = $ *bobs-object-4*
|
Figure 3b, *the apple*, $A = $ *mcintosh-variety*, $X = $ *bobs-object-4*, substitution at NP
|
Figure 3c, *red*, $R = $ *has-crimson-blush*, $X = $ *bobs-object-4*, adjunction at N′

This derivation is the *trace* of a computational process that plans three steps of action. In each step, the computation proceeds by adding a specific lexical action to the planned utterance, thereby drawing on specific linguistic conventions and contributing specific content. This is an *action* perspective on the utterance. But what does the computation traced out in (21) do? It aggregates specific, fine-grained linguistic knowledge to *produce* a single complex mental representation that can guide language use. This complex representation is precisely the intention representation recorded as (8), which I repeat in the following.

(22) a Since the context now supports
*red* (*has-crimson-peel*) $\wedge$ *apple* (*mcintosh-variety*) $\wedge$ *bruised* (*marred-by-contusion*)

(as communication knowledge) and supports
*mcintosh-variety* (*bobs-object-4*) $\wedge$ *has-crimson-blush* (*bobs-object-4*)

(in the current information state),

b *Bob's* action of uttering *the red apple is bruised*, as analyzed in (17)

c must, in view of (6)
(with *U* = *Bob, R* = *has-crimson-peel, A* = *mcintosh-variety, B* = *marred-by-contusion* and *X* = *bobs-object-4)*

d effect *assert* (*Bob, marred-by-contusion* (*bobs-object-4*)).

The key insight of this paper, then, is that a computational process can derive representations like (22) through the systematic instantiation and assembly of linguistically motivated grammatical structures like those outlined in Fig. 3. A language user may therefore be characterized correctly as making incremental decisions of actions, as traced out in (21), and at the same time be characterized correctly as accessing detailed linguistic knowledge and integrating it into complex representations, as specified in (22).

### *3.3. Derivation, search and choice*

The development thus far has offered a reconciliation between views that characterize language use as collaborative action and views that characterize utterance construction as inference from knowledge of language. The demonstration opens the door to the development of explanations of utterances in conversation that combine the insights of both approaches.

I and my colleagues have been pursuing such analyses since implementing an intention-based model of grammatical derivation for natural language generation in the SPUD system (Cassell, Stone, & Yan, 2000; Stone, Bleam, Doran, & Palmer, 2000; Stone et al., 2003). The details of our system-building efforts require too much background to be successfully presented in the scope of this short article. I can, however, distill some of our experience in developing these case studies, and some of the results that might be forthcoming from comparable investigations in a computational cognitive science of language, by considering the problem of reference. How is it possible, in principle, for language users to integrate the disparate constraints of grammar and context to successfully identify objects in their environment to one another? With the computational analysis I have presented so far, we have no answer to this question. For example, although a formal grammar may specify an inventory of possible linguistic intentions, it will not thereby offer speakers any assurance that they will be understood. Reference is built into communicative intentions like those diagramed in (22).

Conversely, although a collaborative account can predict possible content for reference, it will not thereby offer speakers a useful guide to language use. This needs to be argued more carefully. By the collaborative principle, an identifying description must contain enough information to distinguish the intended referent from its salient alternatives in context. See Reiter and Dale (2000). However, when one explores grammar and logic systematically, in a computational setting, one discovers a plethora of options for content that would identify an object but do not correspond directly to the semantics of any natural utterance (Meteer, 1991; van Deemter, 2002). In general, one cannot transform these descriptions into correct logical forms while working purely at the level of conceptual semantics (Shieber, 1993). And even in constrained cases, when you know that a logical form has a derivation in the grammar, it can be computationally intractable to find it (Brew, 1992; Koller & Striegnitz, 2002). Collectively, this research provides a strong theoretical argument that *no* computational system can solve the problem of reference by *first* reasoning collaboratively to construct conceptual content and *then* reasoning linguistically to realize that content in words.

To resolve the paradox, we *have* to recognize that collaborative and linguistic reasoning describe a *single* process. Knowledge of language determines the inventory of moves available to the speaker at each step of planning a referring expression, as sketched thus far in Section 3. But the requirement that participants in dialogue coordinate, as sketched in Section 2, explains how the speaker makes the decision among these options.

To see how this might work in principle, recall the relationship between understanding and generation, repeated as (23) below:

(23) For all resources $r$ and contexts $c$,
for any $i$ obtained by GENERATE $(r, c, m)$,
we also have $i =$ UNDERSTAND $(r, c, a(i))$.

This relationship can be straightforwardly realized by incremental, coordinated reasoning. Specific proposals for this coordinated reasoning abound in the literature (Asher & Lascardies, 2003; Hobbs, Stickel, Appelt, & Martin, 1993; Sperber & Wilson, 1986). Ultimately, though, what is required is simply for processes of production to use (23) to guide the formulation of utterances.

As a provisional idealization, we assumed in SPUD that the understanding module can correctly recognize a schematic derivation structure, as in (18), from the use of any utterance. In other words, understanding handles lexical and syntactic ambiguity correctly. (This assumption can be relaxed; van Deemter, 2003.) However, understanding must still recover the intended reference of the utterance using shared linguistic and communicative resources and the shared information state of the dialogue. For example, in recognizing the intention behind *the red apple is bruised*, understanding starts from a schematic structure as provided in (24).

(24) a Since the context now supports
$red\ (R) \wedge apple\ (A) \wedge bruised\ (B)$

(as communication knowledge) and supports
$A(X) \wedge R(X)$

(in the current information state),

b *Bob's* action of uttering *the red apple is bruised*

c must, in view of (6)

d effect *assert* (*Bob*, $B(X)$)

By matching the constraints in (24a) against the context, understanding should be able to arrive at the same communicative intention that motivated the speaker's utterance. (This constraint-satisfaction view of coordination in dialogue has many other computational implications; Haddock, 1989; Mellish, 1985; Schuler, 2001.) Correspondingly, by anticipating this process of understanding, the speaker can work to construct an utterance sufficiently explicit that other interlocutors could be expected to recognize it. The speaker thereby ensures that (23) is met. In so doing, processes of language use seamlessly bridge grammatical derivation and collaborative reasoning.

As an example, we return again to the interaction *which apple is bruised—the red apple is bruised* explored generally in Fig. 2. We return to the context $c_2$ where the discussion has raised the question of which apple is bruised, and consider the stage of computation just after *Bob* has selected the message $m_2 =$ *assert* (*Bob, marred-by-contusion* (*bobs-object-4*)) as the next move to realize in this context. *Bob* must generate a recognizable utterance in context to convey this message.

Now to be more specific, we assume that there are in fact two apples in the context: the small mcintosh apple and a large golden delicious apple. Assume that *Bob* represents this second apple conceptually as *bobs-object-3*. Informally, here, *Bob* has to take coordination into account because a default description, *the apple is bruised*, say, will not distinguish between the two possible apples, and so cannot be expected to be understood as intended. The formalism that we have developed in this section allows us to restate this intuitive observation directly in terms of the linguistic knowledge and linguistic representations maintained in processes of language use.

To highlight the role of coordination in choice, we assume that *Bob* has in fact already committed to utter *the apple is bruised*. In light of the correspondence sketched in Sections 3.1 and 3.2, these commitments ground out in a specific intention representation, as outlined in (25).

(25) a Since the context now supports
*apple* (*mcintosh-variety*) $\wedge$ *bruised* (*marred-by-contusion*)

(as communication knowledge) and supports
*mcintosh-variety* (*bobs-object-4*)

(in the current information state),
b *Bob's* action of uttering *the apple is bruised*
c must, in view of (6)
(with $U =$ *Bob*, $A =$ *mcintosh-variety*,
$B =$ *marred-by-contusion*, $X =$ *bobs-object-4*)
d effect *assert* (*Bob, marred-by-contusion* (*bobs-object-4*)).

However, when *Bob* anticipates how this utterance will be understood, there is another possibility, as outlined in (26).

(26) a Since the context now supports
*apple* (*golden-delicious-variety*) $\wedge$ *bruised* (*marred-by-contusion*)

(as communication knowledge) and supports
*golden-delicious-variety* (*bobs-object-3*)

(in the current information state),
b *Bob's* action of uttering *the apple is bruised*

c must, in view of (6)
(with $U$ = *Bob*, $A$ = *golden-delicious-variety*, $B$ = *marred-by-contusion*, $X$ = *bobs-object-3*)
d effect *assert* (*Bob*, *marred-by-contusion* (*bobs-object-3*)).

In (26), we take the description *the apple* to refer to the golden delicious apple, not the mcintosh. By carrying out this assessment—by reasoning about coordination—*Bob* can recognize that the provisional intention from (25) is not yet satisfactory. Further action is required. But this assessment requires only the involvement of low-level processes of conversation: planning and interpretive processes that construct linguistic structures using simple operations of grammatical derivation.

Now the grammar allows *Bob* to incorporate an additional linguistic element into the provisional utterance, using the entry Fig. 3a for the adjective *red*. Extending the interpretation in tandem gives the structure of (22). Again, *Bob* must coordinate, by assessing the expected results of understanding. In this case, because the golden delicious apple is not in fact small, *Bob* can conclude that (22) can be recognized as intended.

In characterizing generation as a process of constructing intentions, we make it possible to treat generation as a planning problem. We can even implement generation using mechanisms for deliberation that are continuous with more general mechanisms for deciding what to do by refining intentions (Appelt, 1985; Heeman & Hirst, 1995; Pollack, 1992; Stone et al., 2003). The effect of this reasoning is to construct ways of using language—here, the referring expression *the red apple*—incrementally, using steps of derivation that add syntax, semantics and pragmatics simultaneously. This process never faces a problem of searching to realize predefined semantic content, and so defuses the many associated computational problems.

## 4. Conclusion

In this paper, I have sketched a direct computational role for the rules of language in planning contributions to collaborative conversation. The account is inspired and informed by implemented generation systems from computational linguistics, but ultimately responds only to the general constraints placed on theoretical cognitive models by our intuitions and skills in conversation.

The account assumes that the grammar specifies the form and meanings of the linguistic actions that speakers use in conversation as an inventory of atomic elements. These specifications offer precise and natural formulations of our knowledge of language. The generation process acts by selecting and combining these elements. As it proceeds, the process constructs detailed representations of utterance interpretation which map out the form of utterances, their links to context and the resulting updates to the information state of the conversation. Since the process proceeds through steps of grammatical derivation, the resulting representations are direct products of the system's knowledge of language. At the same time, what these representations do is to map out a speaker's commitments in using a planned utterance. Thus, they serve as a resource for action, deliberation and coordination throughout conversation. In this sense, we can explain the function of these representations, and the choices the generator makes to construct them, only in the context of a view of language use as collaborative action.

Compatibility between intentional and interpretive processing is an objective that not even all agent theories or plan-based generators attain. For example, in accounts of collaborative intention that specify agent's mental states and behaviors, like that of Cohen and Levesque (1991), the theory characterizes the commitments of collaborating agents through a direct specification of the underlying attitudes of the agent, as a flat list of possibly interrelated beliefs and goals. In such an approach, there may be no explicit place for richly structured abstract representations of the sort that we associate with grammatical derivation, which can summarize an agent's reasons to act. Conversely, models of linguistic inference that carry out staged reasoning about successive levels of representation, including the simple cascading pipeline commonly implemented in applied language generation systems (Reiter, 1994), may not easily be understood as constructing and coordinating plan representations. For example, such processes may first construct semantic representations from specifications of desired dialogue moves, then discard the dialogue moves, and go on to construct syntax from semantics using the grammar. Such a system need never have a unified pragmatic representation for its utterances. Considered on their own, such alternative views of collaborative agency and linguistic inference may be sensible. Nevertheless, to the extent that the theory of conversation can only account for dialogue by appealing to both kinds of explanation, we must ultimately prefer a unified framework.

The framework here, in particular, understands grammatical reasoning as implementing collaboration. There is only one process of language use, only one thing happening. Both theories describe that process, in different terms. If that is right, it raises a rich new set of questions about the specific interrelationships between the collaborative effort we see in dialogue and the linguistic competence that interlocutors call on to achieve it. To start, we must treat syntactic, semantic and pragmatic phenomena simultaneously with the precision we now find in standalone accounts of specific linguistic phenomena. I observed a number of directions for such integrations during the course of the paper. On the one hand, we need richer modeling of task context, perhaps along the lines suggested by Rich et al. (2001), as well as richer modeling of the linguistic context, perhaps along the lines suggested by Ginzburg and Cooper (2004). At the same time, we need more detailed accounts of linguistic syntax, along the lines suggested by Frank (2002) and more detailed accounts of linguistic semantics, perhaps along the lines suggested by Gardent and Kallmeyer (2003).

But we can also imagine genuinely new challenges emerging from linguistic and collaborative analyses of dialogue pragmatics. For example, dynamic approaches to semantics often start from a semanticist's notion of context change, in which utterances contribute propositions and objects that witness their truth. So far, theories classify such updates in terms of abstract categories, such as questions, answers or acknowledgments, by applying further communication knowledge (Asher & Lascardies, 2003; Poesio & Traum, 1997). It seems as though there are still two levels of representation, the linguistic and the collaborative. Indeed, the interpretations presented here, such as (8), continue to retain this duality; their results have been to generate dialogue moves, not directly to change the context. But what if the linguistic analysis of utterances did directly provide a characterization of context change, in terms of the pragmatic dimensions that determine interlocutors' options and obligations for the conversation? Such an analysis promises to allow us to simplify the explanations and implementations of planning utterances, and of operations of selection and update in dialogue. And such an analysis promises to align knowledge of language more tightly with the understanding of one anothers' actions we need to work together effectively. Attractive as it may be, the structure and principles of such an analysis remain very much open.

## Notes

1. It is important to underscore that while these insights have been earned in part through the successes and failures of artificial intelligence research, they are genuine features of our commonsense notions of intentionality (Malle & Knobe, 1997) and have proved indispensable ingredients of collaborative accounts of language use in the action tradition (Clark, 1996; Lochbaum, 1998).
2. Here I use the standard terminology from computational linguistics (Larsson & Traum, 2000; Reiter & Dale, 2000).
3. Even the judgment of color, as associated with the word *red* in particular, is notoriously indirect. To recover reflectance of a surface from incident light to the eye, we must correct for the way the object is illuminated. The reflectance we get is not color either, but combines a range of other material properties, including specularity and texture. Anyway our judgments of color describe objects not pixels: the red of an apple correlates the distribution of patterns across its surface with its variety and its state of ripeness. In (2), *red* is used to convey a property *has-crimson-blush* which indicates that the apple's exterior surface is covered with a crimson blush characteristic of certain kinds of apples, but not all kinds. When we take the goal of pragmatics to be the modeling of human language use, the complexity and abstraction of our background conceptual resources seems inescapable.
4. This is not necessarily to endorse or subscribe to any particular approach to the philosophical puzzles involved.

## Acknowledgments

This research reflects the influence and encouragement of my collaborators in dialogue research, particularly Justine Cassell, Aravind Joshi, Mark Steedman, Rich Thomason and Bonnie Webber. But this paper reflects my own perspective, so any errors are my own. This paper benefits from comments from anonymous reviewers at *Cognitive Science*, Vasil Daskalopoulos, David DeVault, Raymond Gibbs, Robert Goldstone, Paul Tepper, and audiences at the Carleton Semantics Conference, the University of Rochester and MIT. This research was supported in part by NSF HLC 0308121.

## References

Abeille, A., & Rambow, O. (Eds.). (2000). *Tree adjoining grammars: Formalisms, linguistic analyses and processing*. CSLI Publications.

Agre, P. E. (1997). *Computation and human experience*. Cambridge University Press.
Allen, J., Byron, D., Dzikovska, M., Ferguson, G., Galescu, L., & Stent, A. (2001). Towards conversational human-computer interaction. *AI Magazine*, *22*(4), 27–37.
Allen, J. F., & Perrault, C. R. (1980). Analyzing intention in utterances. *Artificial Intelligence*, *15*, 143–178.
Appelt, D. (1985). *Planning English sentences*. Cambridge, England: Cambridge University Press.
Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.
Beaver, D. (2001). *Presupposition and assertion in dynamic semantics.* Ph.D. Thesis, University Of Edinburgh (CSLI Publications, revision of 1995).
Birner, B. J., & Ward, G. L. (1998). *Information status and noncanonical word order in English*. Philadelphia, PA: John Benjamins Publishing.
Brachman, R., McGuinness, D., Schneider, P. P., Resnick, L. A., & Borgida, A. (1990). Living with CLASSIC: when and how to use a KL-ONE-like language. In J. Sowa (Ed.), *Principles of semantic networks*. Morgan Kaufmann.
Bratman, M. E. (1987). *Intention, plans, and practical reason.* Cambridge, MA: Harvard University Press.
Bratman, M. E., Israel, D. J., & Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence*, *4*, 349–355.
Breiter, P., & Sadek, M. D. (1996). A rational agent as the kernel of a cooperative spoken dialogue system. In J. P. Muller, M. J. Wooldridge, & N. R. Jennings (Eds.), *Intelligent agents III*. Springer.
Brennan, S. E. (1990). *Seeking and providing evidence for mutual understanding.* Ph.D. thesis, Stanford University.
Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*(6), 1482–1493.
Brew, C. (1992). Letting the cat out of the bag: Generation for shake-and-bake MT. In *Proceedings of COLING*, pp. 610–616.
Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language use*. Cambridge University Press.
Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2004). Reference resolution in the wild: Circumscription of referential domains by naive participants during an interactive problem solving task. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.
Bunt, H. (2000). Dialogue pragmatics and context specification. In H. Bunt & W. Black (Eds.), *Abduction, belief and context in dialogue* (pp. 81–150). Benjamin.
Cassell, J. (2000). Embodied conversational interface agents. *Communications of the ACM*, *43*(4), 70–78.
Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., & Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *SIGGRAPH* (pp. 413–420).
Cassell, J., Stone, M., & Yan, H. (2000). Coordination and context-dependence in the generation of embodied conversation. In *First International Conference on Natural Language Generation* (pp. 171–178).
Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.
Clark, H. H. (1996). *Using language.* Cambridge University Press.
Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge: Cambridge University Press.
Cohen, P. R., & Levesque, H. J. (1990a). Intention is choice with commitment. *Artificial Intelligence*, *42*, 213–261.
Cohen, P. R., & Levesque, H. J. (1990b). Rational interaction as the basis for communication. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 221–256). Cambridge, MA: MIT Press.
Cohen, P. R., & Levesque, H. J. (1991). Teamwork. *Nous*, *24*(4), 487–512.
Cohen, P. R., & Perrault, C. R. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science*, *3*(3), 177–212.
Copestake, A., Lascarides, A., & Flickinger, D. (2001). An algebra for semantic construction in unification-based grammars. In *ACL* (pp. 132–139).
Dekker, P. (2002). Pronouns in a pragmatic semantics. *Journal of Pragmatics*, *34*, 815–827.
DeVault, D., & Stone, M. (2004). Interpreting vague utterances in context. In *Proceedings of COLING*.

Doran, C., Egedi, D., Hockey, B. A., Srinivas, B., & Zaidel, M. (1994). XTAG System—a wide coverage grammar for English. In *Proceedings of COLING* (pp. 922–928).

Doran, C., Hockey, B. A., Sarkar, A., Srinivas, B., & Xia, F. (2000). Evolution of the XTAG system. In A. Abeille & O. Rambow (Eds.), *Tree adjoining grammars: Formal, computational and linguistic aspects* (pp. 371–403). CSLI Publications.

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about knowledge*. Cambridge MA: MIT Press.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.

Frank, R. (2002). *Phrase structure composition and syntactic dependencies*. Cambridge, MA: MIT Press.

Gardent, C., & Kallmeyer, L. (2003). Semantic construction in feature-based TAG. In *Proceedings of the 10th Meeting of the European Chapter of the Association for Computational Linguistics*.

Ginzburg, J. (1995). Resolving questions, I & II. *Linguistics and Philosophy, 18*(5; 6), 459–527, 567–609.

Ginzburg, J., & Cooper, R. (2004). Clarification, ellipsis and the nature of contextual updates in dialogue. *Linguistics and Philosophy, 27*.

Ginzburg, J., & Sag, I. A. (2002). *Interrogative investigations: The form, meaning, and use of English interrogatives*. Stanford, CA: CSLI Publications.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

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 P. Cole & J. Morgan (Eds.), *Syntax and semantics III: Speech acts* (pp. 41–58). New York: Academic Press.

Groenendijk, J., & Stokhof, M. (1990). Dynamic predicate logic. *Linguistics and Philosophy*, *14*, 39–100.

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

Grosz, B. J., & Sidner, C. L. (1990). Plans for discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 417–444). Cambridge MA: MIT Press.

Haddock, N. (1989). *Incremental semantics and interactive syntactic processing*. Ph.D. thesis, University of Edinburgh.

Halpern, J. Y., & Moses, Y. (1985). A guide to the modal logics of knowledge and belief: preliminary draft. In *Proceedings of IJCAI* (pp. 480–490).

Hanna, J., & Tanenhaus, M. K. (2004). The use of perspective during referential interpretation. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.

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

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

Hobbs, J. R. (1985). Ontological promiscuity. In *Proceedings of ACL* (pp. 61–69).

Jackendoff, R. S. (1983). *Semantics and cognition.* MIT Press.

Jackendoff, R. S. (1990). *Semantic structures*. MIT Press.

Joshi, A., & Vijay-Shanker, K. (1999). Compositional semantics with lexicalized tree-adjoining grammar (LTAG). In *International Workshop on Computational Semantics* (pp. 131–145).

Joshi, A. K. (1987). The relevance of tree adjoining grammar to generation. In G. Kempen (Ed.), *Natural language generation* (pp. 233–252). Dordrecht: Martinus Nijhoff Press.

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

Kamp, H., & Rossdeutscher, A. (1994). DRS-construction and lexically driven inference. *Theoretical Linguistics*, *20*, 97–164.

Koller, A., & Striegnitz, K. (2002). Generation as dependency parsing. In *Proceedings of ACL* (pp. 17–24).

Larsson, S., & Traum, D. (2000). Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. *Natural Language Engineering*, *6*, 323–340.

Levin, B. (1993). *English Verb Classes and Alternations: A preliminary investigation*. Chicago, IL: University of Chicago Press.

Lewis, D. (1979). Scorekeeping in a language game. In R. Bauerle, U. Egli, & A. von Stechow (Eds.), *Semantics from different points of view* (pp. 172–187). Berlin: Springer Verlag.

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

Malle, B. F., & 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.

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

Mellish, C. S. (1985). *Computer interpretation of natural language descriptions*. Chichester, UK: Ellis Horwood.

Meteer, M. W. (1991). Bridging the generation gap between text planning and linguistic realization. *Computational Intelligence*, *7*(4), 296–304.

Muskens, R. (1996). Combining Montague semantics and discourse representation. *Linguistics and Philosophy*, *19*(2), 143–186.

Newell, A. (1980). *Unified theories of cognition*. Harvard University Press.

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

Poesio, M., & Traum, D. (1998). Toward an axiomatization of dialogue acts. In J. Hulstijn & A. Nijholt (Eds.), *Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues (13th Twente Workshop on Language Technology)* (pp. 207–222).

Poesio, M., & Traum, D. R. (1997). Conversational actions and discourse situations. *Computational Intelligence*, *13*, 309–347.

Pollack, M. E. (1990). Plans as complex mental attitudes. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 77–103). Cambridge MA: MIT Press.

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

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

Prince, E. (1986). On the syntactic marking of presupposed open propositions. In *Proceedings of the 22nd Annual Meeting of the Chicago Linguistic Society* (pp. 208–222). Chicago. CLS.

Pylyshyn, Z. (2000). Situating vision in the world. *Trends in Cogntive Sciences*, *4*(5), 197–207.

Pylyshyn, Z. W. (1986). *Computation and cognition: Toward a foundation for cognitive science.* Cambridge, MA: MIT Press.

Reiter, E. (1994). Has a consensus NL generation architecture appeared, and is it psycholinguisti-cally plausible? In *Seventh International Workshop on Natural Language Generation* (pp. 163–170).

Reiter, E., & Dale, R. (2000). *Building natural language generation systems.* Cambridge University Press.

Rich, C., Sidner, C. L., & Lesh, N. (2001). COLLAGEN: Applying collaborative discourse theory to human-computer interaction. *AI Magazine*, *22*(4), 15–26.

Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Prentice Hall.

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

Schabes, Y. (1990). *Mathematical and computational aspects of lexicalized grammars*. Ph.D. thesis, Computer Science Department, University of Pennsylvania.

Schabes, Y., & Shieber, S. M. (1994). An alternative conception of tree-adjoining derivation. *Computational Linguistics*, *20*(1), 91–124.

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

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

Searle, J. R. (1975). Indirect speech acts. In P. Cole & J. Morgan (Eds.), *Syntax and semantics III: Speech acts* (pp. 59–82). New York: Academic Press.

Shieber, S. M. (1993). The problem of logical form equivalence. *Computational Linguistics*, *19*(1), 179–190.

Sidner, C. (1994). An artificial discourse language for collaborative negotiation. In *AAAI* (pp. 814–819).

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

Stalnaker, R. (1998). On the representation of context. *Journal of Logic, Language and Information*, *7*, 3–19.

Steedman, M. (2000a). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, *31*(4), 649–689.

Steedman, M. (2000b). *The syntactic process*. MIT Press.

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

Stone, M. (2003). Linguistic representation and Gricean inference. In *International Workshop on Computational Semantics* (pp. 5–21).

Stone, M. (2004). Communicative intentions and conversational processes in human-human and human-computer dialogue. In J. Trueswell & M. K. Tanenhaus (Eds.), *World-situated language use: Psycholinguistc, linguistic and computational perspectives on bridging the product and action traditions*. MIT Press.

Stone, M., Bleam, T., Doran, C., & Palmer, M. (2000). Lexicalized grammar and the description of motion events. In *TAG+5: Workshop on tree-adjoining grammar and related formalisms* (pp. 199–206).

Stone, M., Doran, C., Webber, B., Bleam, T., & Palmer, M. (2003). Microplanning with communicative intentions: The SPUD system. *Computational Intelligence*, *19*(4), 311–381.

Stone, M., & Thomason, R. H. (2002). Context in abductive interpretation. In *EDILOG 2002: Proceedings of the sixth workshop on the semantics and pragmatics of dialogue* (pp. 169–176).

Stone, M., & Thomason, R. H. (2003). Coordinating understanding and generation in an abductive approach to interpretation. In *DIABRUCK2003: Proceedings of the seventh workshop on the semantics and pragmatics of dialogue* (pp. 131–138).

Traum, D. R. (1994). *A computational theory of grounding in natural language conversation*. Ph.D. thesis, University of Rochester.

Traum, D. R., & Allen, J. F. (1994). Discourse obligations in dialogue processing. *ACL* (pp. 1–8).

Travis, C. (1997). Pragmatics. In B. Hale & C. Wright (Eds.), *A companion to the philosophy of language* (pp. 97–107). Oxford University Press.

van Deemter, K. (2002). Generating referring expressions: boolean extensions of the incremental algorithm. *Computational Linguistics*, *28*(1), 37–52.

van Deemter, K. (2003). OT for NLP: Towards a computationally useful notion of superoptimality. In *Fifth International Workshop on Computational Semantics*.

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

Vijay-Shanker, K. (1987). A *study of tree adjoining grammars.* Ph.D. thesis, Department of Computer and Information Science, University of Pennsylvania.

Wooldridge, M. (2000). *Reasoning about rational agents*. MIT Press.