1

Introduction

1.1 Toward a theory of language generation and communication

A primary goal of natural-language generation research in artificial intelligence is to design a system that is capable of producing utterances with the same fluency as that of a human speaker. One could imagine a "Turing Test" of sorts in which a person was presented with a dialogue between a human and a computer and, on the basis of the naturalness of its use of the English language, asked to identify which participant was the computer. Unfortunately, no natural-language generation system yet developed can pass the test for an extended dialogue.

A language-generation system capable of passing this test would obviously have a great deal of syntactic competence. It would be capable of using correctly and appropriately such syntactic devices as conjunction and ellipsis; it would be competent at fitting its utterances into a discourse, using pronominal references where appropriate, choosing syntactic structures consistent with the changing focus, and giving an overall feeling of coherence to the discourse. The system would have a large knowledge base of basic concepts and commonsense knowledge so that it could converse about any situation that arose naturally in its domain.

However, even if a language-generation system met all the above criteria, it might still not be able to pass our "Turing Test" because to know only about the syntactic and semantic rules of the language is not enough. For this reason, calling such a system a "language-generation system" is in truth a misnomer, because it would be more accurately called a *communication* system. One must constantly bear in mind that language behavior is part of a coherent plan and is directed

toward satisfying the speaker's goals. Furthermore, sentences are not straightforward actions that satisfy only a single goal. The utterances people produce are crafted with great sophistication to satisfy multiple goals at different communicative levels. For example, in a single utterance a speaker may inform a hearer of two or more propositions, make a request, shift the focus of the discourse, and flatter the hearer. On the surface, this does not argue that anything more than the above criteria is needed to produce natural-sounding utterances — all that is necessary is to allow for greater complexity. Things are not that straightforward, however, because recognizing how an utterance satisfies multiple goals often requires that the hearer know about the speaker's plan, and reason about how the utterance fits into it. A speaker attempting to plan such an utterance must reason about what the hearer knows and how the hearer can interpret the speaker's intentions. When a speaker produces an utterance, he is reasoning not only about his language, but about the entire communicative process.

Consider the situation in Figure 1.1, which is typical of two agents cooperating on a task in which one has to make a request of the other. The speaker points to one of the tools on the table and says, "Use the wheelpuller to remove the flywheel." The hearer, who is observing the speaker while he makes the request, and knows that the speaker is drawing his attention to a particular tool, thinks to himself, "Ah, so that's a wheelpuller. I was wondering how I was going to get that flywheel off."

In this situation, the speaker's utterance affects the hearer far beyond a simple analysis of the propositional content of the utterance. Most obviously, the speaker is requesting the hearer to carry out a particular action, since the use of the imperative strongly suggests that a request is intended. However, the speaker includes using the wheelpuller as part of his request. If he knew that the hearer did not know that he was supposed to use the wheelpuller to remove the flywheel, then his utterance also serves to inform the hearer of what tool to use for the task. In addition, the fact that the speaker *points* to a particular object communicates his intention to refer to it with the noun phrase "the wheelpuller." (In fact, pointing may be the only way to refer successfully to an object when the only mutually believed description of it is that it is some sort of thing.) Since the intention to

refer has been communicated, the noun phrase also communicates the fact that the intended referent is a wheelpuller. Thus the speaker has performed a "labeling" action that will enable him to refer more easily to the object in subsequent utterances. The speaker could have just said, "Use *that thing* to remove the flywheel," if he had no goal of informing the hearer that the tool was a wheelpuller.

Utterances that are intended by the speaker to satisfy multiple goals are very common. Here are a few examples of common situations in which the speaker plans an utterance to satisfy multiple goals:

- A rock climber says to a friend, "Joe and I are going to climb the Snake Dike Route on Half Dome next weekend." The speaker does not use the prepositional phrase "on Half Dome" to pick out a particular Snake Dike Route from several that he and the hearer mutually know about. Most likely, he is informing the hearer that he is going to climb a particular route, and the additional descriptor "on Half Dome" is intended to inform the hearer of the location of a route that he never heard of before, and perhaps impress him as well.

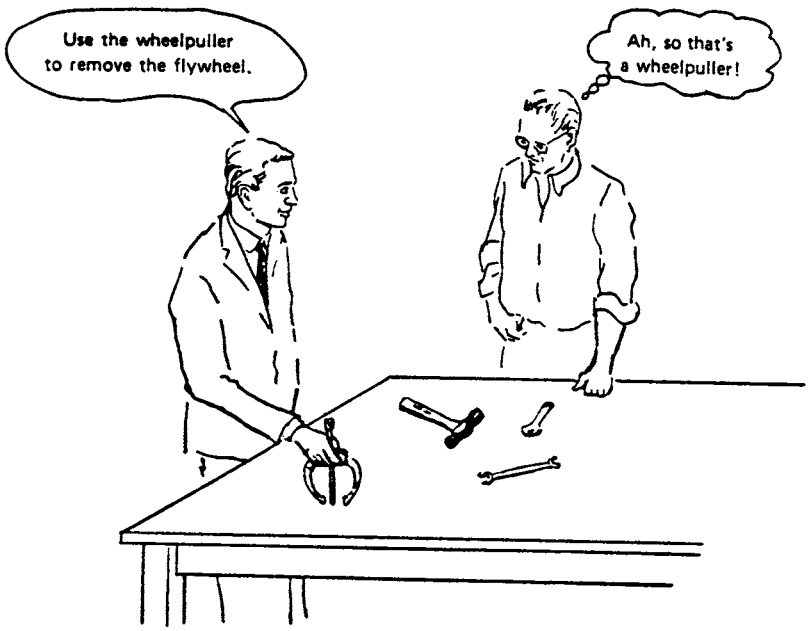

Figure 1.1: Satisfying multiple goals with a request

- The rabidly anti-American Ayatollah Khomeini says "The Great Satan shall not defeat us." It is obvious from the context of the utterance that the description "Great Satan" is intended to refer to the United States, even though the description is not objectively true of its intended referent. In this case, the speaker exploits a noun phrase to simultaneously refer and communicate his emotional attitude toward the referent.
- Multiple-goal satisfaction even plays a role in such conventional utterances as "Could you tell me what time it is?" In this case the speaker chooses the indirect speech act to satisfy a goal of demonstrating politeness toward the hearer, although the more direct but less polite "What time is it?" would have conveyed the request equally well.

1.2 The need for a general planning mechanism

Figure 1.1 illustrates how understanding a speaker's physical actions can be important for understanding an utterance. The speaker's action of grasping the wheelpuller is interpreted by the hearer to have communicative intent, like pointing. The speaker assumes that the hearer knows the connection between such communicative gestures and the linguistic act of uttering an noun phrase. Since linguistic acts and physical acts can be interpreted together in reasoning about a speaker's intentions, a language-generation system that treats physical and linguistic actions as uniformly as possible will enable the production of utterances that, like the one in Figure 1.1, satisfy multiple goals.

In Figure 1.2 the agents are faced with a problem similar to that of Figure 1.1, but the agent making the request happens to be holding a large box, which prevents him from grasping the wheelpuller as he did in Figure 1.1. If he says the same thing as he did in Figure 1.1 to achieve realization of his request, he will not succeed; since the hearer does not know what a wheelpuller is, the speaker will not have established his intention to refer, as he did in the previous example.

One option open to the speaker is to devise some description of the object that does not require pointing, and perhaps to inform the hearer later in a different utterance that the object is a wheelpuller. However, since the only mutually believed descriptors of the intended referent are those resulting from perception, the resulting description will probably be awkward (e.g., "the thing with two arms and a large screw in the middle"). The speaker could also attempt to describe the object first and then refer to it; however, this tactic too can be awkward. Of

course, if an agent does not have physical actions at his disposal, these techniques are his only alternatives.

Another alternative that could be planned when the speaker has both physical and linguistic actions at his disposal is for him to set down the box, which would free his hands, and then proceed as in Figure 1.1. As this example illustrates, relatively low-level linguistic planning, such as deciding what description to use to refer to something, can lead to the planning of physical actions. Such interaction provides support for the argument in favor of closely correlating the planning of physical and linguistic actions.

Many communicative situations arise in which the speaker and the hearer are not in proximity to each other, and therefore physical actions are not relevant to the planning process. A common example is the writing of text, whereby the reader who will interpret the text is not only normally unknown to the writer, but will almost certainly be at a different location. In such a case, a model of

Figure 1.2: The need for integrating physical and linguistic actions

language generation based on planning is still applicable because, even though there is really just one agent, the writer must still plan his text as though there were another agent directly involved. Instead of reasoning about a particular person's beliefs, a "typical reader" is assumed to be the other agent, and all the problems of intention communication and reasoning about the beliefs of the typical reader are likely to apply. Considerations such as his knowledge of the topic and his technical sophistication come into play.

A hypothesis of this research is that an agent's behavior is controlled by a general goal satisfaction process. Agents are assumed to have goals that are satisfied by constructing plans from available actions. Given that an agent's overall behavior is controlled by such a planning process, it is advantageous for his linguistic behavior to be controlled by such a process as well. The reasons for this conclusion are that (1) agents have to plan *both* physical and linguistic actions to achieve their goals, (2) linguistic and physical actions interact with each other, and (3) actions such as informing and requesting interact and can be realized simultaneously in the same utterance. Since a language-generation system must reason about these interactions to produce natural-sounding utterances, a uniform process that plans both physical and linguistic actions is needed.

1.3 A theory of language generation based on planning

Generating natural language by means of a general planning mechanism is a reasonable approach to the problem for a variety of reasons discussed in the previous section. However, this approach requires adopting a different view of language and communication from the one that has usually been characterized in past language-generation research. Many previous theories and their related systems have regarded language processing in a manner analogous to that depicted in Figure 1.3, which illustrates a view that has been labeled the *conduit metaphor* by Reddy (1979). The conduit metaphor refers to the treatment of language as a pipeline or conduit that relays information from the speaker to the hearer. The speaker, starting with some idea of what he wants to say, "packages" it in natural language and sends the package through the conduit to the hearer, who then unwraps the package and removes the contents. This metaphor is quite pervasive in our

commonsense intuitions about language and is reflected in many of our prevalent sayings, for example "He got his ideas across very well," or "He couldn't put his thoughts into words."

The disadvantage of this general view is that it forces one to postulate a very strong separation between two stages of the utterance-planning process: deciding *what* to say and deciding *how* to say it. Most language-generation research to date has adopted the what-how separation as a design principle to be saliently reflected in the structure of systems to be designed. Usually there is an "expert system" which draws upon a very large corpus of knowledge about its task domain when interacting with a user. When the expert system needs to communicate, it encodes a fragment of its knowledge in some internal language and then passes it on to a "generation module" that translates the message from the internal language into English.

In contrast to the language-as-conduit view outlined above, the approach advocated by this theory (represented in Figure 1.4) treats language not as something to be transferred through a conduit, but rather as a set of actions available to agents that affects the mental states of other agents. This approach regards decisions about "what to say" and "how to say it" as two facets of the same overall

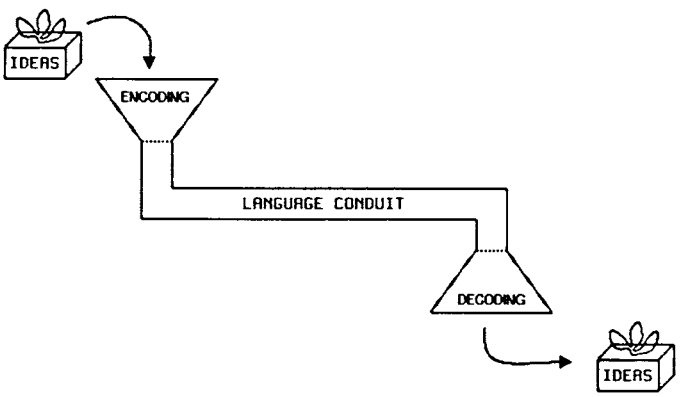

Figure 1.3: The conduit metaphor

process, and recognizes the interactions between them. The planning of an action appropriate for a given situation necessitates consideration of several crucial elements: different kinds of goals that are satisfied by utterances; the knowledge of the hearer; general knowledge about the world; the constraints imposed by the syntax of the language. The utterance planner can integrate these diverse knowledge sources to arrive at a plan involving abstract specifications of speech acts and then, finally, produce English sentences. Instead of regarding the hearer as the mere consumer of a message, the utterance planner treats him as an active participant in the communication process.

The planning system developed as a part of this research is called KAMP, which is an acronym for **K**nowledge **A**nd **M**odalities **P**lanner. KAMP is a hierarchical-planning system that uses a nonlinear representation of plans, called a *procedural network* by Sacerdoti (1977). A hierarchical design for a utterance-planning system was selected because it provides for separating the planning of domain-level goals and actions and the planning of low-level linguistic actions, as well as for intermediate levels of abstraction that facilitate the integration of multiple goals

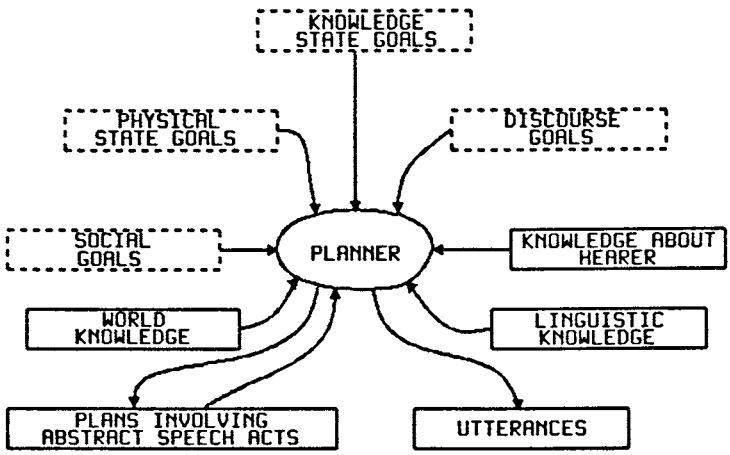

Figure 1.4: Overview of an utterance planner

into utterances. The hierarchy of linguistic actions used by KAMP is represented in Figure 1.5. The planner can focus its attention on domain-level and high-level linguistic actions while ignoring details about the choice of syntactic structures and descriptions for referring expressions. However, the uniformity of treatment of linguistic actions allows higher-level goals and actions to be influenced by the expansion of low-level linguistic actions. The mechanism KAMP uses to accomplish this is described in Chapter 6.

The highest-level linguistic actions are called *illocutionary acts*, which are such speech acts as informing or requesting represented at a very high level of abstraction, without any consideration given to an action's ultimate linguistic realization. The next level consists of *surface speech-acts*, which are abstract representations of sentences with particular syntactic structures. At this level, specific linguistic knowledge becomes important. One surface speech act can realize one or more illocutionary acts. The next level consists of *concept activation actions*, which

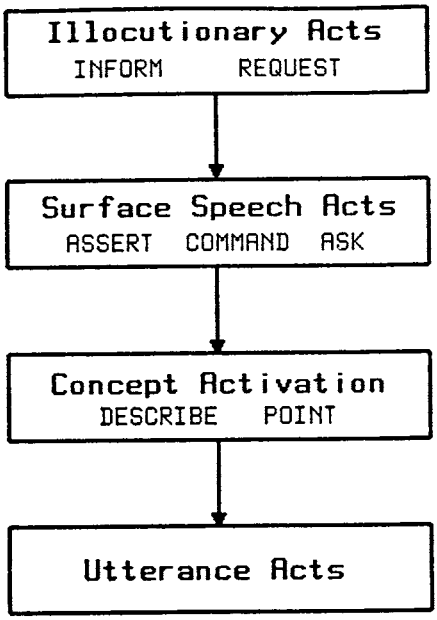

Figure 1.5: A hierarchy of actions related to language

entail the planning of descriptions that are mutually believed by both the speaker and the hearer to refer to objects in the world. Finally, concept activation actions are expanded as *utterance acts*, at which point specific words and syntactic structures are chosen to realize the descriptors chosen for the concept activation actions. These syntactic structures have to be compatible with the sentential syntactic structure selected when the surface speech act is planned. Concept activation actions can also be expanded partially as physical actions that establish the speaker's intention to refer, such as pointing. The detailed axiomatization and treatment by KAMP of each of these action types is described in detail in Chapters 5, 6 and 7.

1.4 The goals of this research

For any research to merit the label "scientific," there must be a clear idea of just what constitutes the problem that is being addressed, and what is to be regarded as a potential solution. The problem that confronts us here is essentially one of explanation of behavior. Agents are observed to exhibit consistent behavior; consequently some explanation is required to account for why they do what they do. This research accepts as given the mentalistic hypothesis that agents have mental states that causally determine their actions. What this research must explicate, at least in part, is how particular mental states account for particular actions. Since this work focuses on language, it is particularly concerned with the question of how holding a particular set of beliefs and intentions results in an agent's making a particular utterance.

Mental states are difficult objects to deal with because they are not directly observable. Therefore, this research proceeds by constructing a formal, computational theory (along with a system that embodies this theory) in which it is possible to represent mental states and provide an explicit mechanism whereby these states determine actions. To the extent that such a model can indeed account for observed behavior, one has succeeded in providing a computational theory.

To provide a complete computational theory of speech acts and the use of natural language is a very tall order, and this research falls far short of this ultimate goal. The principal significant contributions until now have been the establishment