THE REPRESENTATION OF MULTIMODAL USER INTERFACE DIALOGUES USING DISCOURSE PEGS

Susann Luperfoy
MITRE Corporation
7525 Colshire Blvd. W418
McLean, VA 22102
luperfoy@starbase.mitre.org

and

ATR
Interpreting Telephony Research Laboratories
Kyoto, Japan

ABSTRACT

The three-tiered discourse representation defined in (Luperfoy, 1991) is applied to multimodal human-computer interface (HCI) dialogues. In the applied system the three tiers are (1) a linguistic analysis (morphological, syntactic, sentential semantic) of input and output communicative events including keyboard-entered command language atoms, NL strings, mouse clicks, output text strings, and output graphical events; (2) a discourse model representation containing one discourse object, called a peg, for each construct (each guise of an individual) under discussion; and (3) the knowledge base (KB) representation of the computer agent's 'belief' system which is used to support its interpretation procedures. I present evidence to justify the added complexity of this three-tiered system over standard two-tiered representations, based on (A) cognitive processes that must be supported for any non-idealized dialogue environment (e.g., the agents can discuss constructs not present in their current belief systems), including information decay, and the need for a distinction between understanding a discourse and believing the information content of a discourse; (B) linguistic phenomena, in particular, context-dependent NPs, which can be **partially** or **totally** anaphoric; and (C) observed requirements of three implemented HCI dialogue systems that have employed this three-tiered discourse representation.

THE THREE-TIERED FRAMEWORK

This paper argues for a three-tiered computational model of discourse and reports on its use in knowledge based human-computer interface (HCI) dialogue. The first tier holds a linguistic analysis of surface forms. At this level there is a unique object (called a linguistic object or LO) for each linguistic referring expression or non-linguistic communicative gesture issued by either participant in the interface dialogue. The intermediate tier is the discourse model, a tier with one unique object corresponding to each concept or guise of a concept, being discussed in the dialogue. These objects are called **pegs** after Landman's theoretical construct (Landman, 1986a).[1] The third tier is the knowledge base (KB) that describes the belief system of one agent in the dialogue, namely, the backend system being interfaced to. Figure 1 diagrams a partitioning of the information available to a dialogue processing agent. This partitioning gives rise to the three discourse tiers proposed, and is motivated, in part, by the distinct processes that transfer information between tiers.

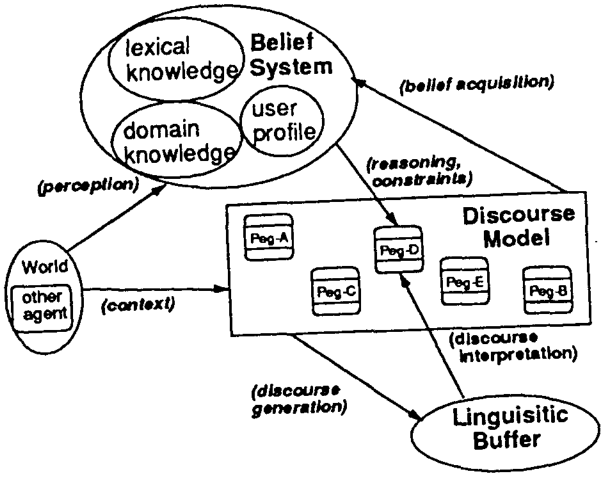

FIGURE 1. Partitioned Discourse Information

The linguistic tier is similar to the linguistic representation of Grosz and Sidner (1985) and its LO's are like Sidner's NP bundles (Sidner, 1979), i.e., both encode the syntactic and semantic analyses of surface forms. One difference, however, is that NP bundles specify database objects directly whereas LOs are instead "anchored" to pegs in the discourse model tier and make no direct connection to entries in the static

[1] The discourse peg functions differently from its namesake but the term provides the suitable metaphor (also suggested by Webber): an empty hook on which to hang properties of the real object. For more background on the Data Semantics framework itself see (Landman 1986b) and (Veltman, 1981).

knowledge representation. LOs are also like Discourse Referents (Karttunen, 1968), Discourse Entities ((Webber, 1978), (Dahl and Ball, 1990), (Ayuso, 1989), and others), File Cards (Heim, 1982), and Discourse Markers (Kamp, 1981) in at least two ways. First, they arise from a meaning representation of the surface linguistic form based on a set of generation rules which consider language-specific features, and facts about the logical form representation: quantifier scope assignments, syntactic number and gender markings, distributive versus collective reading information, ordering of modifiers, etc. Janus (Ayuso, 1989) allows for DE's introduced into the discourse context through a non-linguistic (the haptic) channel. But in Janus, a mouse click on a screen icon is assigned honorary linguistic status via the logical form representation of a definite NP, and that introduces a new DE into the context. WML, the intensional language used, also includes time and possible world parameters to situate DE's. These innovations are all important attributes of objects at what I have called the linguistic tier.

Secondly, the discourse constructs listed above all correspond either directly (Discourse Referents, File Cards, Discourse Entities of Webber) or indirectly after collapsing of referential equivalence classes (Discourse Markers, DE's of Janus) with referents or surrogates in some representation of the reference world, and it is by virtue of this mapping that they either are assigned denotations or fail to refer. While I am not concerned here with referential semantics I view this linguistic tier as standing in a similar relation to the reference world of its surface forms.

The pegs discourse model represents the world as the current discourse assumes it to be only, apart from how the description was formulated, apart from the true state of the reference world, and apart from how either participant believes it to be. This statement is similar to those of both Landman and Webber. The discourse model is also the locus of the objects of discourse structuring techniques, e.g., both intentional and attentional structures of Grosz and Sidner (1985) are superimposed on the discourse model tier. A peg has links to every LO that "mentions" it, the mentioning being either verbal or non-verbal and originating with either dialogue participant.

Pegs, like File Cards, are created on the fly as needed in the current discourse and amount to dynamically defined guises of individuals. These guises differ from File Cards in that they do not necessarily correspond 1:1 to individuals they represent, i.e., a single individual can be treated as two pegs in the discourse model, if for example the purpose is to contrast guises such as Superman and Clark Kent, without requiring that there also be two individuals in the knowledge structure. In comparing the proposed representation to those of Heim, Webber, and others it is also helpful to note a difference in emphasis. Heim's theory of definiteness defines semantic values for NPs based on their ability to add new File Cards to the discourse state, their "file change potential." Similarly, Webber's goal is to define the set of DE's justified by a segment of text. Examples of a wide range of anaphoric phenomena are used as evidence of which DEs had to have been generated for the antecedent utterance. Thus, the definition of Invoking Descriptions but no labels for subsequent mention of a DE or discussion of their affect on the DE.

In contrast, my emphasis is in tracking these representations over the course of a long dialogue; I have nothing to contribute to the theory of how they are originally generated by the logical form representation of a sentence. I am also concerned with how the subsequent utterance is processed given a possibly flawed or incomplete representation of the prior discourse, a possibly flawed or incomplete linguistic representation of the new utterance, and/or a mismatch between KB and discourse. The purpose here is to manage communicative acts encountered in real dialogue and, in particular, HCI dialogues in which the interpreter is potentially receiving information from the other dialogue participant with the intended result of an altered belief structure. So I include no discussion of the referential value of referring expressions or discourse segments, in terms of truth conditions, possible worlds, or sets of admissible models. Neither is the aim a descriptive representation of the dialogue as a whole; rather, the purpose is to define the minimal representation of one agent's egocentric view of a dialogue needed to support appropriate behavior of that agent in real-time dialogue interaction.

The remainder of this paper argues for the additional representational complexity of the separate discourse pegs tier being proposed. Evidence for this innovation is divided into three classes (A) cognitive requirements for processing dialogue, (B) linguistic phenomena involving context-dependent NPs, and (C) implementation-based arguments.

EVIDENCE FOR THREE TIERS

A. COGNITIVE PROCESSING CONSTRAINTS

This section discusses four requirements of discourse representation based on the cognitive limitations and pressures faced by any dialogue participant.

1.Incompleteness: The information available to a dialogue agent is always incomplete; the belief system, the linguistic interpretation, the prior discourse representation are partial and potentially flawed representations of the world, the input utterances, and the information content of the discourse, respectively. The distinction between discourse pegs and KB objects is important because it allows for a clear separation between what occurs in the discourse, and what is encoded as beliefs in the KB. The KB is viewed as a source of information consulted by one agent during language processing, not as the locus of referents or referent surrogates. Belief system incompleteness means it is common in dialogue to discuss ideas one is unfamiliar with or does not believe to be true, and to reason based on a partial understanding of the discourse. So it often happens that a discourse peg fails to correspond to anything familiar to the interpreting agent. Therefore, no link to the KB is required or entailed by the occurrence of a peg in the discourse model.

There are two occasions where the interpreter is unable to map the discourse model to the KB. The first is where the class referenced is unfamiliar to the interpreting agent, e.g., when an unknown common noun occurs and the interpreter cannot map to any class named by that common noun, e.g., "The picara walked in." The second is where the class is understood but the particular instance being referenced cannot be identified at the time the NP occurs. I.e., the interpreter may either not know of any instances of the familiar class, Picaras, or it may not be able to determine which of those picara instances that it knows of is the single individual indicated by the current NP. The pegs model allows the interpreter to leave the representation in a partial state until further information arrives; an underspecified peg for the unknown class is created and, when possible, linked to the appropriate class. As the dialogue progresses subsequent utterances or inferences add properties to the peg and clarify the link to the KB which becomes gradually more precise. But that is a matter between the peg and the KB; the original LO is considered complete at NP processing time and cannot be revisited.

2. Contradiction: Direct conflicts between what an agent believes about the world (the KB) and what the agent understands of the current discourse (the discourse model) are also common. Examples include failed interpretation, misunderstanding, disagreement between two negotiating parties, a learning system being trained or corrected by the user, a tutorial system that has just recognized that the user is confused, errors, lies, and other hypothetical or counterfactual discourse situations. But it is often an important service of a user interface (UI) to identity just this sort of discrepancy between its own KB information and the user's expressed beliefs. How the UI responds to recognized conflicts will depend on its assigned task; a tutoring system may leave its own beliefs unchanged and engage the user in an instructional dialogue whereas a knowledge acquisition tool might simply correct its internal information by assimilating the user's assertion.

To summarize 1 and 2, since dialogue in general involves transmission of information the interpreting agent is often unfamiliar with individuals being spoken about. In other cases, familiar individuals will receive new, unfamiliar, and/or controversial attributes over the course of the dialogue. Thirdly, on the generation side, it is clear that an agent may choose to produce NL descriptions that do not directly reflect that agent's belief system (generating simplified descriptions for a novice user, testing, game playing, etc.). In all cases, in order to distinguish what is said from what is believed, KB objects must not be created or altered as an automatic side effect of discourse processing, nor can the KB be required to be in a form that is compatible with all possible input utterances. In cases of incompleteness or contradiction the underspecified discourse peg holds a tentative set of properties that highlight salient existing properties of the KB object, and/or others that add to or override properties encoded in the KB.

3. Dynamic Guises: Landman's analysis of identity statements suggests a model (in a model-theoretic semantics) that contains pre-defined guises of individuals. In the system I propose, these guises are instead defined dynamically as needed in the discourse and updated non-monotonically. These are the pegs in the discourse model. Grosz (1977) introduced the notion of focus spaces and vistas in a semantic net representation for the similar purpose of representing the different perspectives of nodes in the semantic net that come into focus and affect the interpretation of subsequent NPs. What is in attentional focus in Grosz's system and in mine, are not individuals in the static belief system but selected views on those individuals and these are unpredictable, defined dynamically as the discourse progresses. I.e., it is impossible to know at KB creation time which guises of known individuals a speaker will present to the discourse. My system differs from the semantic net model in the separation it posits between static knowledge and discourse representation; focus spaces are, in effect, pulled out of the static memory and placed in the discourse model as a structuring of pegs. This eliminates the need to ever undo individual effects of discourse processing on the KB; the entire discourse model can be studied and either cast away after the dialogue or incorporated into the KB by an independent operation we might call "belief incorporation."

4. Information Decay: In addition to monotonic information growth and non-monotonic changes to the discourse model, the agent participating in a dialogue experiences information decay over the course of the conversation. But information from the linguistic, discourse, and belief system tiers decays at different rates and in response to different cognitive forces/limitations. (1) LOs become old and vanish at an approximately linear rate as a function of time counted from the point of their introduction into the discourse history, i.e., as LOs get older, they fade from the discourse and can no longer serve as linguistic sponsors[2] for anaphors; (2) discourse pegs decay as a function of attentional focus, so that as long as an individual or concept is being attended to in the dialogue, the discourse peg will remain near the top of the focus stack and available as a potential discourse sponsor for upcoming dependent referring expressions; (3) decay of static information in the KB is analogous to more general forgetting of stored beliefs/information which occurs as a result of other cognitive processes, not as an immediate side-effect of discourse processing or the simple passing of time.

B. LINGUISTIC EVIDENCE

This section sketches an analysis of context-dependent NPs to help argue for the separation of linguistic and discourse tiers. (Luperfoy, 1991) defines four types of context-dependent NPs and uses the pegs discourse framework to represent them: a dependent (anaphoric) LO must be **linguistically sponsored** by another LO in the linguistic tier or **discourse sponsored** by a peg in the discourse model and these two categories are subdivided into **total anaphors** and **partial anaphors**. Total anaphors are typified by coreferential, (totally dependent), definite pronouns, such as "himself" and "he" below, both of which are sponsored by "Karl."

Karl saw himself in the mirror. He started to laugh.

Partial anaphors depend on but do not corefer with their sponsors. Examples of partial anaphors have been discussed widely under other labels, by Karttunen, Sidner, Heim, and others, in examples like this one from (Karttunen, 1968)

I stopped the car and when I opened the hood I saw that the radiator was boiling.

where knowledge about the world is required in order to make the connection between dependent and sponsor, and others like Carlson's (1977)

Nancy hates racoons because they ate her corn last year.

where associating dependent to sponsor requires no specific world knowledge, only a general discourse principle about the ability of generic references to kinds (signalled by a bare plural NP in English) to sponsor dependent references to indefinite instances. (Substitute "picaras" for "racoons" in Carlson's example to demonstrate the independence of this phenomenon from world knowledge about the referent of the NP.)[3] This holds for mass or count nouns and applies in either direction, i.e., the peg for a specific exemplar can sponsor mention of the generic kind.

Nancy ate her oatmeal this morning because she heard that it lowers cholesterol.

The two parameters, partial/total dependence and linguistic/discourse sponsoring, classify all anaphoric phenomena (independently of the three-tiered framework) and yield as one result a characterization of indefinite NPs as potentially partially anaphoric in exactly the same way that definite NPs are.

I stopped the car and when I opened the hood I saw that a spark plug wire was missing.

The distinction between discourse sponsoring and linguistic sponsoring, plus the differential information decay rates for the three tiers discussed in Section A, together predict acceptability conditions and semantic interpretation of certain context-dependent NP forms. For example, the strict locality of one-anaphoric references is predicted by two facts:

(a) one-anaphors must always have a linguistic sponsor (i.e., an LO in the linguistic tier).

(b) these linguistic sponsor candidates decay more rapidly than pegs in the discourse model tier.

In contrast, definite NPs can be discourse sponsored. And the sponsoring peg may have been first introduced into the discourse model by a much earlier LO mention and kept active by sustained attentional focus. Thus, discourse- versus linguistic sponsoring helps explain why definite NPs can reach back to distant segments of the discourse history while one-anaphors cannot.[4]

Figure 2 illustrates the four possible discourse configurations for context-dependent NPs. The KB interface is omitted in the diagrams in order to show only the interaction between linguistic and discourse

[2] Discussed in next section.

[3] Compare this partial anaphor to the total anaphoric reference in, *Nancy hates racoons because they are not extinct.*

[4] For a detailed description of the algorithms for identifying sponsors and assigning pegs as anchors, for all NP types see (Luperfoy 1991) and (Luperfoy and Rich, 1992).

tiers, and dark arrows indicate the sponsorship relation. In each case, LO-1 is non-anaphoric and mentions Peg-A, its anchor in the discourse model. For the two examples in the top row LO-2 is linguistically sponsored by LO-1. Discourse sponsorship (bottom row) means that the anaphoric LO-2 depends directly on a peg in the discourse model and does not require sponsoring by a linguistic form. The left column illustrates total dependence, LO-1 and LO-2 are co-anchored to Peg-A. Whereas, in partial anaphor cases (right column), a new peg, Peg-B, gets introduced into the discourse model by the partially anaphoric LO-2.

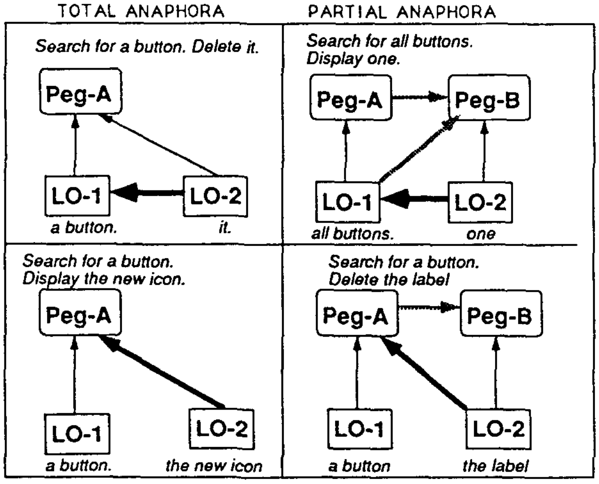

FIGURE 2. Four Possible Discourse Configurations For Anaphoric NPs

The classification of context-dependence is made explicit in the three-tiered discourse representation which also distinguishes incidental coreference from true anaphoric dependence. It supports uniform analysis of context-dependent NPs as diverse as reflexive pronouns and partially anaphoric indefinite NPs. The resulting relationship encodings are important for long-term tracking of the fate of discourse pegs. In File Change Semantics this would amount to recording the relation that justifies accommodation of the new File Card as a permanent fact about the discourse.

Furthermore, relationships between objects at different levels inform each other and allow application of standard constraints. The three tiers allow you to uphold linguistic constraints on coreference (e.g., syntactic number and gender agreement) at the LO level but mark them as overridden by discourse or pragmatic constraints at the discourse model level., i.e. apparent violations of constraints are explained as transfer of control to another tier where those constraints have no jurisdiction. In a two-tiered model coreferential LOs must be equated (or collapsed into one) or else they are distinct. Here, the discourse tier is not simply a richer analysis of linguistic tier information nor a conflation of equivalence classes of LOs partitioned by referential identity.

C. EVIDENCE BASED ON AN IMPLEMENTED SYSTEM

The discourse pegs approach has been implemented as the discourse component of the Human Interface Tool Suite (HITS) project (Hollan, et al. 1988) of the MCC Human Interface Lab and applied to three user interface (UI) designs: a knowledge editor for the Cyc KB (Guha and Lenat, 1990), an icon editor for designing display panels for photocopy machines, and an information retrieval (IR) tool for preparing multi-media presentations. All three UIs are knowledge based with Cyc as their supporting KB. An input utterance is normally a command language operator followed by its arguments. And an argument can be formulated as an NL string representation of an NP, or as a mouse click on presented screen objects that stand for desired arguments. Output utterances can be listed names of Cyc units retrieved from the knowledge base in response to a search query, self-narration statements simultaneous with changes to the screen display, and repair dialogues initiated by the NL interpretation system.

Input and output communicative events of any modality are captured and represented as pegs in the discourse model and LOs in the linguistic history so that either dialogue participant can make anaphoric reference to pegs introduced by the other, while the source agent of each assertion is retained on the associated LO.

The HITS UIs endeavor to use NL only when the added expressive power is called for and allow input mouse clicks and output graphic gestures for occasions when these less costly modalities are sufficient. The respective strengths of the various UI modalities are reviewed in (P. Cohen et al., 1989) which reports on a similar effort to construct UIs that make maximal benefit of NL by using it in conjunction with other modalities.

Two other systems which combine NL and mouse gestures, XTRA (Wahlster, 1989) and CUBRICON (Neal, et al., 1989), differ from the current system in two ways. First, they take on the challenge of ambiguous mouse clicks, their primary goal being to use the strengths of NL (text and speech) to disambiguate these deictic references. In the HITS system described here only presented icons can be clicked on and all uninterpretable mouse input is ignored. A second, related difference is the assumption by CUBRICON and XTRA of a closed world defined by the knowledge base representation of the current screen state. This makes it a reasonable strategy to attempt to coerce any uninterpretable mouse gesture into its nearest approximation from the finite set of target icons. In rejecting the closed world assumption I give up the constraining power it offers, in exchange for the ability to tolerate a partially specified discourse representation that is not fully aligned with the KB. In general, NL systems assume a closed world, in part because the task is often information retrieval or because in order for NL input to be of use it must resolve to one of a finite set of objects that can be acted upon. Because the HITS systems intended to generate and receive new information from the user, it is not possible to follow the approach taken in Janus for example, and resolve the NP "a button" to a sole instance of the class #%Buttons in the KB. Ayuso notes that this does not reflect the semantics of indefinite NPs but it is a shortcut that makes sense given the UI task undertaken.

In human-human dialogue many extraneous behaviors have no intended communicative value (scratching one's ear, picking up a glass, etc.). Similarly, many UI events detectable by the dialogue system are not intended by either agent as communicative and should not be included in the discourse representation, e.g., the user moving the mouse cursor across the screen, or the backend system updating a variable. In the implemented system NL and non-NL knowledge sources exchange information via the HITS blackboard (R. Cohen et al., 1991) and when a knowledge source communicates with the user a statement is put on the blackboard. Only those statements are captured from the blackboard and recorded in the dialogue. In this way, all non-communicative events are ignored by the dialogue manager.

Many of the interesting properties of this system arise from the fact that it is a knowledge-based system for editing the same KB it is based on. The three-tiered representation suits the needs of such a system. The HITS knowledge editor is itself represented in the KB and the UI can make reference to itself and its components, e.g., #%Inspector3 is the KB unit for a pane in the window display and can be referred to in the UI dialogue. Secondly, ambiguous reference to a KB unit versus the object in the real world is possible. For example, the unit #%Joseph and the person Joseph are both reasonable referents of an NP: e.g., "When was he born?" requests the value in the #%birthdate slot of the KB unit #%Joseph, whereas "When was it created?" would access a bookkeeping slot in that same unit. Finally, the need to refer to units not yet created or those already deleted would occur in requests such as, "I didn't mean to delete them" which require that a peg persist in focus in the discourse model independent of the status of the corresponding KB unit. These example queries are not part of the implementation but do exemplify reference problems that motivate use of the three-tiered discourse representation for such systems.

The dialogue history is the sequences of input and output utterances in the linguistic tier and is structured according to (Clark and Shaeffer 1987) as a list of contributions each of which comprises a presentation and an acceptance. This underlying structure can be displayed to the user on demand. The following example dialogue shows a question-answer sequence in which queries are command language atom followed by NL string or mouse click.

user:	SEARCH FOR **a Lisp programmer who speaks French**
system:	*#%Holm, #%Ebihara, #%Jones, #%Baker.*
user:	FOLLOWUP **one who speaks Japanese**
system:	*#%Ebihara*
user:	FOLLOWUP **her creator**
system:	*#%Holm*
user:	INSPECT **it**
system:	*#%Holm displayed in #%Inspector3*

Here, output utterances are not true generated English but rather canned text string templates whose blanks are filled in with pointers to KB units. The whole output utterance gets captured from the HITS blackboard and placed in the discourse history. The objects filling template slots generate LOs and discourse pegs which are then used by discourse updating algorithms to modify the focus stack. For example,

output-template:
#%Holm displayed in #%Inspector3.

causes the introduction of LOs and pegs for #%Holm and #%Inspector3. Those objects generated as system output can now sponsor anaphoric reference by the user.

A collection of discourse knowledge sources update data structures and help interpret context dependent utterances. In this particular application of the three-tiered representation, context-dependence is exclusively a fact about the arguments to commands since command names are never context-sensitive. Input NPs are first processed by morphological, syntactic, and semantic knowledge sources, the result being a 'context-ignorant' (sentential) semantic analysis with relative scope assignments to quantifiers in NPs such as "Every Lisp programmer who owns a dog." This analysis would in principle use the DE generation rules of Webber and Ayuso for introducing its LOs. Discourse knowledge sources use the stored discourse representation to interpret context-dependent LO's, including definite pronouns, contrastive one-anaphors,[5] reference with indexical pronouns (e.g. you, my, I, mouse-clicks on the desktop icons), and totally anaphoric definite NPs.[6] The discourse module augments the logical form output of semantic processing and passes the result to the pragmatics processor whose task is to translate the logical form interpretation into a command in the language of the backend system, in this case Cycl, the language of the Cyc knowledge base system.

Productive dialogue includes subdialogues for repairs, requests for confirmations, and requests for clarification (Oviatt et al., 1990). The implemented multimodal discourse manager detects one form of interpretation failure, namely, when a sponsor cannot be identified for an input pronoun. The discourse system initiates its own clarification subdialogue and asks the user to select from a set of possible sponsors or to issue a new NP description as in the example

user: EDIT **it**.

system: *The meaning of "it" is unclear.*

Do you mean one of the following?

<#%Ebihara> <#%Inspector3>

user: (mouse clicks on #%Inspector3)

system: *#%Inspector3 displayed in #%Inspector3*

The user could instead type "yes" followed by a mouse click at the system's further prompting or "no" in which case the system prompts for an alternative descriptive NP which receives from-scratch NL processing. During the subdialogue, pegs for the actual LO <LO-it> (the topic of the subdialogue) and for the two screen icons for #%Ebihara and #%Inspector3 are in focus in the discourse model.

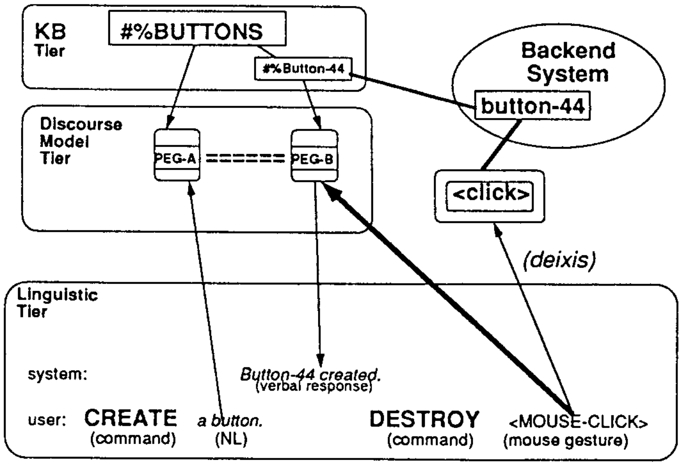

FIGURE 3. Three Tiers Applied to a Display Panel Design Tool

Figure 3 illustrates the arrangement of information structures in one multimodal HCI dialogue setting.[7] In this example, the user requests creation of a new button. Peg-A represents that hypothetical object. The system responds by (1) creating Button-44, (2) displaying it on the screen, and (3) generating a self-narration statement "Button-44 created." After the non-verbal event a followup deictic pronoun or mouse click, e.g., "Destroy that (button)" or "Destroy <mouse-click on Button-44>," could access the peg directly, but a pronominal reference, e.g., "Destroy it" would require linguistic sponsoring by the LO from the system's previous output statement. Because the system responded with both a graphical result and simultaneous self-narration statement in this example, either dependent reference type is possible. The knowledge based graphical knowledge source creates the KB unit #%Button44 as an instance of #%Buttons, but in this UI the user is unaware of the underlying KB and so cannot make or see references to KB units directly.

Note that Pegs A and B cannot be merged in the discourse model. The followup examples above only refer to that new Button-44 that was created. Alternatively (in some other UI) the user might have made total- and partial anaphoric re-mention of Peg-A by saying "Create a button. And make it a round one." The relationship between the two pegs is not identity. However this is not just a fact about knowledge acquisition interfaces, since the IR system might have allowed similar elaborated queries, "Search for a button, and make sure it's a round one."[8] The relationship between Pegs A and B arises from their being objects in a question-response pair in the structured dialogue history.

Finally, if the system is unable to map the word, say it were "knob," to any KB class then that constitutes a missing lexical item. Peg-A still gets created but it is not hooked up to #%Buttons (yet). In response to a 'floating' peg a UI system could choose to engage the user in a lexical acquisition dialogue, leave Peg-A underspecified until later (especially appropriate for text understanding applications), or associate it with the most specific possible node

[5] Luperfoy 1989 defines contrastive one-anaphora as one of three semantic functions of one-anaphora.

[6] Each anaphoric LO triggers a specialized handler to search for candidate sponsors (Rich and Luperfoy, 1988).

[7] Examples are representative of those of the actual system though simplified for exposition.

[8] Analogous to the issue in Karttunen's *John wants to catch a fish and eat it for supper.*

temporarily (e.g., #%Icons or #%PhysicalObjects). The eventual response may be to acquire a new class, #%Knobs, as a subclass of icons, or acquire a new lexical mapping from "knob" to the class #%Buttons.

The implemented systems which test the discourse representation were built primarily to demonstrate other things, i.e., to show the value of combining independent knowledge sources via a centralized blackboard mechanism and to explore options for combining NL with other UI modalities. Consequently, the NL systems were exercised on only a subset of their capabilities, namely, NP arguments to commands, which could be interpreted by most NLU systems. The dialogue situation itself is what argues for the separation of tiers.

CONCLUSION

The three-tiered discourse representation was used to model dialogue interaction from one agent's point of view. The discourse pegs level is independent of both the surface forms that occur and the immediate condition of the supporting belief system. In the implemented UI systems the discourse model provided a necessary buffer between the Cyc KB undergoing revision and the ongoing dialogue. However, most of the relevant considerations apply to other HCI dialogues, to human-human dialogues, and to NL discourse processing in general. I summarize the advantages of the pegs model under the original three headings and close with suggestions for further work.

(A) Cognitive considerations:

The belief system (KB) can serve dialogue processes as a source of information about the reference world without being itself modified as a necessary side effect of discourse interpretation. This means that understanding is not equated with believing, i.e., mismatch between pegs and KB objects is tolerated. Separate processes are allowed to update the KB in the background during discourse processing as the represented world changes and afterward, 'belief acquisition' can take care of assimilating pegs into the KB where appropriate.

The separation of tiers allows for differential rates of information decay. The linguistic tier fades from availability rapidly and as a function of time, discourse tier decay is conditioned by attentional focus, and the KB represents a static belief structure in which forgetting, if represented at all, is not affected by discourse processing.

Interpretation can be accomplished incrementally. The meaning of an NP is not defined as a KB object it corresponds to but as the peg that it mentions in the discourse model, and that peg is always a partial representation of the speaker's intended referent. How partial it is can vary over time and it can be of use for sponsoring dependent NPs, generating questions, etc., even in its partial state. Indeed, feedback from such use is what helps to further specify the peg.

(B) Linguistic phenomena:

In English, all NPs have the potential of being context-dependent. The separation of tiers allows for the distinction between true anaphoric dependence and incidental coreference, encoded as the co-anchoring of multiple LOs to a single peg without sponsorship. Partial and total anaphors are explicitly represented, with linguistic sponsoring distinguished from discourse sponsoring, and these relationships are stored as annotated links in the permanent discourse representation so that internal NL and non-NL procedures may query the discourse structure for information on coreference, KB property values, justifications for later links, etc.

The distinction between discourse and linguistic sponsoring allows language-specific syntactic and semantic constraints to be upheld at the LO level and overridden by pragmatic and discourse considerations at the discourse pegs level, thereby providing a mechanism which addresses well-known violations of linguistic constraints on coreference without relaxing the constraints themselves.

Input and output are distinguished at the linguistic tier but merged at the discourse model tier. The user can make anaphoric reference through any channel to pegs introduced by the backend system through any channel. Yet it remains part of the discourse history record in the linguistic tier, who made which assertions about which pegs. In the HCI dialogue environment this means that NL and non-NL modalities are equally acceptable as surface forms for input and output utterances, i.e., voice input could be added without extension to the current system as long as the speech recognizer output forms that could be used to generate LOs.

(C) Evidence from a trial implementation:

In knowledge-based UIs, the strict separation of tiers means that the KB can be incomplete or incorrect throughout the discourse, it can remain unaffected by discourse processing, and it can be updated by other knowledge acquisition procedures independently of simultaneous discourse processing. Nevertheless, it is possible and may be computationally efficient to implement the discourse model as a specialized, non-static (and potentially redundant) region of the KB so that KB reasoning mechanisms can be applied to the hypothetical state of affairs depicted by pegs in the discourse model.

The guise of an individual has just those properties assumed by the current discourse. Using pegs as dynamically defined guises in effect suppresses non-salient properties of the accessed KB unit. Thus Grosz's requirement that the discourse representation encode relations in focus as well as entities in focus is supported at the pegs level. Moreover, the three-tiered design can represent conflict between interpreted discourse information and the agent's static beliefs because KB values can be overridden in the discourse by ascription of contrary properties to corresponding pegs. A related benefit is that the external dialogue participant is allowed to introduce new pegs and new information into the discourse and this does not require creation of a new KB object during discourse interpretation.

Because pegs are used to accumulate tentative properties on (actual or hypothetical) individuals without editing the KB either permanently or only for the duration of the discourse, belief acquisition can be postponed until a sufficiently complete understanding has been achieved, so the discourse model can serve as an agenda for later KB updating. Meanwhile, partial and incorrect discourse representations are useful and non-monotonic repair operations make it easy to correct interpretation errors by changing links between LO and peg or between peg and KB unit without disturbing other links.

Some pegs are not associated with the linguistic tier at all. Graphical events in the physical environment that make an object salient can inject a peg directly into the discourse model. However, only pegs introduced via the linguistic channel can sponsor linguistic anaphora, e.g., "What is it" requires the presence of an LO, but "What is that" can be sponsored directly by the peg for an icon that just appeared on the screen.

Further Research

Dependents can sponsor other dependents, and in general, there is complex interaction between sequences of NPs in a discourse. For example, in the sentence

Delete the buttons if one of them is missing its label.

its label is partially dependent on *one*, and *it* is totally dependent on *one* which is partially dependent on *them* which is totally dependent on *the buttons* which is presumably a total anaphoric reference to a discourse peg for some set of buttons currently in focus. The present algorithm attempts pseudo-parallel processing of LOs, taking repeated passes through the new utterance, left to right by NP type, (proper nouns, definite NPs,..,reflexives). One-anaphors modified by partitive PPs are exceptional in that they are processed after the pronoun or definite NP (the object of the preposition) to their right. Further work is needed to describe the ways that various NP types interact as this was a technique for coping with the absence of a theory of the possible relationships between sequences of partial and total anaphoric NPs.

LOs for events are created by the semantic processing module and so sequences such as:

You deleted that unit. I didn't want to do that.

could in theory be handled analogously with other partial and total anaphors. However, they are not of use in the current application UIs and so their theory and implementation have remained undeveloped here.

Ambiguous mouse clicks of the sort explored in XTRA and CUBRICON plus the ability of the user to introduce new pegs for regions of the screen, or for events of moving a pane or icon across the screen, or encircling a set of existing icons to place their pegs in attentional focus should all be attempted using the pegs discourse model as a source of target interpretations of mouse clicks and as a place to encode novel, user-defined screen objects.

Finally, with this or other representations of dialogue, a variety of UI metaphors should be explored. The UI can be viewed as a single autonomous agent or as merely the clearing house for communication between the user and a collection of agents, the operating system, the graphical interface, the NL system, or any of the knowledge sources, such as those on the HITS blackboard, which could conceivably want to engage the user in a dialogue.

The three-tiered discourse design is also used in the knowledge based NL system at MCC (Barnett, et al., 1990), and is being explored as one descriptive device for dialogue in voice-to-voice machine translation at ATR.

ACKNOWLEDGEMENTS

This system was designed and developed in cooperation with Kent Wittenburg, Richard Cohen, Paul Martin, Elaine Rich, Inderjeet Mani, and other former members of the MCC Human Interface Lab. I would also like to thank members of the ATR Interpreting Telephony Research Laboratories and anonymous reviewers for valuable comments on an earlier draft of this paper.

REFERENCES

Ayuso, Damaris (1989) Discourse Entities in Janus. Proceedings of the 27th Annual Meeting of the ACL. pp.243-250.

Barnett, James, Kevin Knight, Inderjeet Mani, and Elaine Rich (1990) A Knowledge-Based Natural

Language Processing System. Communications of the ACM.

Carlson, Gregory (1977). A Unified Analysis of the English Bare Plural. *Linguistics and Philosophy*, 1, 413-457.

Clark, Herbert and E. Schaefer. (1987). Collaborating on Contributions to Conversations. *Language and Cognitive Processes*, pp. 19-41.

Cohen, Richard, Timothy McCandless, and Elaine Rich, A Problem Solving Approach to Human-Computer Interface management, MCC Tech Report ACT-HI-306-89, Fall 1989.

Cohen, Philip, Mary Dalrymple, Douglas B. Moran, Fernando C.N. Pereira, Joseph W. Sullivan, Robert A. Gargan, Jon L. Schlossberg and Sherman W. Tyler. (1989) *Synergistic Use of Direct Manipulation and Natural Language.* In Proceedings of CHI, pp. 227-233.

Dahl, Deborah and Catherine N. Ball. (1990). *Reference Resolution in PUNDIT* (Tech. Report). UNISYS.

Grosz, Barbara (1977). *The Representation and Use of Focus in a Systemfor Understanding Dialogs.* In Proceedings of IJCAI 5.

Grosz, Barbara and Candace Sidner (1985) *The Structures of Discourse Structure* (Tech. Report). SRI International

Guha, R. V. and Douglas Lenat. (1990). Cyc: A Mid-Term Report. *AI Magazine*

Heim, Irena (1982) The Semantics of Definite and Indefinite Noun Phrases. U of Massachusetts, PhD Thesis.

Hollan, James, Elaine Rich, William Hill, David Wroblewski, Wayne Wilner, Kent Wittenburg, Jonathan Grudin, and members of the Human Interface Laboratory. (1988). *An Introduction to HITS: Human Interface Tool Suite* (Tech Report). MCC

Karttunen, Lauri (1968) What Makes Definite Noun Phrases Definite? Technical Report, Rand Corp.

Karttunen, Lauri (1976) Discourse Referents. In McCawley, J. (ed.), *Syntax and Semantics.* Academic Press, New York.

Landman, F. (1986) Pegs and Alecs. *Linguistics and Philosophy*, pp. 97-155.

Landman, F. (1986) Data Semantics for Attitude Reports. *Linguistics and Philosophy*, pp. 157-183.

Luperfoy, Susann (1989) The Semantics of Plural Indefinite Anaphors in English. *Texas Linguistic Forum.* pp. 91-136.

Luperfoy, Susann (1991) *Discourse Pegs: A Computational Analysis of Context-Dependent Referring Expressions.* Doctoral dissertation, Department of Linguistics, The University of Texas.

Luperfoy, Susann and Elaine Rich (1992) A Computational Model for the Resolution of Context-Dependent References. (in submission)

Neal, Jeanette, Zuzana Dobes, Keith E. Bettinger, and Jong S. Byoun (1990) Multi-Modal References in Human-Computer Dialogue. Proceedings of AAAI. pp 819-823

Oviatt, Sharon L., Philip R. Cohen and Ann Podlozny (1990) *Spoken Language in Interpreted Telephone Dialogues.* SRI International Technical Note 496.

Rich, Elaine A. and Susann Luperfoy (1988) An Architecture for Anaphora Resolution, *Proceedings of Applied ACL.*

Sidner, Candace L. (1979) *Towards a Computational Theory of Definite Anaphora Comprehension in Discourse.* Doctoral dissertation, Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Veltman, F. (1981) Data Semantics In Groenendijk, J. A. G., T. M. V. Janssen and M. B. J. Stokhof (eds.) *Formal Methods in the Study of Language Part 2*, Amsterdam: Mathematisch Centrum.

Wahlster, Wolfgang. (1989) User and Discourse models for multimodal Communication. In J.W. Sullivan and S.W. Tyler, eds., *Architectures for Intelligent Interfaces: Elements and Prototypes*, Addison-Wesley, Palo Alto, CA.

Webber, Bonnie L. (1978) *A Formal Approach to Discourse Anaphora.* Doctoral dissertation, Division of Applied Mathematics, Harvard University.