Language use in context

Janyce Wiebe, Graeme Hirst, and Diane Horton

THIS IS NOT THE FINAL VERSION

Authors' addresses:
Wiebe: Department of Computer Science and the Computing Research Laboratory, New Mexico State University, Box 30001/Dept. CS, Las Cruces, NM 88003; e-mail wiebe@cs.nmsu.edu; phone (505) 646-6228; fax (505) 646-6218.
Hirst: Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 1A4; e-mail gh@cs.toronto.edu; phone (416) 978-8747; fax (416) 978-1455.
Horton: Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 1A4; e-mail dianeh@cs.toronto.edu; phone (416) 978-5899; fax (416) 978-1931.

Despite the new emphasis on application-driven research in computational linguistics, much fundamental research remains to be done. In this paper, we describe recent computational work investigating language use in context. This means, first, going beyond sentence boundaries and processing *discourse*—treating texts or dialogues as wholes composed of interrelated parts, rather than merely as sequences of isolated sentences. Any piece of discourse establishes a linguistic context against which subsequent utterances must be understood. Second, beyond the linguistic context is the *participatory* context. A speaker or writer directs an utterance or text toward a hearer or reader, and does so with a particular intent or 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 this intent as part of understanding.

We will discuss 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 that is 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 that we shall describe below occur in both kinds of discourse. We will use the terms *speaker* and *writer* almost interchangeably, and similarly *hearer* and *reader*.

The underlying goal of the research described in this special issue is to move beyond "toy" systems and come to grips with "real language". While the research described in the other papers in this issue 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 a paper of this length, we cannot hope to describe all of the recent, important work addressing language use in context. For example, we will not discuss work on pronoun resolution, ellipsis, metaphor, or many aspects of belief ascription.

STRUCTURE BEYOND THE SENTENCE BOUNDARY

Discourse segmentation

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

(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 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 (1.3–5) give specific details that expand on what is said in sentence (1.2). Moreover, text has relationships at many levels: for example, not only are sentences (1.3–5) related to sentence (1.2), but they have relationships among themselves as well. Fortunately, there are sometimes *cue phrases* or textual markings that help the reader figure out the relationships; and in spoken dialogue, intonation can help too. The word *however* in sentence (1.2), for instance, signals some sort of contrast with what was said in (1.1).

Readers have to recover the structure of a discourse not only so that 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. Consider this conversation:

(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) clearly refers to Sheila, even though it was Linda who was just referred to by *she* in the previous sentence. The structure of the dialogue can help the reader to identify the correct referent. Sentences (2.2–5) interrupt what was under discussion in (2.1), but sentence (2.6) returns to the topic of (2.1). (Notice that the cue phrase *anyway*, possibly accompanied by a small pause, or change in pitch patterns, 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 (2.2–5). In fact, the perception of the structure of the discourse and the interpretation of pronouns constrain one another.

In addition to relationships concerning the content of the text, there are relationships concerning the writer's or speaker's intentions, and the two are closely linked. In example (1), the author had a reason for writing sentences (1.3–5), perhaps to clarify for the reader what was written in (1.2). The details given in (1.3–5) serve this purpose of the author.

There is evidence that people do perform this kind of segmentation of discourse during understanding. For example, Passonneau and Litman [19] and Hirschberg and Grosz [7] found

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

statistically significant agreement among subjects who were 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 be used to introduce a new subtopic [8]. But many words that serve as cue phrases also have other uses—for example, *now* can simply mean 'at this time'. To see how these ambiguities could be resolved, Hirschberg and Litman [8] studied cue-phrase usage in speech, and developed a method for disambiguating cue phrases by means of 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.

The research described above has obvious applications in both speech generation and speech understanding. Speech generation systems can use intonation and cue phrases as people do, to help to break the discourse into appropriate segments, thereby assisting the listener in accomplishing other tasks that are crucial to understanding speech, such as determining the referents of noun phrases and recognizing the rhetorical and intentional relationships between segments. And speech recognition systems can use intonational features and cue phrases (and perhaps also textual features, if the speech has been transcribed) to help perform segmentation and infer relationships between segments. In fact, 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 cue phrases in such a way that the intonational features suggest how the cue phrases are being used.

Relationships within discourse segments

We now turn to some recent work that investigates the structure within discourse segments.

Hobbs *et al.* [10] and others have suggested that during understanding, people make defeasible assumptions, that is, assumptions that are consistent with what they believe, but which can be later 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, with the goal of finding interpretations of individual sentences in a paragraph that are all together consistent. Their approach formalizes the intuitive notion that the sentences in a paragraph all are 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 between segments [9] and of *temporal* relations between the events that the discourse refers to. In the absence of particular 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 (3), for example, one assumes that John's closing the door precedes his sitting down on the couch:

(3) 1. John closed the door to the kitchen.
2. He sat down on the couch.

But in (4):

(4) 1. John fell.
2. Max had pushed him.

we infer the stronger relation that Max's pushing John caused John to fall, in which case the second event precedes the first one (notice that the tenses in (4) support this temporal interpretation); the coherence relation in this case is that (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:

(5) 1. John fell.
2. Max pushed him.

Focusing only on background knowledge and ignoring tense, Lascarides *et al.* express defaults, such as that described above for narration, in a nonmonotonic logic. These are overridden if there is information to the contrary; in (5), for example, 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 (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 (6) (from [12]):

(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 at which (6.1) is said; the doctor informing John of something happened after John's going to the hospital, but before the time when the sentences in (6) are said; and John's breaking his ankle took place before the doctor's telling 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 given in [13]. (The notation t_e refers to the time of event e.)

(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 (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}. (There are other relations possible with the past perfect, but this issue will not concern us here.) Further, 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 (7.2–4) thus form a subnarrative embedded within the overall narrative.

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

(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 (7) and (8), yet the temporal relations among events are the same. Notice that (8.3–4) are in the simple past like (8.5), yet (8.3–4) describe events in the embedded narrative, but (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] take up 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 (9.2) below, the time that the past tense refers to ($t_{rideOff}$) depends on the event described in the previous sentence (*getOn*); in particular, $t_{rideOff}$ is after t_{getOn}, whenever t_{getOn} might be.

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

In Kameyama *et al.*'s terminology, a 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}. Now consider (10) (from [13]), where (10.3a) and (10.3b) are two alternative continuations:

(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 Kameyama *et al.*'s analysis, 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 (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 choosing the right one for a given past tense.

Another type of segmentation

In texts, writers often report the mental states (beliefs, knowledge, intentions, hatred, perceptions, and so on) of many different people. The most straightforward way to report someone's mental state is to present it explicitly as such, with a sentence such as (11.2):

(11) 1. Stuart had accomplished his mission.
2. But he knew that by now the enemy was swarming in 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]

But mental states may also be presented *implicitly*, as in (11.3–5). Example (11) is a passage from a non-fiction book about the American Civil War; in fact, Stuart did not turn back, but continued on. In (11.3–5), the writer is presenting Stuart's motivations for doing so, even though the writer does not explicitly indicate that he is presenting them. Thus, we have the problem of segmenting text according to whose beliefs, intentions, and so forth are being presented, a problem made difficult 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 by 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 be an indication that a block of sentences presenting that agent's mental states will follow (as in (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. Thus, expressions of uncertainty or judgement are similar to cue phrases, discussed earlier in this paper, because they can help the reader perform point-of-view segmentation. An example is the phrase *It was almost as if* in (12.1):

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

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

Even though (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 rather describes the eventual historical outcome. The absence or presence of a phrase such as the hedge in (12.1) is one textual feature that can help a text understanding system recognize implicitly presented mental states.

THE SPEAKER AND THE HEARER

We now move to recent research that acknowledges the role played in discourse by the individual knowledge, goals, experiences, etc. of the speaker and hearer (or writer and reader).

Explanation

In advisory dialogues—dialogues in which an expert advises someone on how to assemble some device, for example, or improve his or her C^{++} programs—advisees may lack the knowledge or experience necessary to fully understand the expert's explanations. In such cases, they often ask follow-up questions. A computer system 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] have developed a text planner for advisory dialogues that has these capabilities.

Moore and Paris integrate two main approaches to discourse in their work. The first, seen earlier in this paper, focuses on the rhetorical or coherence relationships among the segments of the discourse; the other, *intentional*, approach focuses 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 will 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 rhetorical knowledge in the system, which is based on Mann and Thompson's Rhetorical Structure Theory [16], consists of strategies for achieving communicative goals by means of establishing 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 that is already known to the user (using the *contrast* relation); another is to tell the user some of the attributes of the object (using the *elaboration-attribute* relation). Thus, there might be more than one strategy that can achieve a particular goal; as well, a single strategy may be used in service of more than one type of goal.

[3] *ibid*, page 205.

Moore and Paris's system maintains a record of why it said what it said; thus, 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 using a different strategy. Consider (13)(from [18]):

(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?
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 a goal to persuade the hearer to replace SETQ with SETF (in order to enhance the readability and maintainability of his or her Lisp program). In utterances (13.4–5), it therefore motivates 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, in (13.6) the system explains the concept, by 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 the goal of knowing what a generalized-variable is. 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 therefore selects an alternative strategy and tries again: in (13.8), it gives examples of generalized-variables.

Collaboration in discourse

Although the work on explanation just described takes the hearer into account in some ways, discourse is fundamentally collaborative in more ways than these. "The participants in a discourse work together to satisfy various of their individual and joint needs" [5, page 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—*i.e.,* utterances—and actions in the domain of discourse), while the system uses these actions as a basis for attributing intentions or plans to the agent, and is otherwise passive. In many cases, the system had no plans of its own, nor did it consider that 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 pre-formulated. All of these assumptions are incompatible with the collaborative nature of discourse. In truth, all agents involved in a discourse may have plans, they may all infer each other's plans, they often share joint plans (*e.g.,* that one will explain something to another), 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 attempted to embrace these important facets of real discourse.

If one wishes to model situations in which agents share joint plans, one must ask what it means for two agents to have a joint plan to do something. Grosz and Sidner [5] are concerned with precisely this question. They extend Pollack's earlier definition of having a plan [20], which was designed only for single-agent plans, to the case of joint plans. We paraphrase their definition as follows:

DEFINITION: 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;

(2) and 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. Of course, there are a myriad of unresolved philosophical issues that such a definition raises, but Grosz and Sidner's definition is a good starting point for discussion.

The difficult question still remains: how do the agents acquire the intentions and very strong mutual beliefs required to have a shared plan? This is the problem that Lochbaum, Grosz, and Sidner [15] tackle. First, however, they modify the definition in two important ways. Clause (1a) is generalized to say that the agents have a "recipe" for doing A. (A recipe for A, although defined formally, can be thought of 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 allows for any level of detail, that agents may have a shared plan with any level of detail. Clauses (1d) and (2b) are modified to require not that the subaction accomplishes A, but merely that it contributes *somehow* to A (the notion of 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, which 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 where two agents have a shared plan to do A, if one of them says something about some action of type Γ, then the hearer may conclude that the speaker believes that Γ

contributes somehow to A; further, if the hearer can recognize some way in which Γ actually does contribute to A, he may conclude that they mutually believe it contributes.

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 help formulate plans that will meet their goals. The system must be able to discuss goals, and to form plans incrementally as the system and user interact. Building such a system makes it necessary to create and integrate 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 the concepts of planning and plan execution. It also takes an intentional approach, with utterances viewed as linguistic actions; hence, they 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.

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

(14) **User:** 1. We have to make OJ. There are oranges at I and an OJ factory at B. Engine E3 is scheduled to arrive at I at 3PM. 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 should be shipped from I to B, using engine E3. Yet the system infers this, and can therefore determine which oranges the user is referring to when she says *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 that will arrive at 3 p.m. 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; that is, people are seen as faultless language processors whose skills AI research strives mightily to emulate. Even the work discussed above on inferring joint plans, though it brings the speaker and hearer into full consideration, does not 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. We now turn to research that takes a similar perspective on human–*computer* conversation.

Misunderstandings in conversation might occur, for example, because the hearer takes an unintended sense of an ambiguous expression, because the hearer does not have the necessary background knowledge to interpret the utterance or draw the right expected inferences from it, or simply because of an error in typing or speech recognition. If the result is either no interpretation at all or several possible interpretations from which a choice cannot be made, the hearer can then ask for clarification, can remain silent (hoping that subsequent utterances will 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 instance, or an imprecise use of tense.

But if a conversant finds a single reasonable, albeit erroneous, interpretation, the misunderstanding will manifest itself only later, if at all, when the conversants find themselves talking at cross purposes. There are thus two parts to the problem: first, noticing that there has been an earlier misunderstanding—either by oneself or by the other conversant—and secondly, generating an utterance that will repair the misunderstanding. For the first part, Eller and Carberry suggest that if the heuristics described above 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 then 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 (*e.g.,* given Q and $P \Rightarrow Q$, guess that P is also true) and the generation of an utterance as default reasoning (see [10] and above). In this model, conversants abductively form defeasible expectations as to what kind of utterance is likely to occur next in the conversation, and use these expectations to monitor for differences in understanding. When speakers make an utterance, 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 lead the speaker to produce it. If McRoy and Hirst's model finds more than one possibility, it makes a choice 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 at the same time turned a friendly letter into a threatening one, or vice versa.

Indeed, the selection of what is to be said at all depends upon complex inter-personal 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 one doesn't wish to offend. Hovy's natural-language generation system PAULINE [11] was the first to account, in an integrated manner, for inter-personal 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 attempted to find more-general principles for the selection of content, words, and syntactic structures.

One particular problem is the construction of *referring expressions*. A referring expression is a word or phrase that a speaker uses to denote some particular object or entity. The problem is that 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:

(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 of these is best depends on the previous utterances—the last two options require a previous referring expression to act 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 that the speaker could possibly be referring to, detailed descriptions such as (15.4–6) 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 attribute–value pairs that will serve to discriminate the referent from other entities. These methods also take account of 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; that is, they can collaborate on the task of reaching a common understanding of the reference. Heeman and Hirst [6] have modeled 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; and 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, has sought mostly 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; *e.g.,* either *the pitted olives* or *the olives that have been pitted.* But, as Hovy's work showed, form is just as important. DiMarco and Hirst [3] sought to correlate a writer's use of various syntactic constructions with his or her higher-level stylistic goals, with a view to ensuring that an automatic translation retains these goals even if that requires a different syntactic structure in the target language. For example:

(16) Ils se livrent alors, <u>sous des dehors irrésistibles de drôlerie</u>, à une lutte sournoise et passionnée.

(17) And, <u>from behind a cover of irresistibly funny wit</u>, they open fire in an artful and passionate battle.[4]

The underlined phrase that interrupts the main clause in (16) was moved by the translator to a position before the main clause in (17). Though (16) is quite natural in French, the movement of the phrase was necessary to prevent what would otherwise be a somewhat unnatural sentence in the corresponding 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.

[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: Problémes de traduction*, Éditions Ophrys, 1981, p. 125), who lists and explains some of the differences in the naturalness of various structures in English and French.

CONCLUSION

The common core of the research that we have discussed in this paper is a concern for the full complexity of language as people use it: in particular, the complexity that arises in creating and comprehending units larger than a single sentence, in deciding upon the best word or expression that fits the present context and intent, and in 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.

Acknowledgements

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

References

[1] Allen, James F., Schubert, Lenhart K., Ferguson, George, Heeman, Peter, Hwang, Chung Hee, Kato, Tsuneaki, Light, Marc, Martin, Nathaniel G., Miller, Bradford W., Poesio, Massimo, and Traum, David. The TRAINS Project: A case study in building a conversational planning agent. To appear in *Journal of Experimental and Theoretical Artificial Intelligence*, 1995.

[2] Dale, Robert. *Generating Referring Expressions.* The MIT Press, Cambridge, Mass., 1992.

[3] DiMarco, Chrysanne and Hirst, Graeme. A computational theory of goal-directed style in syntax. *Computational Linguistics 19*, 3 (September 1993), 451–499.

[4] Eller, Rhonda and Carberry, Sandra. A meta-rule approach to flexible plan recognition in dialogue. *User Modeling and User-Adapted Interaction 2*, 1–2 (1992), 27–53.

[5] Grosz, Barbara J. and Sidner, Candace L. Plans for discourse. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, Eds. *Intentions in Communication.* The MIT Press, Cambridge, Mass., 1990, pp. 417–444.

[6] Heeman, Peter and Hirst, Graeme. Collaborating on referring expressions. *Computational Linguistics 21*, 3 (September 1995).

[7] Hirschberg, Julia and Grosz, Barbara. Intonational features of local and global discourse structure. In *Proceedings of the Speech and Natural Language Workshop* (Feb. 23–26, Harriman, N.Y.). DARPA, Software and Intelligent Systems Technology Office, 1992, pp. 441–446.

[8] Hirschberg, Julia and Litman, Diane. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics 19*, 3 (September 1993), 501–530.

[9] Hobbs, Jerry R. Coherence and coreference. *Cognitive Science 3*, 1 (January–March 1979), 67–90.

[10] Hobbs, Jerry R., Stickel, Mark, Martin, Paul, and Edwards, Douglas. 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, pp. 95–103.

[11] Hovy, Eduard H. *Generating Natural Language under Pragmatic Constraints.* Lawrence Erlbaum Associates, Hillsdale, N.J., 1988.

[12] Hwang, Chung Hee and Schubert, Lenhart 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, pp. 232–240.

[13] Kameyama, Megumi, Passonneau, Rebecca, and Poesio, Massimo. 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, pp. 70–77.

[14] Lascarides, Alex, Asher, Nicholas, and Oberlander, Jon. 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, Karen E., Grosz, Barbara J., and Sidner, Candace L. Models of plans to support communication: an initial report. In *Proceedings of the 9th National Conference on Artificial Intelligence* (July 29–Aug. 3, Boston, Mass.). American Association for Artificial Intelligence, Calif., 1990, vol. 1, pp. 485–490.

[16] Mann, William and Thompson, Sandra. Rhetorical structure theory: toward a functional theory of text organization. *Text 8*, 3 (1988), 243–281.

[17] McRoy, Susan and Hirst, Graeme. The repair of speech act misunderstandings by abductive inference. *Computational Linguistics, 21*, 3 (September 1995). [Probably]

[18] Moore, Johanna and Paris, Cécile. Planning text for advisory dialogues: capturing intentional and rhetorical information. *Computational Linguistics 19*, 4 (December 1993), 651–694.

[19] Passonneau, Rebecca and Litman, Diane. 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, pp. 148–155.

[20] Pollack, Martha 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, pp. 207–214.

[21] Reiter, Ehud. 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, pp. 257–285.

[22] Sproat, Richard, Hirschberg, Julia, and Yarowsky, David. A corpus-based synthesizer. In *Proceedings of the International Conference on Spoken Language Processing* (October, Banff), 1992.

[23] Webber, Bonnie Lynn. Tense as discourse anaphor. *Computational Linguistics 14*, 2 (June 1988), 61–73.

[24] Wiebe, Janyce M. Tracking point of view in narrative. *Computational Linguistics 20*, 2 (June 1994), 233–287.

[25] Zadrozny, Wlodek and Jensen, Karen. Semantics of paragraphs. *Computational Linguistics 17*, 2 (June 1991), 171–209.

CR Categories and Subject Descriptors: I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems—*Natural Language Interfaces*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing.

Additional Key Words and Phrases: Discourse, Collaboration, Style

About the authors:

JANYCE M. WIEBE works in the areas of pragmatics and discourse. She is an assistant professor in the Department of Computer Science and a principle investigator in the Computing Research Laboratory at New Mexico State University. She received her Ph.D. in computer science from the State University of New York at Buffalo.

Author's present address: Department of Computer Science, New Mexico State University, Box 30001/Dept. CS, Las Cruces, NM 88003; e-mail wiebe@cs.nmsu.edu

GRAEME HIRST, a professor in the artificial intelligence group at the University of Toronto, is the author of two books in computational linguistics. He is book-review editor of the journal *Computational Linguistics*.

Author's present address: Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 1A4; e-mail gh@cs.toronto.edu

DIANE HORTON's research interests are in the inference of plans, beliefs, and other mental attitudes from natural language discourse. She is a lecturer at the University of Toronto.

Author's present address: Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 1A4; e-mail dianeh@cs.toronto.edu