*Because using language involves so much more than creating and comprehending isolated utterances, computers, like people, must accommodate linguistic and interpersonal context if they are to use language in a natural way.*

Janyce Wiebe

Graeme Hirst

Diane Horton

# Language Use in Context

Any text or dialogue establishes a linguistic context within which subsequent utterances must be understood. And beyond the linguistic context is the participatory context. A speaker or writer directs an utterance or text toward a hearer or reader with a particular purpose—to inform, to amuse, to collaborate in a task, perhaps. The form and content of the utterance are chosen accordingly, and the listener or reader must infer the underlying intent as part of their understanding.

This article explores recent research on language use in context, going beyond sentence boundaries and processing discourse—treating texts or dialogues as whole units composed of interrelated parts, not merely as sequences of isolated sentences. The article discusses the comprehension and production of language, looking at both texts and dialogues. A text to be processed might be, for example, a newspaper or magazine article being translated into another language or whose content is to be "understood" or abstracted in an information storage and retrieval system. A dialogue to be processed might be a conversation, spoken or typed, between a human and a computer, in service of some collaborative task. Many of the problems described here occur in both kinds of discourse. We use the words "speaker" and "writer," as well as "hearer" and "reader," almost interchangeably.

The underlying goal of the research described in this special section is to move beyond "toy" systems and come to grips with "real language." While the research described in the other articles in this section focuses on robustly processing massive amounts of text, the work described here focuses on understanding, in computational terms, the complexities and subtleties of language as people really use it.

In an article of this length, we cannot hope to describe all of the recent important work addressing language use in context. For example, we will not cover pronoun resolution, ellipsis, metaphor, or many aspects of belief ascription.

## Discourse Segmentation

Discourse has a rich structure. Sentences group together, and they can be related to one another in a variety of ways. Understanding the meaning of a discourse requires determining how the various pieces fit together. Consider the following excerpt from the introduction to a textbook on programming in C:

**Example 1**

1. One of the central goals of this text is to enable teachers to manage C's inherent complexity.
2. Managing complexity, however, is precisely what we do as programmers.
3. When we are faced with a problem that is too complex for immediate solution, we divide it into smaller pieces and consider each one independently.
4. Moreover, when the complexity of one of those pieces crosses a certain threshold, it makes sense to isolate that complexity by defining a separate abstraction that has a simple interface.
5. The interface protects clients from the underlying details of the abstraction, thereby simplifying the conceptual structure.
6. The same approach works for teaching programming.
7. To make the material easier for students to learn, this text adopts a library-based approach that emphasizes the principle of abstraction . . .[1]

When a person reads this excerpt, he or she gets more from it than just the individual meanings of the individual sentences. Understanding the rhetorical, or coherence, relationships among pieces of the text is an important aspect of understanding the text as a whole. For example, sentences 1.3 to 1.5 give specific details that expand on what is said in sentence 1.2. Moreover, text involves relationships at many levels; for example, not only are sentences 1.3 to 1.5 related to sentence 1.2, but they have relationships with each other as well. Fortunately, there are often cue phrases or textual markings to help the reader figure out the relationships; in spoken dialogue, intonation also helps. The word "however" in sentence 1.2, for instance, signals some sort of contrast with 1.1.

Readers must recover the structure of a discourse not only so they can infer the relationships among the pieces, but also because the structure constrains other essential aspects of understanding, such as figuring out what a pronoun refers to. The following conversation illustrates the constraining nature of such relationships:

**Example 2**

**A:** 1. Sheila wants you to call her about the bicycle.
**B:** 2. Has she found a roommate yet?
**A:** 3. Yeah.
4. Her old friend Linda is moving here from Waterloo to start a new job.
5. She moves in next week.
6. Anyway, you should phone her today.

The pronoun "her" in sentence 2.6 refers to Sheila, even though it was Linda who was referred to by "she" in the previous sentence. The structure of the dialogue helps the reader identify the correct referent. Sentences 2.2 to 2.5 interrupt the topic of discussion in 2.1, but sentence 2.6 returns to the topic. Notice that the cue word "anyway," possibly accompanied by a small pause, or change in pitch, gives a strong hint of this structure. The structure constrains the possible referents for the pronoun "her" in 2.6; the referent is more likely to come from sentence 2.1 than from 2.2 to 2.5. Thus, the perception of the structure of the discourse and the interpretation of pronouns constrain one another.

Relationships concerning the content of the text and relationships concerning the writer's or speaker's intentions are closely linked. In Example 1, the author had a reason for writing sentences 1.3 to 1.5, perhaps to clarify for the reader what was written in 1.2. The details given in 1.3 to 1.5 serve this purpose of the author.

There is evidence that people perform this kind of discourse segmentation during understanding. For example, Passonneau and Litman [19] and Hirschberg and Grosz [7] found statistically significant agreement among subjects asked to perform a discourse-segmentation task. How might a computer perform such segmentation or produce language from which such segments can be recovered? In their study, Hirschberg and Grosz also investigated the relationship between features of intonation, such as pitch range and timing, and the structure of the discourse. They found that, at both the local and global levels of discourse, there are statistically significant correlations between certain features of discourse structure and certain intonational features.

Another important kind of indication is cue phrases, such as "anyway," "however," "well," "still," "for example," and "now," which often provide explicit information about the structure of a discourse. For example, "now" can introduce a new subtopic [8].

---

[1] Eric S. Roberts, *The Art and Science of C: A Library-Based Introduction to Computer Science*, Addison-Wesley, 1995, page xv.

But many words that can serve as cue phrases also have other uses; the word "now," for example, can mean "at this time." To see how these ambiguities could be resolved, Hirschberg and Litman [8] studied cue-phrase use in speech, and developed a method for disambiguating cue phrases through intonational features of speech. They also discovered textual features of transcribed speech, such as punctuation, that are relatively easy to extract from transcriptions and can be used as additional aids in cue-phrase disambiguation. Much computational work in discourse processing has assumed that cue-phrase disambiguation is feasible; Hirschberg and Litman's findings provide empirical support for this assumption.

This research has applications in speech generation and speech understanding:

- Speech generation systems can use intonation and cue phrases as people do, helping break the discourse into appropriate segments, thereby assisting the listener in accomplishing other tasks crucial to understanding speech, such as determining the referents of noun phrases and recognizing the rhetorical and intentional relationships between segments.
- Speech recognition systems can use intonational features, cue phrases, and perhaps textual features (if the speech has been transcribed) to help perform segmentation and infer relationships between segments. AT&T Bell Laboratories' Text-to-Speech System [22], based on Hirschberg and Litman's work, does both; it disambiguates cue phrases in text on the basis of textual features, and then generates a spoken form of the text so the intonational features suggest how the cue phrases are being used.

### Relationships Within Discourse Segments

Investigating the structure within discourse segments, Hobbs et al. [10] and others suggest that during understanding, people make defeasible assumptions—assumptions that are consistent with what they believe, but which can later be overridden by contrary evidence. Such assumptions lead them to a plausible, coherent interpretation of the discourse. Zadrozny and Jensen [25] apply this approach holistically to paragraphs, seeking interpretations of individual sentences in a paragraph that are consistent together. Their approach formalizes the intuitive notion that the sentences in a paragraph are all related in some way to the same topic.

Lascarides, Asher, and Oberlander [14], whose approach is also founded on defeasible reasoning, address the inference of certain types of *coherence* relations [9] between segments and of *temporal* relations between the events that the discourse refers to. In the absence of information to the contrary, the default coherence relation between two sentences $s_1$ and $s_2$ that describe respectively events $e_1$ and $e_2$ is simple *narration*, in which case $e_1$ occurs before $e_2$. In the following example, one assumes that John's closing the door precedes his sitting down on the couch:

**Example 3**
1. John closed the door to the kitchen.
2. He sat down on the couch.

This discourse contrasts with the following example:

**Example 4**
1. John fell.
2. Max had pushed him.

We infer from Example 4 that Max's pushing John caused John to fall, so the event mentioned second precedes the first (notice that the tenses in Example 4 support this temporal interpretation); the coherence relation in this case is that sentence 4.2 is an explanation of 4.1. But people are sometimes sloppy in their use of tense (more technically, tense and aspect); one might have come to the same conclusion even if both sentences were in the simple past because of one's knowledge about pushing and falling.

**Example 5**
1. John fell.
2. Max pushed him.

Focusing only on background knowledge and ignoring tense, Lascarides et al. express defaults, such as those described here for narration, in a nonmonotonic logic. These are overridden if there is information to the contrary; in Example 5, the default is overridden by the specific knowledge that pushing someone causes them to fall. Their mechanism is also sensitive to the linguistic context; one could imagine a context for Example 5 in which it is taken to mean that John fell, and then Max pushed him. In this case, the relation is narration after all.

Hwang and Schubert [12] focus on interpreting tense to create a representation of the temporal relations among events described in the discourse, including implicit temporal relations across clause and sentence boundaries. To get an idea of what sorts of relations are recognized, consider Example 6:

**Example 6**
1. John went to the hospital.
2. The doctor told John that he had broken his ankle.

Hwang and Schubert's mechanism derives the following relations, among others: John's going to the hospital took place before the moment sentence 6.1 is said; the doctor's talking to John happened after John went to the hospital, but before the time when the sentences in Example 6 are said; and John's breaking his ankle took place before the doctor told him he broke it.

Hwang and Schubert's mechanism also works for longer narrative, provided that tense is used "literally." Consider the following modified version of an example in [13]. (The notation $t_e$ refers to the time of event *e*.)

**Example 7**
1. John went over to Mary's house. [$t_{goOver}$]
2. On the way, he had stopped by the flower shop for some roses. [$t_{stop}$]
3. He had picked out five red ones, three white ones, and one pale pink. [$t_{pickOut}$]
4. Then he had chosen a vase to put them in. [$t_{choose}$]
5. Unfortunately, they failed to cheer her up. [$t_{failToCheer}$]

Tense (technically, tense, aspect, and the aspectual classes of the events and states) implies the following temporal relations: time $t_{goOver}$ is before the time at which sentence 7.1 is said since 7.1 is in the simple past. The past perfect tense of "had stopped" in 7.2, "had picked out" in 7.3, and "had chosen" in 7.4 implies that $t_{stop}$, $t_{pickOut}$, and $t_{choose}$ are before the end of $t_{goOver}$. (Other relations are possible with the past perfect, but this issue does not concern us here.) Furthermore, with the simple past tense of "failed to cheer her up" in 7.5, we return to the time when John is at Mary's house. Sentences 7.2 to 7.4 constitute a subnarrative embedded in the overall narrative.

Of course, tenses are not always used as literally as they are in Example 7. It is quite natural for the simple past, rather than the past perfect, to be used once perspective shifts to the subnarrative, as in Example 8:

**Example 8**
1. John went over to Mary's house. [$t_{goOver}$]
2. On the way, he had stopped by the flower shop for some roses. [$t_{stop}$]
3. He picked out five red ones, three white ones, and one pale pink. [$t_{pickOut}$]
4. Then he chose a vase to put them in. [$t_{choose}$]
5. Unfortunately, they failed to cheer her up. [$t_{failToCheer}$]

Some of the tenses are different in Example 8 from Example 7, yet the temporal relations among events are the same. Notice that sentences 8.3 and 8.4 are in the simple past, like 8.5, yet 8.3 and 8.4 describe events in the embedded narrative, and 8.5 resumes the main narrative. Thus, as Hwang and Schubert and others discuss, tense alone is not sufficient in such cases to determine the temporal relations among events.

Kameyama, Passonneau, and Poesio [13] address this problem. Their approach is based on the idea, proposed by Webber [23] among others, that determining the time that a past tense refers to is similar to determining which entity a pronoun refers to, in that they both depend on things mentioned in the previous discourse. In sentence 9.2, the time referred to by the past tense ($t_{rideOff}$) depends on the event described in the previous sentence (*getOn*); in particular, $t_{rideOff}$ is after $t_{getOn}$, whenever $t_{getOn}$ might be.

**Example 9**
1. The ranger got on his horse. [$t_{getOn}$]
2. He rode off into the sunset. [$t_{rideOff}$]

In the terminology of Kameyama et al., past tense is understood with respect to a discourse reference time, which is established by the linguistic context. For the second sentence, the discourse reference time is $t_{getOn}$. In Example 10 [13], which follows, 10.3a and 10.3b are alternative continuations:

**Example 10**
1. John went over to Mary's house. [$t_{goOver}$]
2. On the way, he had stopped by the flower shop for some roses. [$t_{stop}$]

3a. He picked out five red ones, three white ones, and one pale pink. [$t_{pickOut}$]
3b. Unfortunately, they failed to cheer her up. [$t_{failToCheer}$]

Both continuations are in the simple past, yet the first is part of the embedded narrative, while the second is part of the main narrative. In the analysis of Kameyama et al., the past perfect "had stopped" introduces two discourse reference times, and a following past-tense sentence might be understood with respect to either one of them. The two times introduced by sentence 10.2 are $t_{goOver}$ (actually, a time inferred to be equal to the end of $t_{goOver}$) and $t_{stop}$; the past tense of 10.3a is understood with respect to $t_{stop}$, while the past tense of 10.3b is understood with respect to the end of $t_{goOver}$. Kameyama et al. discuss how to keep track of discourse reference times as the discourse proceeds, and how to choose the right one for a given past tense.

### Another Type of Segmentation

In texts, writers often report the mental states—beliefs, knowledge, intentions, hatreds, perceptions, and more—of various people. The most straightforward way to report someone's mental state is to present it explicitly with a sentence such as 11.2 in the following discourse:

**Example 11**
1. Stuart had accomplished his mission.
2. But he knew that by now the enemy was swarming to his rear.
3. To return the way he had come would invite trouble.
4. To continue on, to make a complete circuit around McClellan's army, might foil the pursuit.
5. Besides, it would be a glorious achievement.[2]

[2] James M. McPherson, *Battle Cry of Freedom*, Oxford University Press, page 463.

But mental states may also be presented implicitly, as in sentences 11.3 to 11.5. Example 11 is a passage from a nonfiction book about the American Civil War; in fact, Stuart did not turn back, but continued on. In sentences 11.3 to 11.5, the writer presents Stuart's motivations for doing so, even though the writer does not explicitly indicate he is presenting them. Thus, we have the problem of segmenting text according to whose beliefs and intentions are presented, a problem complicated by such implicitly presented mental states. Wiebe [24] developed an algorithm for performing this segmentation in third-person narrative texts. The algorithm is based on regularities—found through extensive examination of naturally occurring text—in the ways that writers manipulate point of view. For example, an explicit report of an agent's mental state can indicate that a block of sentences presenting that agent's mental states will follow, as in Example 11, but this indication does not typically hold if the sentence contains an expression of uncertainty or judgment toward the mental state; such textual markings suggest the point of view of either another person mentioned in the text or the writer. Expressions of uncertainty or judgement are similar to cue phrases because they help the reader perform point-of-view segmentation. An example is the phrase "It was almost as if" in sentence 12.1 in the following passage from the same book:

**Example 12**

1. It was almost as if he [Brown] knew that failure with its ensuing martyrdom would do more to achieve his ultimate goal than any "success" could have done.
2. In any event, that was how matters turned out. [3]

Even though sentences 11.2 and 12.1 both mention a person's knowledge, the hedge in 12.1 suggests that Brown's point of view does not continue in 12.2 as Stuart's does in 11.3. Sentence 12.2 does not present Brown's mental state, but describes the historical outcome. The absence or presence of a phrase such as the hedge in 12.1 is a textual feature that can help a text understanding system recognize implicitly presented mental states.

### The Speaker and the Hearer

Recent research explores the role played in discourse by the individual knowledge, goals, and experience of speakers, hearers, writers, and readers.

In advisory dialogues—in which an expert advises someone on assembling a device or improving a C++ program, for example—the hearer may lack the knowledge or experience necessary to fully understand the expert's explanations. In such cases, hearers often ask follow-up questions. A computer playing the role of the expert should be able to participate in a dialogue with the user, providing justifications for its recommendations, descriptions of its problem-solving strategies, and definitions of the terms it used. Moore and Paris [18] developed a text planner for advisory dialogues with these capabilities.

Moore and Paris integrate two main approaches to discourse in their work:

- One, discussed earlier in this article, focusing on the rhetorical or coherence relationships among the segments of the discourse; and
- An intentional approach, focusing on the intentions that motivate speakers' utterances and on the relationships among them.

With the intentional approach, generation is cast as the process of planning a sequence of utterances that achieve one's goals, and understanding is cast as the process of inferring the speaker's intentions from his or her utterances.

Most theories of discourse include both intentional and rhetorical knowledge to some extent. Moore and Paris focus on explicitly representing and using both kinds during processing.

The system's rhetorical knowledge, based on Mann and Thompson's Rhetorical Structure Theory [16], consists of strategies for achieving communicative goals through rhetorical links between discourse segments. For example, one way to achieve the goal of enabling the user to identify an object is to contrast it with an object already known to the user—through the contrast relation; another is to tell the user some of the attributes of the object—through the elaboration-attribute relation. Thus, more than one strategy might be able to achieve a particular goal, and a single strategy may be used in service of more than one type of goal.

Moore and Paris's system maintains a record of why it said what it said; when the user indicates that an explanation was not completely understood, it can determine which of its goals failed and attempt to achieve it again through a different strategy. Consider Example 13 [18]:

**Example 13**

**System:**

1. What characteristics of the program would you like to enhance?

**User:**

2. Readability and maintainability.

**System:**

3. You should replace (SETQ X 1) with (SETF X 1).
4. SETQ can only be used to assign a value to a simple-variable.
5. In contrast, SETF can be used to assign a value to any generalized-variable.
6. A generalized-variable is a storage location that can be named by any accessor function.

**User:**

7. What is a generalized-variable?

[3] ibid page 205.

**System:**

8. For example, the car of a cons is a generalized-variable named by the access function CAR, and the cdr of a cons is a generalized-variable named by the access function CDR.

The system has the goal of persuading the user to replace SETQ with SETF to enhance the readability and maintainability of his or her Lisp program. In sentences 13.4 and 13.5, it encourages the user to make the replacement by describing relevant differences between the object being replaced and the object replacing it. Having reasoned that the listener might not know what a generalized-variable is, the system explains the concept, stating its class membership (a storage location) and describing an attribute (it can be named by any accessor function). Yet in 13.7, the user expresses interest in knowing the definition of a generalized variable. The system already tried to satisfy this goal in 13.6; because the system explicitly represents its own goals, it can now realize that it did not succeed. It then selects an alternative strategy and tries again. In 13.8, it gives examples of generalized variables.

## Collaboration in Discourse

Although the work described here on explanations acknowledges the hearer in some ways, discourse is fundamentally collaborative in more ways. "The participants in a discourse work together to satisfy various of their individual and joint needs" [5, p. 418]. Most early work on inferring the intentional structure behind discourse did not consider this. Many theories modeled only situations in which one agent performs actions (both linguistic actions—utterances—and actions in the domain of discourse), while the system used these actions as a basis for attributing intentions or plans to the agent and was otherwise passive. In many cases, the system had no plans of its own, nor did it consider the agent might attempt to attribute any plans to it. There certainly were no joint plans. Furthermore, the agent's plans were presumed to be preformulated. All of these assumptions are incompatible with the collaborative nature of discourse. All agents involved in a discourse may have plans, they may all infer each other's plans, they often share joint plans, and their plans may be formulated on the fly. Indeed, the need to formulate a plan may be precisely why a user consults a system. Some recent work on plan inference has sought to embrace these facets of real discourse.

To model situations in which agents share joint plans, one must first ask what it means for two agents to have a joint plan. Grosz and Sidner [5] are concerned with precisely this question, extending to joint plans Pollack's earlier definition of having a plan [20], which was designed for single-agent plans. We paraphrase their definition as follows: Two agents have a shared plan to do action A if and only if for each subaction involved in doing A:

1. They mutually believe
   (a) That the subaction relates in a particular way to A;
   (b) That one of them can do that action;
   (c) That he intends to do it; and
   (d) That he intends, by doing it, to accomplish A; and
2. The agent who is to do the action:
   (a) Does in fact intend to do it; and
   (b) Intends, by doing it, to accomplish A.

While a shared plan is under construction, the two agents will hold only some of these intentions and mutual beliefs. At such times, they are said to have a partial shared plan. Although their definition raises a myriad of unresolved philosophical issues, Grosz and Sidner provide a good starting point for discussion.

How do the agents acquire the intentions and very strong mutual beliefs required to have a shared plan? Lochbaum, Grosz, and Sidner [15] tackle this problem. First they modify the definition in two important ways. Clause 1a is generalized to say the agents have a "recipe" for doing A. (A recipe for A, although defined formally, can be viewed simply as a way of doing A.) This change means that one definition suffices for any sort of shared plan, and also, since their notion of a recipe permits any level of detail, that agents can have a shared plan with any level of detail. Clauses 1d and 2b are modified to require not that the subaction accomplishes A, but that it merely contributes somehow to A; contribution is defined as the transitive closure of a number of basic relations between actions. By not requiring that the subaction contribute in any one particular way, the definition permits agents to have a shared plan that is vague in this regard. These modifications permit agents to share joint plans that are vague or lacking in detail; sharing is certainly important during plan construction when the plan has not yet been fully refined.

Lochbaum, Grosz, and Sidner offer an algorithm for inferring the mutual beliefs expressed in their improved versions of clauses 1a and 1d. The key part of the algorithm says that, in a context in which two agents have a shared plan to do A, if one of them says something about some action of type T (**need gamma**), the hearer may conclude that the speaker believes that T (**need gamma**) contributes somehow to A; furthermore, if the a hearer recognizes some way in which T (**need gamma**) does contribute to A, the hearer may conclude that the hearer and the speaker mutually believe that T (**need gamma**) contributes to A.

The TRAINS project, led by Allen and Schubert [1], also embraces the collaborative nature of discourse. The project's aim is to build a system that acts as a planning assistant, collaborating with the user to formulate plans that meet his or her goals. The system must be able to discuss goals and to form plans incrementally as the system and user interact. Building such a system requires creating and integrating components that handle problems as varied as parsing, reasoning about the domain, and reasoning about the beliefs, goals, and plans currently held by the system and the user.

The project's approach to discourse centers on planning and plan execution. It also takes an intentional approach, with utterances viewed as linguistic actions. Therefore, utterances can be treated in the same framework as domain actions, that is, planned for, executed (i.e., spoken or written), reasoned about (i.e., understood), and so on.

TRAINS provides a testbed for research on many problems in discourse, as well as in such other areas as temporal reasoning. The results of this research are being integrated into demonstration systems that participate in interesting dialogues, such as the one from which the following excerpt is taken. The domain of discourse is the shipment of commodities by rail:

**Example 14**

**User:**

1. We have to make orange juice. There are oranges at I and an orange juice factory at B. Engine E3 is scheduled to arrive at I at 3 P.M. Shall we ship the oranges?

**System:**

2. Yes. Shall I start loading oranges in the empty car at I?

The user does not explicitly propose that the oranges be shipped from I to B, using engine E3, yet the system infers this and can determine which oranges the user is referring to when saying "Shall we ship the oranges?" The system simultaneously answers the question and implicitly accepts the plan by replying "Yes." It then uses its knowledge of the domain to identify two possible ways to complete their now-mutual plan: one is to use a boxcar already located at I, and the other is to wait and use the boxcar that comes with the engine due at 3 pm. The system then poses a question to find out whether the first alternative is acceptable to the user.

## Fallibility of Conversants

Most research in language understanding has assumed a somewhat idealized notion of human linguistic abilities; people are viewed as faultless language processors whose skills AI research strives mightily to emulate. Even the work discussed in this article on inferring joint plans, though it brings speakers and hearers into full consideration, fails to address their fallibility. In fact, people are frequently unclear and imprecise in what they say and write, and as comprehenders, they frequently reach no understanding, or worse, a mistaken understanding. However, people make up for this by their flexibility. They are, for example, adept at detecting when a misunderstanding has set a conversation awry and at saying the right thing to correct it.

Misunderstandings in conversation might occur because the hearer takes an unintended sense of an ambiguous expression, because the hearer does not have the background knowledge needed to interpret the utterance or draw the expected inferences from it, or simply because of an error in typing or speech recognition. If the result is no interpretation at all or several possible interpretations from which a choice cannot be made, the hearer can ask for clarification, can remain silent (hoping that subsequent utterances resolve matters), or can invoke additional processing to try to recover. Eller and Carberry [4] take the third approach; the interpretation of an utterance that cannot be coherently integrated into the current context is *relaxed* by a set of heuristics. Relaxation permits the system to consider a somewhat unlikely shift of focus, for example, or an imprecise use of tense.

But if a conversant finds a single reasonable, albeit erroneous, interpretation, the misunderstanding manifests itself later, if at all, when the conversants find themselves talking at cross purposes. There are thus two parts to the problem:

- Noticing that there has been an earlier misunderstanding—either by oneself or by the other conversant; and
- Generating an utterance that will repair the misunderstanding.

For the first part, Eller and Carberry suggest that if these heuristics do not serve to interpret a problematic utterance, the cause might be an earlier misunderstanding and so apply the heuristics to earlier utterances, creating alternative contexts in which the current utterance can be considered. They do not address the second part.

McRoy and Hirst [17] have shown that both parts of the problem can be accounted for in a model of conversation in which the interpretation of an utterance is characterized as abductive reasoning (for example, given $Q$ and $P \Rightarrow Q$, one may guess that $P$ is also true) and the generation of an utterance as default reasoning [10]. In this model, conversants abductively form defeasible expectations as to the kind of utterance likely to occur next in the conversation, and use these expectations to monitor for differences in understanding. When speakers utter something, they use their beliefs about the discourse context, about the other participant's beliefs, and about conventions of discourse to select an utterance appropriate to their goals. The other participant attempts to retrace this selection process abductively, trying to identify the goal, expectation, or misunderstanding that might have led the speaker to produce it. If McRoy and Hirst's model finds more than one possibility, it chooses one at random, the model not accounting for differing likelihoods of the various interpretations. If a misunderstanding on either side, self or other, is found, a conversation participant will re-interpret earlier utterances to find another interpretation and utter an appropriate correction.

The model is implemented in an extension of Prolog that performs abduction and default reasoning. The model includes both interpretation and generation, so two copies of the program with different beliefs and knowledge can converse with one another.

### Nuance and Style in Language

The exact choice of words, phrases, and sentence structure all affect the precise meaning and effect of an utterance. A writer or speaker chooses (consciously or not) such goals as whether to be formal or friendly, persuasive or dismissive, clear or obscure. These aspects of an utterance are as much a part of its message as its literal meaning, and any sophisticated natural language system needs to be sensitive to them. A machine translation system, for example, would be inadequate if, in translating a business letter from English to French, it preserved the literal meaning but turned a friendly letter into a threatening one or vice versa.

Indeed, the selection of what is to be said at all depends on complex interpersonal concerns. One might choose material that supports one's own position or that might appeal to the listener, while omitting material that undermines one's position or that might annoy a listener whom does not wish to offend. Hovy's natural-language generation system PAULINE [11] was the first to account in an integrated manner for interpersonal concerns in linguistic nuance and in the selection of material. Although PAULINE was impressive, it had no theoretical basis; it employed a wide variety of rules that were little more than an ad hoc collection of heuristics. Subsequent research has sought more general principles for the selection of content, words, and syntactic structures.

One particular problem is the construction of referring expressions—words or phrases a speaker uses to denote some particular object or entity. The problem is there are usually many ways to do this, and in any given situation, some are better than others. For example, any of the following might serve to fill the space in the sentence shown:

**Example 15**
I'll meet you near______in an hour.

1. the tree
2. the tall tree
3. the larch
4. the tree with the soft, light-green needles
5. the tall conifer next to the old well
6. the larch that is about 50 feet tall
7. the big one
8. it

Which alternative is best depends on the previous utterances. For example, the last two options require a previous referring expression as an antecedent, as well as the listener's assumed knowledge and the circumstances of the utterance. Alternative 15.3 is no good if the listener doesn't know enough about trees to identify a larch, though 15.4 might serve. If there is only one tree the speaker could possibly be referring to, detailed descriptions, such as 15.4 to 15.6, are misleading, as they spuriously imply that some kind of contrast is being made.

Dale [2] and Reiter [21] have developed methods for constructing referring expressions. The first consideration is whether a pronominal reference is possible; if not, a definite noun phrase must be constructed. Such a noun phrase must be both efficient and adequate, identifying the referent with the least amount of information necessary to do so unambiguously. Dale uses the notion of minimal distinguishing description—the smallest set of attributes and their values that will serve to discriminate the referent from other entities. These methods also account for the preference in language for descriptions that use so-called basic categories. For example, in the context "Will you please take the_______for a walk?", the word "animal" might be sufficient to uniquely identify the thing to be taken for a walk, but "dog" is still the more natural expression. Reiter points out that using an inappropriate category, or a description more complex than necessary, creates a false implicature, and he presents an algorithm for generating referring expressions that is able to avoid such situations.

The problem is somewhat different in interactive discourse because, if a referring expression fails to pick out a unique entity, the participants can immediately try to correct it; they can collaborate on the task of reaching a common understanding of the reference. Heeman and Hirst [6] model this computationally as the construction and recognition—by two agents with possibly differing beliefs—of plans to refer to something. In this model, the generation of a referring expression is viewed as the construction of a plan to bring the referent to the attention of the other conversant; comprehension of a referring expression is viewed as recognition of this (possibly faulty) plan. The model accounts for both production and comprehension of referring expressions. Thus, as with the model of McRoy and Hirst, two copies of the program with different beliefs can talk to one another, negotiating a common understanding of a referring expression.

Much of the work on language generation, including that on referring expressions, seeks to determine the content to be expressed. For example, in Dale's system, a syntactic structure is chosen at random to express the content of the description; for example, either "the pitted olives" or "the olives that have been pitted." But, as Hovy's work shows, form is equally important. DiMarco and Hirst [3] sought to correlate a writer's use of various syntactic constructions with his or her higher-level stylistic goals to ensure that an automatic translation retains these goals even if that requires a different syntactic structure in the target language. For example:

**Example 16**
Ils se livrent alors, sous des dehors irrésistibles de drôlerie, à une lutte sournoise et passionnée.

**Example 17**
And from behind a cover of irresistibly funny wit, they open fire in an artful and passionate battle.[4]

The underlined phrase that interrupts the main clause in Example 16 was moved by the translator to a position before the main clause in Example 17. Though Example 16 is quite natural in French, the movement of the phrase was necessary to prevent what would otherwise be a somewhat unnatural sentence in English. To capture this kind of linguistic intuition, DiMarco and Hirst developed the idea of a *grammar of style*, which correlates the syntactic structures of a language with a set of language-independent stylistic goals. In translation, these goals can then be determined in the source text and used in the generation of the new text.

## Conclusion

The research discussed in this article is generally concerned with the full complexity of language as people use it, particularly the complexity of creating and comprehending units larger than a single sentence, determining the best word or expression to fit the present context and intent, and accommodating the fallibility of language users. Although the results of some of this work have already been incorporated into application systems, much of it is exploratory, basic research, a necessary precursor to the development of practical systems.

**Acknowledgments**
We wish to thank Yorick Wilks, Lisa Smith, and an anonymous referee for helpful comments on earlier drafts of this article. The preparation of this article was supported in part by a grant to Graeme Hirst by the Natural Sciences and Engineering Research Council of Canada.

[4] This example is from the bilingual Programme du Festival du Cinéma Américain, Festival de Deauville, 1976, quoted by Jacqueline Guillemin-Flescher, (*Syntaxe Comparée du Français et de l'anglais: Problèmes de Traduction*, Éditions Ophrys, 1981, p. 125), who explains some of the differences in the naturalness of various structures in English and French.

**References**

1. Allen, J. F., Schubert, L. K., Ferguson, G., Heeman, P., Hwang, C. H., Kato, T., Light, M., Martin, N. G., Miller, B. W., Poesio, M., and Traum, D. The TRAINS project: A case study in building a conversational planning agent. *Journal of Experimental and Theoretical Artificial Intelligence 7,* 1 (January-March 1995), 7–48.
2. Dale, R. *Generating Referring Expressions.* The MIT Press, Cambridge, Mass., 1992.
3. DiMarco, C. and Hirst, G. A computational theory of goal-directed style in syntax. *Computational Linguistics 19,* 3 (September 1993), 451–499.
4. Eller, R. and Carberry, S. A meta-rule approach to flexible plan recognition in dialogue. *User Modeling and User-Adapted Interaction 2,* 1–2 (1992), 27–53.
5. Grosz, B. J. and Sidner, C. L. Plans for discourse. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, eds. *Intentions in Communication.* The MIT Press, Cambridge, Mass., 1990, 417–444.
6. Heeman, P. and Hirst, G. Collaborating on referring expressions. *Computational Linguistics 21,* 3 (September 1995), 351–382.
7. Hirschberg, J. and Grosz, B. Intonational features of local and global discourse structure. In *Proceedings of the Speech and Natural Language Workshop* (February 23–26, Harriman, N.Y.). DARPA, Software and Intelligent Systems Technology Office, 1992, 441–446.
8. Hirschberg, J. and Litman, D. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics 19,* 3 (September 1993), 501–530.
9. Hobbs, J. R. Coherence and coreference. *Cognitive Science 3,* 1 (January–March 1979), 67–90.
10. Hobbs, J. R., Stickel, M., Martin, P., and Edwards, D. Interpretation as abduction. In *Proceedings of the 26th Meeting of the Association for Computational Linguistics* (June 7–10, Buffalo, N.Y.). Association for Computational Linguistics, N.J., 1988, 95–103.
11. Hovy, E. H. *Generating Natural Language under Pragmatic Constraints.* Lawrence Erlbaum Associates, Hillsdale, N.J., 1988.
12. Hwang, C. H. and Schubert, L. K. Tense trees as the "fine structure'' of discourse. In *Proceedings of the 30th Meeting of the Association for Computational Linguistics* (June 28–July 2, Newark, Del.). Association for Computational Linguistics, N.J., 1992, 232–240.
13. Kameyama, M., Passonneau, R., and Poesio, M. Temporal centering. In *Proceedings of the 31st Meeting of the Association for Computational Linguistics,* (June 22–26, Columbus, Ohio). Association for Computational Linguistics, N.J., 1993, 70–77.
14. Lascarides, A., Asher, N., and Oberlander, J. Inferring discourse relations in context. In *Proceedings of the 30th Meeting of the Association for Computational Linguistics* (June 28–July 2, Newark, Del.). Association for Computational Linguistics, N.J., 1992, pp. 1–8.
15. Lochbaum, K. E., Grosz, Barbara J., and Sidner, C. L. Models of plans to support communication: an initial report. In *Proceedings of the 9th National Conference on Artificial Intelligence* (July

29–August 3, Boston, Mass.). American Association for Artificial Intelligence, Calif., 1990, vol. 1, 485–490.
**16.** Mann, W. and Thompson, S. Rhetorical structure theory: toward a functional theory of text organization. *Text 8*, 3 (1988), 243–281.
**17.** McRoy, S. and Hirst, G. The repair of speech act misunderstandings by abductive inference. *Computational Linguistics 21*, 4 (December 1995).
**18.** Moore, J. and Paris, C. Planning text for advisory dialogues: Capturing intentional and rhetorical information. *Computational Linguistics 19*, 4 (December 1993), 651–694.
**19.** Passonneau, R. and Litman, D. Intention-based segmentation: Human reliability and correlation with linguistic cues. In *Proceedings of the 31st Meeting of the Association for Computational Linguistics*, (June 22–26, Columbus, Ohio). Association for Computational Linguistics, N.J., 1993, 148–155.
**20.** Pollack, M. E. A model of plan inference that distinguishes between the beliefs of actors and observers. In *Proceedings of the 24th Meeting of the Association for Computational Linguistics* (June 10–13, New York, N.Y.). Association for Computational Linguistics, N.J., 1986, 207–214.
**21.** Reiter, E. Generating descriptions that exploit a user's domain knowledge. In Robert Dale, Chris Mellish, and Michael Zock, eds. *Current Research in Natural Language Generation.* Academic Press, London, 1990, 257–285.
**22.** Sproat, R, Hirschberg, J, and Yarowsky, D. A corpus-based synthesizer. In *Proceedings of the International Conference on Spoken Language Processing* (October, Banff), 1992.
**23.** Webber, B. L. Tense as discourse anaphor. *Computational Linguistics 14*, 2 (June 1988), 61–73.
**24.** Wiebe, J. M. Tracking point of view in narrative. *Computational Linguistics 20*, 2 (June 1994), 233–287.
**25.** Zadrozny, W. and Jensen, K. Semantics of paragraphs. *Computational Linguistics 17*, 2 (June 1991), 171–209.

**About the Authors:**
**JANYCE M. WIEBE** is an assistant professor in the Department of Computer Science and a principal investigator in the Computing Research Laboratory at New Mexico State University. Her principal areas of research include pragmatics and discourse. **Author's Present Address:** Department of Computer Science, New Mexico State University, Box 30001/Dept. CS, Las Cruces, NM 88003; email: wiebe@cs.nmsu.edu

**GRAEME HIRST** is a professor in the artificial intelligence group at the University of Toronto and the author of two books in computational linguistics. **Author's Present Address:** Department of Computer Science, University of Toronto, Toronto, Ontario, M5S 1A4 Canada; email: gh@cs.toronto.edu

**DIANE HORTON** is a lecturer at the University of Toronto. Her principal areas of research include the inference of plans, beliefs, and other mental attitudes from natural language discourse. **Author's Present Address:** Department of Computer Science, University of Toronto, Toronto, Ontario, M5S 1A4 Canada; email: dianeh@cs.toronto.edu