SPEAK OF THE DEVIL: REPRESENTING DEICTIC AND SPEECH ACT KNOWLEDGE IN AN INTEGRATED LEXICAL MEMORY[1]

Michael Gasser
Michael G. Dyer

Artificial Intelligence Laboratory, UCLA

INTRODUCTION

Consider the following dialogue:

```
1)  X:  I heard Sally got tenure.  [Sally approaches.]
    Y:  Well, speak of the devil.
```

The expression *speak of the devil* says, in effect: "What a coincidence! The person that we were just talking about has arrived on the scene." How would this notion and the lexical entry for this expression need to be represented in memory in order for Y to be able to generate it? The problem is that the concept involves a person who is definable only in terms of the ongoing discourse. A speaker should fail to select the expression *speak of the devil*, for example, when a coincidence referred to involves a person not under discussion.

We are developing a program called CHIE which simulates the English and Japanese speech production of a Japanese learner of English in conversational and narrative contexts (Gasser, 1985). For conversation issues like that associated with *speak of the devil* assume importance. The problem is one of lexical representation. We want lexical entries that can make declarative reference to elements of the utterance and the discourse context in which the lexical items occur. While our main concern is generation, we have been guided by the need to have lexical entries usable in parsing as well. In addition, since we are modelling a human language learner, a major consideration in the design of the lexicon has been learnability.

DEIXIS AND SPEECH ACTS IN CONVERSATION

Deixis

In all language use, but in conversational contexts in particular, certain expressions make reference to entities that are part of the discourse itself, i.e., the speaker, the listener, the time and place of the conversation, and the topic of discussion. Such expressions are called **deictics** (Anderson & Keenan, 1985). *Speak of the devil* is an example of a **discourse deictic**, an expression that points to something that is being discussed. Other categories include person, time, place, and social deictics.

Person deictics are expressions that refer to the speaker or listener of the utterance in which they occur. Sentence 2) has two person deictics, *I* and *your Honor*.

```
2)  Your Honor, may I question the witness?
```

Your Honor is an example of a **vocative**, an expression used to get the attention of a potential listener or to emphasize the contact between the speaker and listener. Vocatives are often highly context-dependent, and *your honor*, which occurs in the context of a trial, is a case in point. This context-dependence further complicates the lexical representation. Not only must the person referred to by *your Honor* be specified as the listener of the current utterance but also as the judge of the trial in which the utterance occurs. Context specificity is not confined to vocatives. In some cases particular expressions constitute basic events in

[1]The research reported on here was supported in part by a grant from the ITA Foundation.

scripts (Schank & Abelson, 1977) or other knowledge structures. An example is *I now pronounce you man and wife* from the wedding script.

Time deictics are expressions that are definable in terms of the time of the utterance. Examples include *ago* and *yesterday*. **Place deictics** point to the location of the speaker or listener. Examples are *here* and *come*. *Come*, for instance, refers to a movement in the direction of the speaker or listener.

Social deictics make reference to a social relationship between the speaker and someone else. Consider the following Japanese sentences.

```
3a)  Abe-san  wa     imooto     ni  yubiwa  o       kure-masita.
     Mr.-Abe  TOPIC  my-sister  to  ring    DIROBJ  give-PAST
     'Mr. Abe gave my sister a ring.'

 b)  Imooto     wa     Abe-san  ni  yubiwa  o       age-masita.
     my-sister  TOPIC  Mr.-Abe  to  ring    DIROBJ  give-PAST
     'My sister gave Mr. Abe a ring.'
```

Note that there are two separate verbs corresponding to English *give*. The notion of relative "social distance" is needed to define the distinction. *Kure-* is appropriate only when the speaker is socially "closer" to the recipient than to the donor. *Kure-* is selected in 3a) because the recipient is the speaker's sister, who is closer to the speaker than is Abe, the donor. With the donor and recipient reversed, as in 3b), *kure-* is not possible, and the default *age-* is selected. *Kure-* is not an isolated case; social deixis in various forms runs rampant through the Japanese lexicon.

Speech Acts

Every utterance realizes a **speech act** of some kind; that is, it is an attempt to achieve some goal of the speaker or writer (Searle, 1969). In conversational contexts, however, we deal with a greater variety of speech acts that we find in, say, narrative, and many are achieved through the use of lexical patterns. Speech acts resemble deictics in that their definition requires reference to the current speaker and listener. A **directive**, for example, has the goal on the part of the speaker of getting the listener to perform some action. The speech acts that we are interested in here are **indirect** ones such as 4).

```
4)  I was wondering if you could take out the garbage.
```

What we note about such utterances is that they have both a "literal" and an "indirect" sense. For example, in its "literal" sense 4) simply constitutes an assertion about the speaker's wanting to know something, while in its "indirect" sense it functions as a directive. The lexicon should be organized in such a way that both meanings are available to listeners who are interpreting the sentences (Gibbs, 1980).

THE LEXICAL MEMORY APPROACH

The last ten years have seen a move in theoretical linguistics toward the inclusion of increasing amounts of linguistic knowledge in the lexicon (e.g., Bresnan & Kaplan, 1983; Hudson, 1984). At the same time, because lexical items can be associated directly with concepts and also provide much of the structure of a phrase directly, language processing models have come to be mostly lexically- rather than syntactically-driven (e.g., for parsing, Dyer, 1983; for generation, McKeown, 1985). Finally, there is a new emphasis on idiomaticity in language, with provision for a large number of lexical entries specifying morpho-syntactic **patterns** of varying degrees of specificity. This view stems from work in linguistics (Fillmore, 1979), in psycholinguistics (Pawley & Syder, 1983; Peters, 1983), and within the **phrasal lexicon** framework and related approaches in AI (Becker, 1975; Jacobs, 1985; Wilensky & Arens, 1980; Zernik & Dyer, 1985). The arguments center on the sheer number of lexical patterns which cannot be accounted for in terms of compositional semantics, the productive nature of many of these phrases, and the computational savings that result from the inclusion of frequent phrases in the lexicon.

At the same time, the need to integrate these notions into the rest of the linguistic system has led to the use of hierarchical network or frame-like representation systems for linguistic knowledge (Hudson, 1984; Jacobs, 1985; Langacker, 1986). This approach results in greater modularity and permits the efficient sharing of features through the creation of categories at various levels. In addition, the linguistic knowledge base of a system can be represented using the same general framework as for other knowledge in the system, promoting uniformity.

We assume that linguistic knowledge is acquired by associating relatively fixed patterns with relatively specific concepts or contextual elements (Peters, 1983). When generalizations are made on the basis of instances of similar patterns, higher-level nodes are added to the hierarchy. Each new generalization is either knowledge about how a type of pattern conveys a semantic notion or about how a type of utterance can achieve a type of communicative goal.

Many uses of language are inseparable from the contexts in which they occur. *Your Honor* in 2) above provides one striking example. The implication is not only that lexical items need to be acquired in context (Zernik & Dyer, 1985) but also that lexical items comprising parts of knowledge structures such as scripts (Schank & Abelson, 1977) are learned along with the knowledge structures themselves. This position agrees with research emphasizing the contextualized nature of language acquisition (Hatch, Flashner, & Hunt, 1986).

In CHIE there is a language-independent hierarchy of concepts, including generalized states, events, actions, physical objects, and goals. Knowledge about morphology is organized in a separate hierarchy of words and word classes (Hudson, 1984). An **utterance** associates a concept with a sequence of one or more words, and a **speech act** associates a goal with an utterance. The basic units of linguistic knowledge (above the phonological and morphological level) are utterance and speech act generalizations. We call our scheme the **Lexical Memory** approach because of the importance of lexical information and because of the ways in which linguistic knowledge is integrated into the rest of memory.

INTEGRATING LEXICAL, CONCEPTUAL, AND PRAGMATIC INFORMATION

Generalized utterances (GUs) correspond closely to the pattern-concept pairs of phrasal lexicons. They take the form of frame-like structures and are organized in a hierarchy. In addition to roles (slots) for a **pattern** and a **content** (the concept referred to), each GU has "deictic" roles for **speaker, listener, time,** and **focus set,** the set of elements currently under discussion, plus a **language** role to distinguish patterns in CHIE's two languages.[2] GUs correspond both to lexical entries and to higher-level syntactic patterns. Figure 1 shows how the GU for the Japanese verb *age-* (sentence 3b) joins a node in the conceptual hierarchy to the hierarchy of words. The "G" links connect concepts to their generalizations (types); the "=" links join equivalent concepts. Role names appear in lower-case. The portion of the conceptual hierarchy shown in the figure includes ATRANS, a transfer of something from one person to another (Schank & Abelson, 1977) and a specialization of this concept, ATRANS-FROM-ACTOR, an ATRANS in which the person from whom the object is transferred is viewed as the actor (Jacobs, 1985). U.AGE- represents a class of utterances which refer to an instance of ATRANS-FROM-ACTOR and have a word of type "AGE_" as the head of their patterns. "AGE_" represents verbs with stem *age-*. Two subtypes of "AGE_" are shown in the figure, including "AGEMASITA", the form used in 3b).

Generalized speech acts (GSAs) are prototypical associations of goals with GUs which constitute plans to achieve these goals. They function like the goal-plan associations in Wilensky (1983). Each GSA has roles for an **objective** of the speaker (the goal) and a GU **plan.** The goal in a GSA **intends** the plan; that is, the execution of the plan is a consequence of the planner's desire to achieve the goal (Dyer, 1983). Like GUs, GSAs are arranged in a hierarchy.

---

[2]For simplicity the language role is ignored in the remainder of this paper. See Gasser (1985).

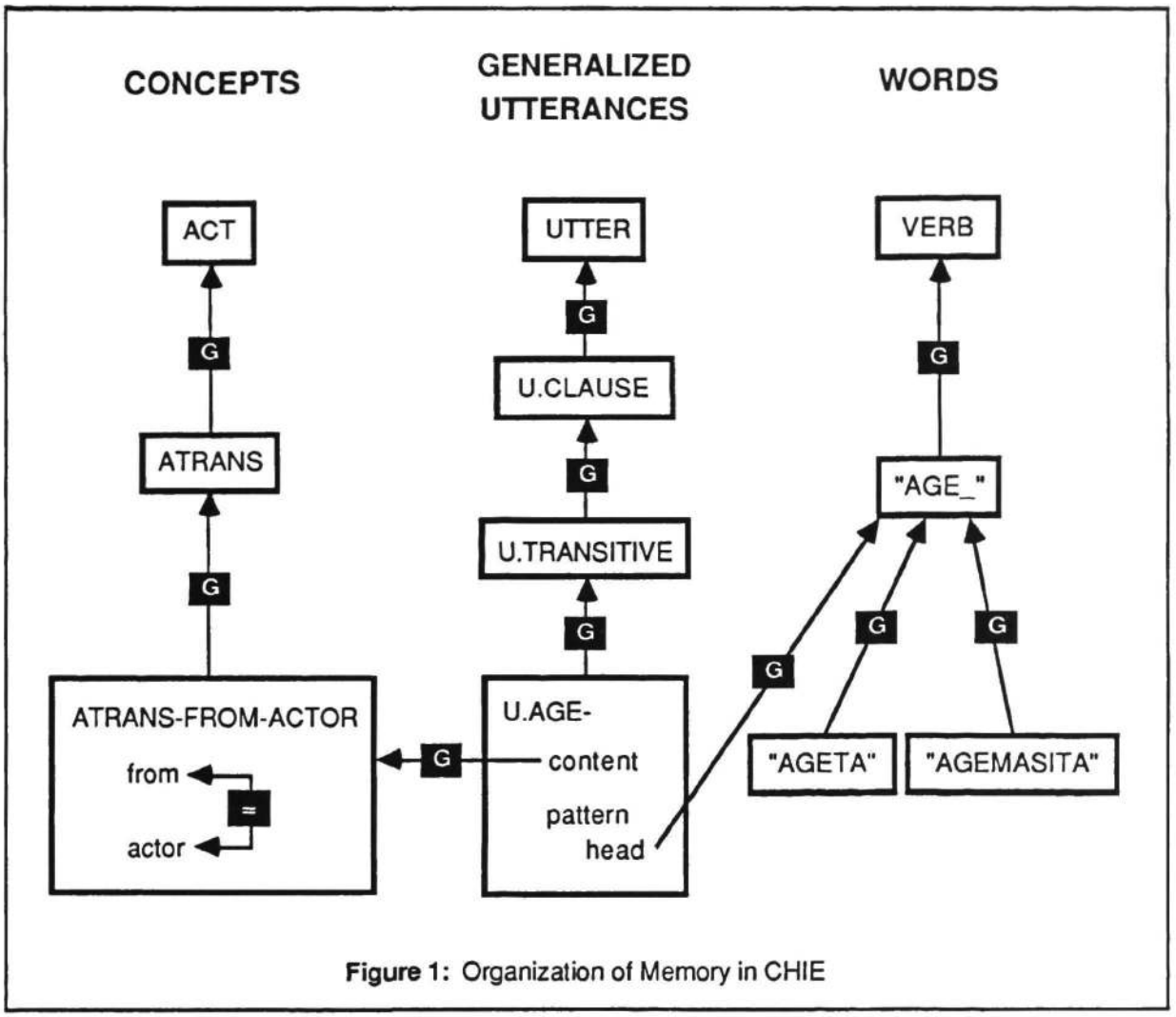

Figure 1: Organization of Memory in CHIE

In addition to the familiar illocutionary acts (Searle, 1969), GSAs include acts such as those realized by vocatives. One use of vocatives is embodied in the GSA SA.ADDRESS, shown in Figure 2. In this and subsequent figures restrictions on roles appear in parentheses following the role names. Thus the plan for SA.ADDRESS must be a kind of U.REFER. What SA.ADDRESS represents is roughly the following: for A to get B to ATTEND (Schank & Abelson, 1977) to an utterance of A's, A produces an utterance which refers to B (a U.REFER).

GUs and GSAs, like other nodes in the network, implicitly inherit all of the role information contained in their generalizations (Fahlman, 1979). For this reason a frame only makes reference to those roles which are restricted in that frame. For example, a general relationship which is true for all GSAs is that the person having the objective is also the speaker of the plan-utterance. This fact is not shown in Figure 2 because it is inheritable from the general GSA, SPEECH-ACT.

SA.ADDRESS is too general to provide specific information about the pattern for its plan-utterance, and it has specializations for particular categories of addressees and particular contexts. One of these is the GSA for addressing a judge during a trial, shown in Figure 3. SA.ADDRESS-TO-JUDGE is one of a set of speech acts which form elements of the trial script. *Your Honor* is appropriate only in the context of a trial and only for addressing the particular judge who is presiding at the trial. Embedding this GSA in the trial script allows us to represent these facts in a straightforward way.

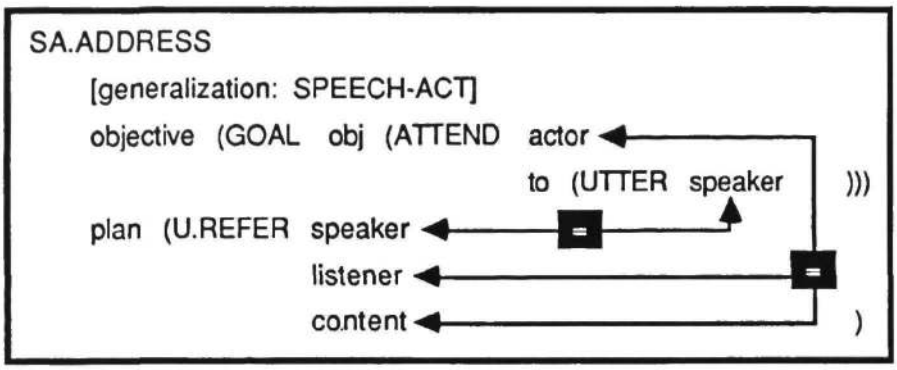

Figure 2: GSA for Address

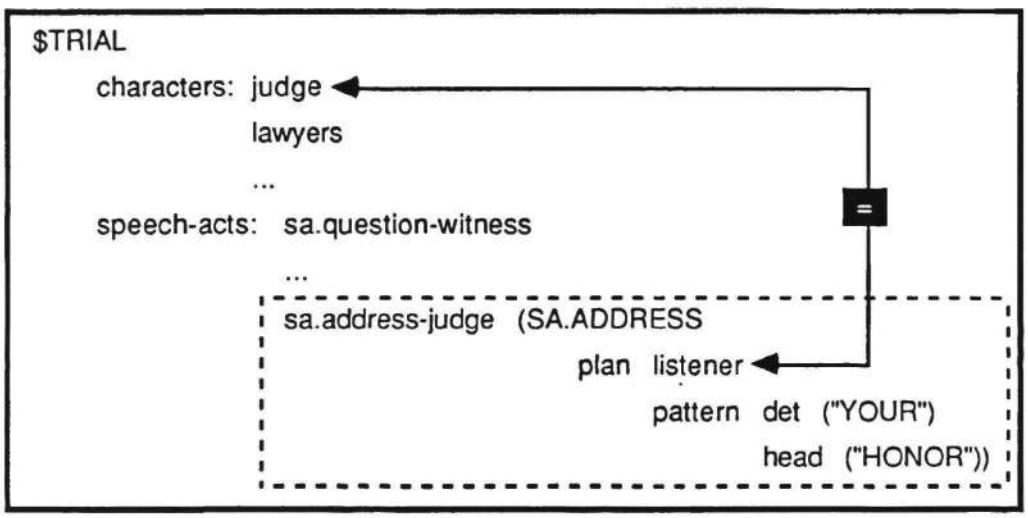

Figure 3: GSA for Addressing Judge

PROCESS OF GENERATION

Generation involves selecting GSAs and GUs in memory which match input goals of the speaker and conceptual content to be referred to. The GSAs and GUs yield patterns which are usually only partially specified. These are combined in a process analogous to **unification** (Kay, 1979) to produce complete patterns.

For the purposes of generation, GSAs are indexed by their objective roles and GUs by their content roles. Both may be indexed further in terms of the contexts in which they occur. Selection of an appropriate GSA or GU often involves classifying the goal or content instance on the basis of successive discrimination. In this sense the nodes in the GSA and GU hierarchies correspond to the nodes in a discrimination network (Goldman, 1975).

Consider the generation of *your Honor* in sentence 2). This occurs in the context of a trial, and we assume that the trial script has already been instantiated along with its various roles. The speaker's goal is that the judge attend to his/her question. This matches the goal in SA.ADDRESS. SA.ADDRESS has

specializations, distinguished on the basis of their contexts, so further classification is possible. One specialization is SA.ADDRESS-TO-JUDGE, which matches the speech act being formulated because it exists within the trial script and because the intended addressee is the current presiding judge. SA.ADDRESS-TO-JUDGE provides the full pattern for the utterance, *your Honor*. In most cases, one or more GUs would also need to be selected to complete the pattern.

DISCOURSE AND SOCIAL DEIXIS

Like other discourse deictics, *speak of the devil* makes reference to an element that is under discussion, i.e., a member of the current focus set (Grosz, 1984; McKeown, 1985). *Speak of the devil* is used to make the listener aware of a coincidence, namely, that a person who is a member of the focus set has just arrived on the scene. Since coincidences are usually worthy of mention, it is not surprising that languages have conventional ways of referring to particular kinds of coincidences. Thus corresponding to English *speak of the devil* is Japanese *uwasa o sureba kage*, literally, 'if (you) make a rumor (about a person), (his) shadow (appears)'. The GU for *speak of the devil* is shown in Figure 4. We represent a coincidence as a fact with three arguments, one for the common element and the others for two facts which are true of the common element. In the case of *speak of the devil* the common element is the person in question, one fact is the membership of this person in the current focus-set, and the other fact is the physical transfer of the person to the location of the speaker.

Consider now sentences 3a) and 3b). The verbs *kure-* and *age-*, like English *give*, both refer to the concept ATRANS-FROM-ACTOR, which is shown in Figure 1. What distinguishes the GUs for these two verbs is the provision regarding relative social distance for *kure-*. For such GUs we need to add the possibility of a **constraint**. A constraint is a role filled by a fact which associates other roles in the frame. Figure 5 shows the GU for this word. The constraint for U.KURE- is that the recipient of the ATRANS is socially nearer to the speaker than is the donor of the ATRANS. In generating 3a) the input concept matches the content roles of both U.KURE- and U.AGE-. The rule we apply in such cases is the **principle of specificity** (Anderson, 1983): when more than one item matches an input, select the one which better

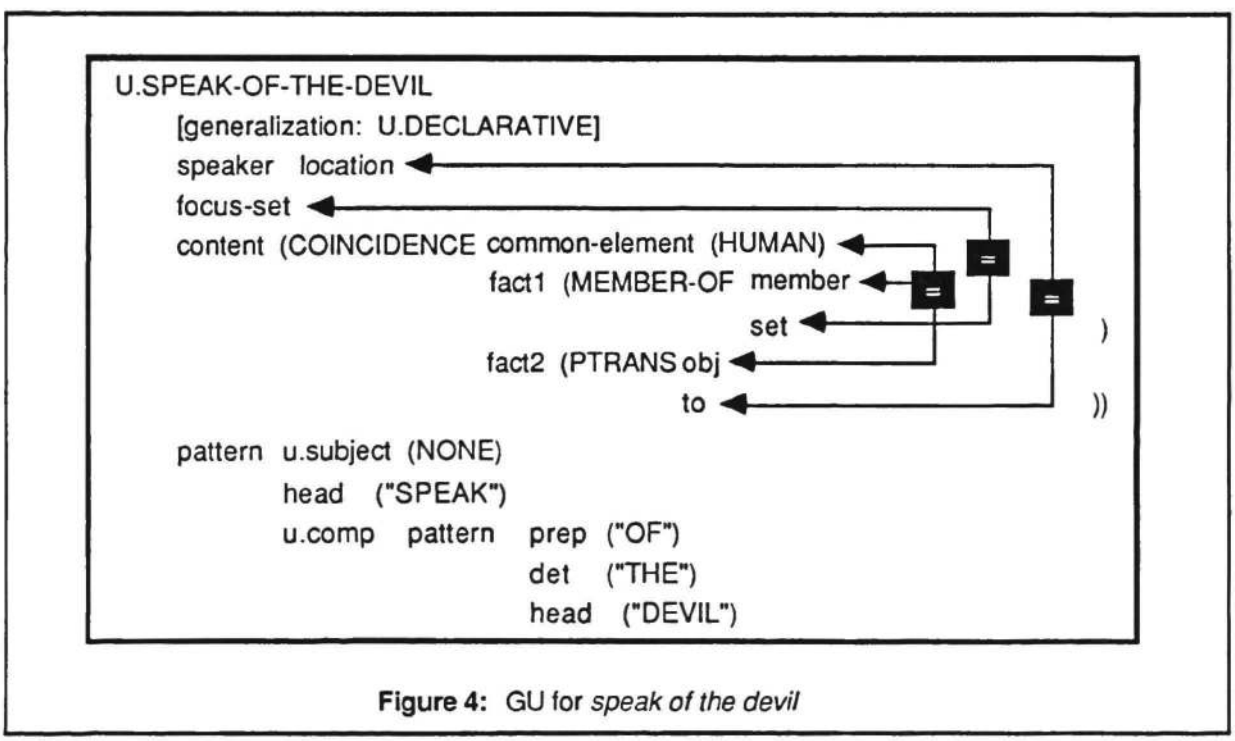

Figure 4: GU for *speak of the devil*

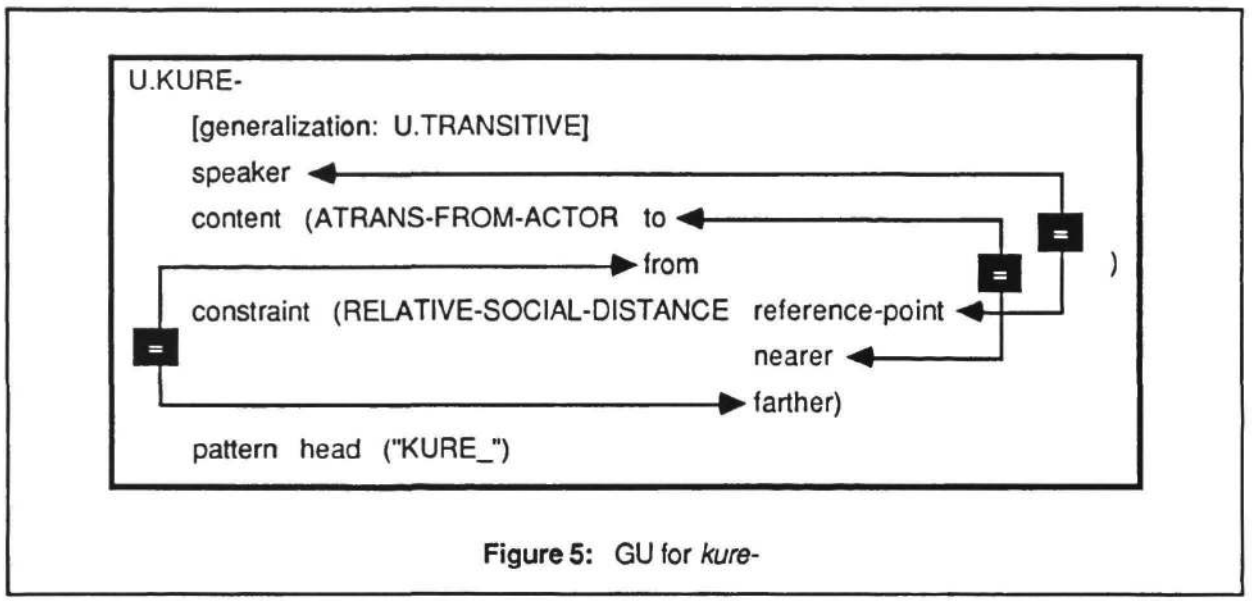

Figure 5: GU for *kure-*

characterizes the input. In this case this is U.KURE-, which matches on one more role than does U.AGE-.

INDIRECT SPEECH ACTS AND LITERAL MEANING

Now consider the directive in 4). For every directive there is a goal which the speaker hopes to achieve by having the listener perform some act. In other words, the plan to satisfy the top-level goal involves an **agency** subgoal (Schank & Abelson, 1977). The immediate goal behind a directive, then, is to have the listener *want* to perform the desired act. The GSA for directives is shown in Figure 6.

The specializations of SA.DIRECTIVE provide alternative ways of achieving the agency goal. These are distinguished on the basis of the speaker's expectation that the listener will perform the requested act. For **commands** the speaker is relatively certain of success. This expectation can derive either from the authority that the speaker has over the listener or the ease of the task. When the expectation of success is

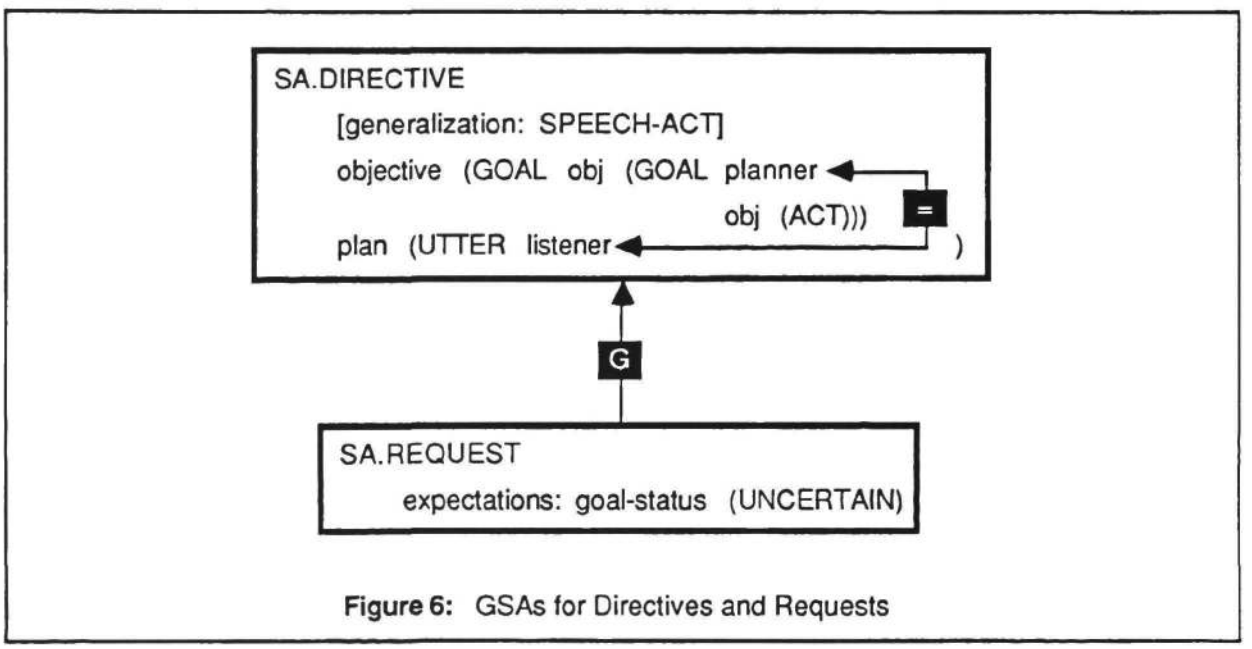

Figure 6: GSAs for Directives and Requests

not high, on the other hand, more polite directives, requests, are called for (Figure 6). Specializations of the request GSA specify particular request patterns. Figure 7 gives one of these, the GSA which is used in generating sentence 4). SA.WONDER associates its pattern with both its "indirect" and "literal" meanings. The indirect meaning is available because the objective of this GSA is that for requests. The literal meaning is accessible because the plan-utterance inherits from U.WONDER, the general GU for sentences with the verb *wonder*. The content of SA.WONDER's plan-utterance is inherited from U.WONDER. This concept, shown in the figure for convenience, is the literal meaning of *wonder*, i.e., a goal of the speaker to know some fact.

Note that the pattern provided by SA.WONDER is more general than *I was wondering if*. Rather it specifies that the subject refer to the speaker, that the head be a form of the verb *wonder*, and that the complement be a nominalized yes/no question. Thus the pattern also accommodates variants such as *we wonder whether*.

COMPARISON WITH OTHER APPROACHES

Existing natural language processing systems have been able to avoid some of the problems described in section 2 because they do not generally deal with unedited dialogue. Two areas that *have* received a good deal of attention, particularly in the language generation literature, are speech act planning (Appelt, 1985; Cohen & Perrault, 1979) and focus (Grosz, 1981; McKeown, 1985). While these researchers have made important progress in handling the pragmatic aspects of generation, all are limited by their failure to integrate the lexicon into the rest of memory. They maintain a functional separation of the grammar from the lexicon, and they make no attempt to incorporate lexical knowledge into other memory structures. This creates problems for representing lexical items that realize deictic or speech act functions or that are closely tied to particular social contexts. Work in the phrasal lexicon framework, on the other

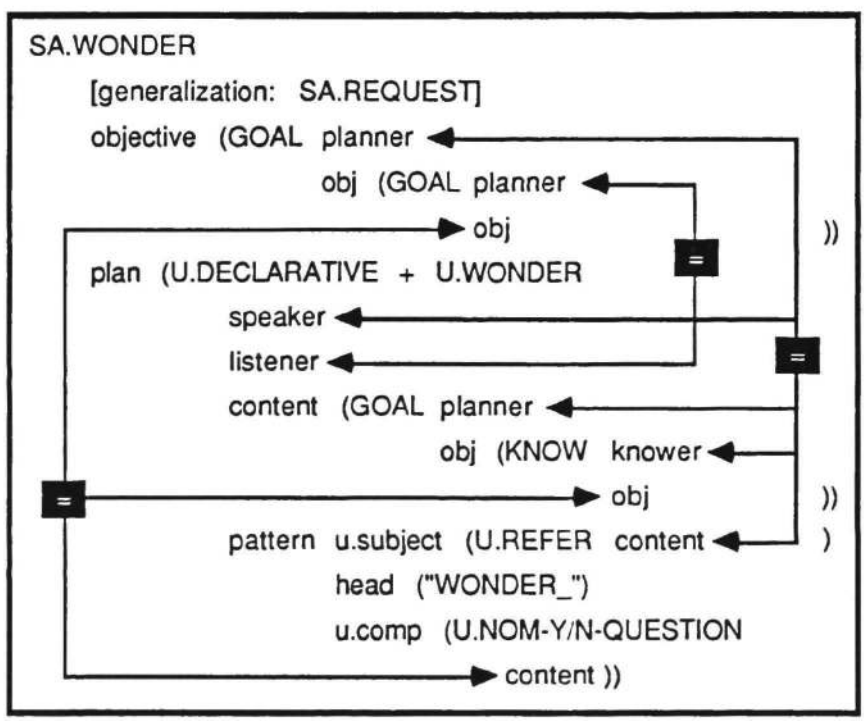

Figure 7: GSA for Requests Using *wonder*

hand, has to a certain extent eliminated the lexicon/phrase-structure distinction but has not generally addressed pragmatic issues.

In the **Lexical Memory** approach most patterns are amalgams of syntactic and lexical information. Those patterns which are purely syntactic, such as those specifying the structure for questions, are part of the same hierarchy as the lower-level "lexical" patterns. Because GUs have their deictic roles built in, any pattern may be assigned a deictic function. At the same time, because GUs fill the plan roles in GSAs, any pattern may also directly satisfy the goal behind a speech act type. The representation of context-oriented expressions like *your Honor* is also facilitated by the possibility of including GUs and GSAs as components of other memory structures, such as scripts.

The integrated nature of our approach distinguishes it in other ways from most mainstream linguistic theories. First, patterns in CHIE do not have a life of their own; they are defined only insofar as they fill a slot in a GU, which may in turn fill the plan slot in a GSA. In other words, they are always intimately associated with either semantic content or pragmatic intent. This contrasts with the autonomy of syntax position but is in agreement with "functional" views of syntax (e.g., Bates & MacWhinney, 1982; Langacker, 1986). Second, the fact that linguistic knowledge is represented in the same general framework as other knowledge and that lexical entries are integrated into conceptual memory is more than a computational convenience. This reflects an attempt to discover the extent to which linguistic knowledge shares properties with other kinds of knowledge. This approach stands in opposition to that of Chomskyan linguistics but is favored by an increasing number of linguists and psychologists (e.g., Anderson, 1983; Hudson, 1984; Langacker, 1986).

CURRENT STATUS AND FUTURE WORK

CHIE is implemented in AMI, a semantic network formalism based loosely on Fahlman's (1979) NETL. The program currently generates sentences for a small number of expressions of the type discussed in this paper. The work is still limited in several ways:

1. There is no account of where discourse goals come from; i.e., the program is not really ready to participate in conversations. A future goal of the project is to integrate the program into a model of the generation of conversation, such as that of Reichman (1985).
2. There is no facility for representing aspects of the listener's current knowledge state. Listener knowledge enters into the definition a number of lexical items, in particular, those of a "presuppositional" nature such as *again*. So that the program can make use of this sort of information in lexical entries, we are developing a procedure for "proving" whether the listener knows a certain fact or knows of a certain entity.
3. GUs and GSAs are *prototypical* associations, and CHIE cannot currently use them in atypical ways. A speaker may make *creative* use of an association in setting up an analogy between two contexts. The expression *your honor*, for example, might be used by a speaker addressing someone other than a presiding judge when a goal is to bring out the addressee's attempt to exercise judge-like authority in some context. We are currently working on a scheme for building analogies to satisfy generation goals.

CONCLUSIONS

In this paper we have presented an approach in which linguistic knowledge is integrated in various ways into the rest of memory:

1. GUs associate morpho-syntactic patterns directly with concept types.
2. GSAs associate goals with discourse plans that are realized as utterances. They are used in the same way as goal-plan associations that do not involve language.
3. Both GUs and GSAs are often embedded in and indexed to particular social situations.

CHIE extends the phrasal lexicon approach to allow lexical entries to make declarative reference to the elements of an utterance itself and to roles in the social context in which an utterance occurs. In addition, our approach has important implications for a model of language acquisition. GUs and GSAs consist of declaratively expressed generalizations about language behavior and thus present good candidates for the basic incremental units of language acquisition. Moreover, since GUs and GSAs may be components of non-linguistic memory structures, the acquisition of the linguistic units can be seen as part of the acquisition of the larger conceptual and situational structures. In this sense language acquisition becomes one aspect of learning about social interaction.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, S. R., & Keenan, E. L. (1985). Deixis. In T. Shopen (Ed.), *Language typology and syntactic description: Vol. 3. Grammatical categories and the lexicon.* Cambridge: Cambridge University Press.

Appelt, D. E. (1985). *Planning English sentences.* Cambridge: Cambridge University Press.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art.* Cambridge: Cambridge University Press.

Becker, J. D. (1975, June). The phrasal lexicon. *Proceedings of the Interdisciplinary Workshop on Theoretical Issues in Natural Language Processing,* Cambridge, MA.

Bresnan, J. W., & Kaplan, R. M. (1983). Lexical-functional grammar: A formal system for grammatical representation. In J. W. Bresnan (Ed.), *The mental representation of grammatical relations.* Cambridge, MA: MIT Press.

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

Dyer, M. G. (1983). *In-depth understanding: A computer model of integrated processing for narrative comprehension.* Cambridge, MA: MIT Press.

Fahlman, S. E. (1979). *NETL: A system for representing and using real-world knowledge.* Cambridge, MA: MIT Press.

Fillmore, C. G. (1979). Innocence: A second idealization for linguistics. *Proceedings of the 5th Annual Meeting of the Berkeley Linguistic Society,* 63-76.

Gasser, M. (1985). Second language production: Coping with gaps in linguistic knowledge (Technical Report UCLA-AI-85-18). Los Angeles: University of California, Los Angeles, Computer Science Department.

Gibbs, R. W., Jr. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory and Cognition, 8,* 149-156.

Goldman, N. M. (1975). Conceptual generation. In R. C. Schank (Ed.), *Conceptual information processing*. Amsterdam: North Holland.

Grosz, B. J. (1981). Focusing and description in natural language dialogues. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding*. Cambridge: Cambridge University Press.

Hatch, E. M., Flashner, V., & Hunt, L. (1986). The experience model and language teaching. In R. Day (Ed.), *Conversation in second language acquisition*. Rowley, MA: Newbury House.

Hudson, R. (1984). *Word grammar*. Oxford: Blackwell.

Jacobs, P. (1985). A knowledge-based approach to language generation. Unpublished doctoral dissertation, University of California, Berkeley.

Kay, M. (1979). Functional grammar. *Proceedings of the Fifth Annual Meeting of the Berkeley Linguistic Society*, 142-158.

Langacker, R. W. (1986). Cognitive grammar. *Cognitive Science, 10*, 1-40.

McDonald, D., & Pustejovsky, J. (1985). Description-directed natural language generation. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, 799-805.

McKeown, K. R. (1985). *Text generation*. Cambridge: Cambridge University Press.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication*. London: Longman.

Peters, A. M. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.

Reichman, R. (1985). *Getting computers to talk like you and me: Discourse context, focus, and semantics (an ATN model)*. Cambridge, MA: MIT Press.

Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Lawrence Erlbaum.

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

Wilensky, R. (1983). *Planning and understanding: A computational approach to human reasoning*. Reading, MA: Addison-Wesley.

Wilensky, R., & Arens, Y. (1980). PHRAN--A Knowledge-based Approach to Natural Language Analysis (Memorandum UCB/ERL M80/34). Berkeley, CA: University of California at Berkeley, Electronics Research Laboratory.

Zernik, U., & Dyer, M. G. (1985). Towards a self-extending lexicon. *Proceedings of the 23rd Annual Conference of the Association for Computational Linguistics*, 284-292.