Research Article

Natural Conversational Interfaces to Geospatial Databases

Guoray Cai
School of Information Sciences and Technology and GeoVISTA Center Pennsylvania State University

Hongmei Wang
School of Information Sciences and Technology and GeoVISTA Center Pennsylvania State University

Alan M. MacEachren
Department of Geography and GeoVISTA Center Pennsylvania State University

Sven Fuhrmann
Department of Geography and GeoVISTA Center Pennsylvania State University

Abstract
Natural (spoken) language, combined with gestures and other human modalities, provides a promising alternative for interacting with computers, but such benefit has not been explored for interactions with geographical information systems. This paper presents a conceptual framework for enabling conversational human-GIS interactions. Conversations with a GIS are modeled as human-computer collaborative activities within a task domain. We adopt a mental state view of collaboration and discourse and propose a plan-based computational model for conversational grounding and dialogue generation. At the implementation level, our approach is to introduce a dialogue agent, *GeoDialogue*, between a user and a geographical information server. *GeoDialogue* actively recognizes user's information needs, reasons about detailed cartographic and database procedures, and acts cooperatively to assist user's problem solving. *GeoDialogue* serves as a semantic 'bridge' between the human language and the formal language that a GIS understands. The behavior of such dialogue-assisted human-GIS interfaces is illustrated through a scenario simulating a session of emergency response during a hurricane event.

Address for correspondence: Guoray Cai, School of Information Sciences and Technology and GeoVISTA Center, Pennsylvania State University, University Park, PA 16802, USA. E-mail: cai@ist.psu.edu

1 Introduction

Today, the majority of geographical information users are not experts in operating a geographical information system (GIS). However, the familiar devices (keyboard and mouse), interface objects (windows, icons, menus, and pointers), and query languages tend to work only for experts in a desktop environment. Practical application environments often introduce an intermediary person to delegate the tasks of communicating with a computer to technical experts (Mark and Frank 1992, Traynor and Williams 1995), but such solutions are not always possible when geographical information needs arise outside of the office environment (in the field or on the move) (Zerger and Smith 2003). Alternatively, human-GIS interfaces can be made more natural and transparent so that people can walk-up to the system and start utilizing geographical information without prior training. Towards this goal, progress has been made in the incorporation of human communication modalities into human-computer interaction systems (Zue et al. 1990; Shapiro et al. 1991; Lokuge and Ishizaki 1995; Oviatt 1996, 2000; Cohen et al. 1997; Sharma et al. 1998; Kettebekov et al. 2000; Rauschert et al. 2002). Designing such interface environments faces a number of challenges, including sensing and recognition, multimodal fusion, as well as semantic mediation and dialogue design. Of these issues, sensing technologies have made the most progress, particularly in the areas of automated speech recognition (Juang and Furui 2000, O'Shaughnessy 2003) and gesture recognition (Sharma et al. 1999, Wilson and Bobick 1999). Totally device-free acquisition of human speech and free-hand gestures has been demonstrated to be feasible for interacting with maps (Sharma et al. 2003). In contrast, multimodal fusion and dialogue management seems more difficult, and solutions are likely to depend on tasks and application domains (Flanagan and Huang 2003). Within the domain of geographical information science, a dialogue-based interface for GIS was envisioned more than a decade ago (see Frank and Mark 1991), but has not been attempted seriously.

This paper introduces the concept of conversational dialogues as a new paradigm of human-GIS interactions. Dialogue-based interaction with GIS differs from the traditional query/response style of interaction in that it requires modeling human-GIS interactions at the discourse level. In addition to taking user's input and extracting commands, the system is expected to actively engage in conversations with the user. The use of "conversation" as a metaphor for human-GIS interaction is particularly attractive in the context of multimodal interfaces for a number of reasons:

1. *Conversations ease the problems of recognition errors.* No current multimodal (speech-gesture) interface is free from *recognition errors*. In human-computer communication, the system should be able to detect its own misrecognitions and initiate dialogues for correcting errors before continuing with the interaction. Wang (2003) demonstrated that speech recognition errors can be 'repaired' using a fuzzy grammar approach. Alternatively, a conversational dialogue system can be a graceful error-correction mechanism.
2. *Conversations make it possible to construct requests incrementally and interactively.* Traditional query-based systems enforce a strict three-phase process: collecting user input, processing query, and generating response, where each phase must be complete and successful before moving to the next phase. However, natural multimodal requests to GIS rarely follow such artificial patterns. It is much easier for humans to specify a complex request in multiple steps, each of which is followed by

a grounding process. Such interactions are best managed as a conversational dialogue where both the human and the system keep track of the dialogue context necessary for grounding the meanings of subsequent inputs.

3. *Conversation is the way to deal with vagueness and ambiguity.* Natural language requests for geographical information often include concepts (spatial or non-spatial) that are vague or ambiguous. The key to processing such requests is to incorporate machine intelligence so that the machine makes an intentional effort to understand the context within which the user makes use of the vague concepts. Through sharing contextual knowledge with the user, the system can avoid misunderstanding of such concepts (see Cai et al. 2003 for an example). A shared visual display can provide both a shared context and *boundary objects* through which meaning is negotiated (MacEachren and Brewer 2004).

4. *Conversations foster human-GIS joint problem-solving.* Professionals are experts in their problem domains and associated tasks, but not in the use of GIS. Conversational interfaces for GIS have the potential of enabling computers to participate in human's problem-solving activities. The goal here would be to reduce the user's cognitive load by communicating in the language of the application domain that the user is familiar with, and by making relevant information available proactively.

The approach we take to enable conversational human-GIS interactions is to add a dialogue agent between the user and a GIS. As part of our research prototype DAVE_G (*Dialogue-Assisted Virtual Environment for GeoInformation*) that supports multimodal interactions with GIS (Rauschert et al. 2002), we have developed a computational agent, *GeoDialogue*, which implements the idea of conversational dialogues for human-GIS interactions. The design of *GeoDialogue* draws an analogy between a conversational dialogue agent and the role of the GIS specialist who traditionally serves as the 'mediator' between professionals (who perform spatial tasks in a domain) and the GIS in many workplaces. As earlier work (Mondschein 1994, Armstrong and Densham 1995, Jones et al. 1997) has shown, the tasks of a GIS specialist in a group decision-making situation are usually to listen to and discuss information needs with other users, and to translate their understanding of the information needs into cartographic and database operations executable by a particular GIS. Having such a GIS specialist (when properly trained) allows professionals to request a map simply by having a natural conversation with a GIS specialist. As appropriate as it can be, Armstrong (2001) and Nyerges et al. (1997) referred to such a GIS specialist as the "chauffeur". Our design of conversational dialogue systems has the goal of capturing the reasoning and communication skills of the GIS "chauffeur" and making them available through a software agent. In the long run, the goal of *GeoDialogue* is to eliminate the human-GIS interaction bottleneck caused by the 'middle person' (Mondschein 1994, Traynor and Williams 1995) so that user interactions with geographical information become more direct.

Like any intelligent agents (Jennings et al. 1998), *GeoDialogue* has the ability of assimilating perceived information (user's multimodal input), performing some cognitive computation, and controlling actions executed on the environment (i.e. spatial databases and information displays). When functioning within a multimodal interface, *GeoDialogue* extracts and integrates semantic information from multiple streams of user input (speech, gesture, gaze etc.), reasons on user's intentions and beliefs using domain and task knowledge, and engages in dialogue with the user utilizing the available store of geospatial data.

Table 1 Human-GIS interactions in three different styles

Query and Answer	Informational Conversation	Problem-solving Conversation
User: Zoom to Florida Sys: [show a map] User: Show roads System: I do not understand User: Show Interstate highways Sys: [show a map] User: Thanks	User: Show me a map of Florida Sys: I have shown State boundaries and highways User: Any population data? Sys: I have 'population by county' and 'population by census block'. Which dataset do you want to see? User: show population by county.	User: There is traffic accident causing a nuclear particulate leaking at this intersection[gesture]. Sys: The wind condition on this location is 15 mph. Would you like to see a Plume model? User: Yes. In addition, tell me what will be impacted. Sys: [show a map] there are three residential areas, one school, and one medical facilities within the pollution zone.

Conversational interfaces provide open-ended opportunities to address human-GIS interaction issues. Our work has followed a road map that transforms human-GIS interaction style from simple 'query-and-answer', to mixed-initiative informational conversations, and eventually to the stage of conversation-enabled human-GIS problem-solving. These three styles of interaction are illustrated in Table 1 using short examples. The first one exemplifies what traditional query-answer systems do. Both 2 and 3 are conversational dialogues, but they are different in the sense that the former are conversations about information retrieval and visualization tasks, while the latter are conversations about user's problem-solving activities. In the remainder of the paper, we will focus on principles and techniques for enabling informational conversations, although our dialogue agent, *GeoDialogue*, provides the infrastructure to deal with the challenges of problem-solving conversations as well. Current implementation and functionalities of *GeoDialogue* will be described in detail.

2 Related Work

A conversation has two major elements: the modalities used and the structure of the discourse. The advances in multimodal interfaces have focused on introducing human modalities into human-computer interaction systems. In order to participate in a conversation, one must maintain a representation of the intentional and attentional structures of the dialogue. This section reviews some relevant work in these two areas.

2.1 Multimodal Interfaces

Using human modalities for interacting with computers has undergone several decades of research (Bolt 1980, Sharma et al. 1998, Juang and Furui 2000, Zue and Glass 2000). Early systems such as *Voyager* (Zue et al. 1990) and *GeoSpace* (Lokuge and Ishizaki

1995) use speech input only. Speech provides an effective and direct way of expressing actions, pronouns and abstract relations. However, using speech alone for interacting with a GIS can be cumbersome when spatial references to regions and features on the map are needed. Gestures offer an effective second modality that is more suitable for expressing spatial relations. Specifications of user information needs using a combination of speech and gesture were shown to be less error prone than those expressed in words alone (Oviatt 1996). *CUBRICON* (Neal et al. 1989, 1998) was the first multimodal system that incorporated natural language and gestures into a GIS query interface. A more recent system, *QuickSet* (Cohen et al. 1997), uses speech and pen-based gestures to enable multimodal interactions with maps, and was shown to be more expressive and efficient than 'traditional' WIMP (windows, icons, menus, and pointers) interfaces, especially in a mobile computing environment (Oviatt and Cohen 2000). Similarly, *Sketch and Talk* (Egenhofer 1996) processed speech and direct sketch on a map display for querying spatial databases.

The systems mentioned above require the use of devices (such as pens) to capture gesture input, which may interfere with the user's ability to focus on the problem itself. Using free hand gestures to interact with maps was envisioned by the *WallBoard* concept (Florence et al. 1996), but did not become a reality until the success of *i*Map (Sharma et al. 1998, 1999; Kettebekov and Sharma 2000), which demonstrated the feasibility of free hand gestures as a new modality for human computer interaction. The general framework of *i*Map has recently been extended by *DAVE_G* (Dialogue-Assisted Virtual Environment for GeoInformation) (Rauschert et al. 2002), which supports speech/gesture interactions with large-screen map displays driven by GIS.

In *i*Map, as well as in earlier versions of *DAVE_G* (MacEachren et al. 2005), human use of speech and gestures are highly constrained to the expression of GIS commands and their parameters. Each request carries a very explicit meaning that directly maps to GIS actions by semantic grammar-based translation rules. Such a simplistic model of multimodal interactions does not reflect the complexity in practical uses of GIS. Human interactions with GIS are part of their problem-solving process that involves not only database and visualization commands, but also steps for defining and discussing a task, exploring ways to perform the task, and collaborating to get it done. Each step of human-GIS interactions is embedded within the larger structure of a problem-solving dialogue that provides the contexts for planning system's actions and for evaluating the effect of such actions. For this reason, it is necessary for computers to have a model of the discourse if a GIS is to be more cooperative and helpful to human problem-solving activities.

2.2 Discourse Models

Research on conversational human-computer interfaces (Zue and Glass 2000, Allen et al. 2001) explicitly models human-computer interactions on the principles of human-human communication. When humans solve problems together, they must communicate their understanding of the problem and construct solutions. Such processes involve extended dialogues where utterances and groups of utterances relate to each other in a coherent manner to form a discourse. The key for discourse processing is the recognition and representation of discourse structure (Lochbaum et al. 2000). Approaches to discourse structures generally fall into two categories: *informational* and *intentional*. Informational approaches model discourse structure as text units and a set of coherence

relations (such as Cause, Evaluation, Background, Elaboration, Purpose, and Effect) (Hobbs 1979, Mann and Thompson 1987) among text units. These works provide solutions to problems such as references and syntactic ambiguities, but they lack the reasoning capability necessary for modeling cooperative behavior of conversations. In contrast to informational approaches, Grosz and Sidner (1986, 1990) argues that discourse is inherently intentional. Their theory of discourse structure recognized three interrelated components of a discourse: a linguistic structure (discourse segments and embedding relations), intentional structure (purposes of discourse segments and their interrelations), and an attentional state (a record of salient entities at any time in the discourse).

A discourse is fundamentally a collaborative behavior (Grosz and Sidner 1990). Based on this notion, Lochbaum (1994, 1998) developed a model of intentional structure using the collaborative planning framework of SharedPlans (Grosz and Kraus 1996). In the SharedPlans formalism, a plan consists of a set of complex mental attitudes (beliefs, intentions, and commitments) towards a joint goal and its subgoals. A set of agents have a full SharedPlan (FSP) when all the mental attitudes required for successful collaboration have been established; otherwise, a SharedPlan is considered partial. The generation of a discourse can be modeled as a process that conversational participants elaborate on a partial SharedPlan towards a full SharedPlan.

There currently exist spoken dialogue interfaces with conversational properties. For example, the MIT Voyager system (Glass et al. 1995) can engage in limited verbal dialogue with users about common geographical knowledge in a region (such as hotels, restaurants, banks, as well as distance, directions, and travel time). AT&T's "How May I Help You" system (Gorin et al. 1997) can automatically route telephone calls to appropriate destinations in a telecommunications environment. A recent survey of existing spoken dialogue research projects can be found in McTear (McTear 2002).

The work reported in this paper builds on the success of conversational dialogue technologies and multimodal GIS systems, with the intention of integrating the two components for the development of more natural interfaces to interactive maps. Our work on *GeoDialogue* integrates both the informational approaches and intentional approaches of discourse structure for the development of conversational human-computer interfaces with geospatial databases. With a collaborative planner embedded in the dialogue engine, *GeoDialogue* shares the same objective with the Copas and Edmonds' interactive planners (Copas and Edmonds 2000) in overcoming the usability difficulties of high-functionality information systems (Fischer 2001).

3 GeoDialogue: Managing Conversations with Interactive Maps

GeoDialogue is a software agent that mediates natural conversational dialogues between users and geographical information systems. By natural, we mean that the system can understand and act upon what people naturally say rather than forcing them to make requests in a formal command-like language. By adopting a human conversation metaphor for information-seeking with a GIS, *GeoDialogue* makes the processes of browsing, discussing, filtering, and summarizing more interactive through conversational acts. The design goal was to enable natural, multimodal dialogue with geographical information displays. We focus initially on geographical information retrieval and visualization activities. In this section, we first introduce the design principles of *GeoDialogue*, and then describe the architecture and functionalities of *GeoDialogue* as implemented.

3.1 Design Principles

In *GeoDialogue*, the process of communication between the system and the user is modeled after the principles of human-human communication. Here, a computer is treated as an intelligent agent capable of rational and cooperative behavior (Bratman 1992). Human-GIS interaction is viewed as a goal-directed activity that involves collaborative planning and coordination between a human agent and a computer agent (Terveen 1995). For such interactions, there exists a direct correspondence between the intentional structure of a discourse and the structure of the tasks (goals and solutions) under discussion. *GeoDialogue* explicitly represents and reasons about the intentional structure and attentional state of human-GIS dialogues and uses such knowledge for interpreting spoken inputs and generating responses (speech output and interactive maps).

Conversations with geographical information through *GeoDialogue* are mixed-initiative (Hagen 1999), meaning that both the user and the system can initiate a new agenda. The system knows when to take, keep, and relinquish control and initiatives, as well as recognizes when the user takes, keeps, and relinquishes control and initiatives. The system may choose to follow a user's initiative or make a new initiative, depending on the need for advancing the agenda. For example, when *GeoDialogue* serves the role of a geographical information assistant, the system will yield control to the user on higher-level (domain related) intentions, and will take controls when the focus of the agenda moves to low-level data retrieval and presentation tasks. In this way, the user offloads some of the cognitive efforts to the computer while still feeling the 'steering' of the interaction. In particular, *GeoDialogue* tends to take initiatives when it detects an opportunity to protect the user from doing erroneous actions (by rejecting these actions), to correct user's misconceptions, and to volunteer choices and constraints while the user is making a decision. We will show how our model of human-GIS dialogues allows the system and the user to alternate the control of dialogue initiatives based on the status of their tasks and collaboration.

3.2 Representation of the Discourse Contexts

In *GeoDialogue*, the discourse context of a human-computer conversation is represented as a plan graph (or PlanGraph). A PlanGraph is similar to the notion of recipe graph (Rgraph) developed by Lochbaum (1994, 1998), except that PlanGraph extends Rgraph on the handling of knowledge-preconditions in collaborative plans. Before describing the structure of PlanGraph, we need to introduce three important concepts: *actions*, *recipes*, and *plans*.

An *action* refers to a specific goal as well as the effort needed to achieve it. In the knowledge store of *GeoDialogue*, an action can be either basic or complex. A *basic* action is directly executable by one or more agents. Examples of basic action may be 'retrieving a named map layer', or 'making a buffer around a known point', which can be directly executed by a GIS. A *complex* action, on the other hand, is not directly executable because certain knowledge pre-conditions or details about performing the action are subject to elaboration and instantiation. For each complex action α, *GeoDialogue* knows one or more possible ways (called *recipes*) to implement it. A recipe of an action encodes the system's knowledge about the abstract and schematic structure of that action. A recipe describes components of an action in terms of parameters,
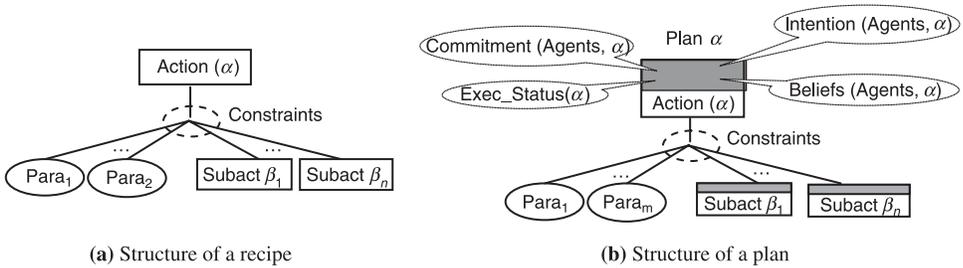
(a) Structure of a recipe   **(b)** Structure of a plan

Figure 1   The concepts of action, recipe, and plan

subactions, and constraints (see Figure 1a). Parameters of a recipe describe the knowledge pre-condition for executing subactions of that recipe. Subactions in a recipe have access to all the parameters in that recipe. All parameters of the recipe must be identified (i.e. instantiated with proper values) before subactions can be executed. *GeoDialogue*'s recipe definition language also supports constraints, which specify any pre- or post-conditions and partial orders of the subactions.

GeoDialogue separates the notion of *recipes* from that of *plans*. A recipe for a complex action α describes how it is decomposed into subgoals in a domain. A *plan*, on the other hand, corresponds to a schema describing not only how to perform an action, but, more importantly, the mental attitudes (beliefs, commitments, and execution status) the participating agents must have towards the action (see Figure 1b). In this sense, our notion of plan follows Pollack's (1990) mental-state view of collaborative plans. A plan represents the mental states of the agents on planning and performing an action, while a recipe represents the knowledge that an agent has about performing an action. To visually distinguish a recipe of an action from a plan of an action, we use slightly different graphical notations for them (cf. Figures 1a and 1b). In case of mediating human-GIS dialogues, two agents (the user and the computer) cooperatively act on α. Then, a plan may include the following components:

- *Intention*(Agents, α) are slots recording the intention of each agent towards action α (which can take a value of 'Intend-To', 'Intend-Not-To', or 'Unknown').
- *Recipe*(α) is a slot holding the recipe selected for the action. It can be empty, indicating that no recipe has been selected for the action. The system may know a number of different recipes for an action α, but only one of them is selected in a particular plan.
- *Beliefs*(Agents, α) are slots for recording what each agent believes about the performance of action α. Agents that participate in a plan on action α must establish beliefs about the ability of the agents to identify a recipe for α and to perform the action α following the recipe.
- *Commitment*(Agents, α) indicate whether the collaborating agents have committed to the success of the action. In many cases, the commitment of an agent to an action means that the agent has allocated resources (e.g. time and money) to perform its share of a collaborative action. For example, if Jim commits to have a lunch with Tom between 1 and 2 p.m., he cannot commit the doing anything else during that time. If conflicts happen, Jim has to re-plan his schedule by canceling or changing other meetings.

- *Exec_Status* (α) indicates the execution status of the plan. A plan can have a status of 'executable,' 'not executable,' 'not executed,' 'executed with success,' or 'executed with failure.'

When two or more agents collaboratively construct a plan, we call it a collaborative plan. As an example of a collaborative plan, consider the case where two persons (husband and wife) need to make a detailed plan for a family vacation. Both of them must *intend* to take a vacation. They must also have the shared *belief* that the two together can figure out all the details of the vacation plan and can carry it through successfully. Among many planning and preparation issues, they must be able to negotiate and agree upon 'where to go' and 'what to do'. Some details of the vacation can be planned either individually or collaboratively. The wife may be responsible for selecting travel modes (by car, train, or airplane) and routes by consulting various information sources (maps, travel agents, and weather reports). The husband may be responsible for preparing clothes and food, shopping, and getting children ready. The husband and wife may work together to decide what activities to do on the destination. During the process, they will keep each other informed and use each other as a source of help. Finally, they need to *commit* themselves to actually carrying out the vacation plan. Sometimes, a commitment can be complicated since an individual may have to re-plan other parts of his/her life in order to create conditions for this collaborative activity (vacation). This example, although intuitive, has all the essential components of collaborative interactions: recognizing and sharing each other's intentions, communicating knowledge to the details of the plan, negotiating agreements, and coordinating actions.

Now we introduce the concept of a plan graph, or PlanGraph. While a plan is a representation of collaboration status on a single goal, a PlanGraph is a schematic representation of all plans and subplans about a large, complex goal. A PlanGraph commonly has a complex goal at its root, which is decomposed recursively as subgoals (or subactions) through the adoption of recipes. For this reason, a PlanGraph is commonly a hierarchically organized set of plans. Figure 2 explains the general structure of PlanGraphs used in *GeoDialogue*. Nodes with oval shape indicate parameters, and nodes with rectangle shape represent subplans. A plan underneath a parameter node is the plan for identifying the parameter. For example, Plan γ_1 is for identifying parameter 'Para$_1$' of plan α.
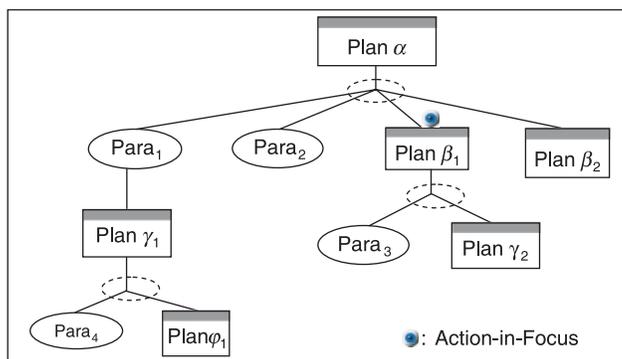

Figure 2 Structure of a PlanGraph

A conversational dialogue is modeled as the process of constructing a collaborative plan by the two participating agents. Before agents start to communicate, the discourse context is none, thus the PlanGraph is initially empty. As agents propose new initiatives during the dialogue, new plan nodes are introduced to the PlanGraph. The 'root plan' of the PlanGraph represents the most encompassing goal mentioned so far. If the action of the root plan is complex, agents will elaborate it on more details by selecting a recipe collaboratively. Then, agents move their attention to the parameters and subactions (as specified in the recipe). If the value of a parameter is unknown, a subplan is formed for identifying the parameter. If a subaction is not directly executable (i.e. a complex action), a subplan will be formed for performing this subaction. These subplans may themselves be complex, and will become the subjects of further elaboration.

A PlanGraph will become a *Full SharedPlan* (FSP) when: (1) participating agents have the shared beliefs that everyone intends and is committed to the whole plan; (2) all actions on the leaf-nodes are basic actions; and (3) for each of the parameters, either that it is already instantiated, or agents have a Full SharedPlan (FSP) for identifying the parameter. If the above conditions are not met, we say that the PlanGraph is only a *Partial SharedPlan* (PSP). A PSP represents an ongoing dialogue, while a FSP represents a complete dialogue. The progression of the dialogue corresponds to the process of evolving a collaborative plan from a PSP towards a FSP (Lochbaum 1998). *GeoDialogue* uses a PlanGraph to capture the discourse context of a dialogue, because it records information about the collaboration states underlying an ongoing dialogue.

Due to the limits of human attention and the linear nature of conversational dialogues, agents commonly talk about a complex task by focusing on one subgoal at one time, and by shifting the attention of the collaboration as one partial goal is accomplished and another partial goal is selected. In *GeoDialogue*, the attention state of a dialogue is represented by a 'cursor' (in the PlanGraph) pointing to the plan that is currently under the focus of the collaboration. We call the action under the cursor an *Action-in-Focus* (AiF). For example, 'Plan β_1' is the Action-in-Focus in the PlanGraph of Figure 2.

In summary, *GeoDialogue* models discourse as collaborative plans (or SharedPlan). The system response is driven by its intention to advance the SharedPlan, and, at the same time, to be responsive and helpful to users.

3.3 Modeling Activities of Geographical Information Dialogues

Human-GIS dialogues commonly happen within the context of an activity. Activities correspond to users' intention and knowledge in a problem domain. Complex activities can be broken down to some intermediate level tasks (called subtasks) and GIS operations. Timpf's (2003) work on geographical activity models is a significant step towards formally describing geographical information processing activities as a set of problem-solving methods, task ontologies, and their dependencies.

In *GeoDialogue*, we use the term 'actions' to refer to both tasks and operations as used by Timpf. We adopt a subset of action ontology developed by Albrecht (1997) and consider four types of actions (Figure 3). Type **I** actions deal with *spatial data retrieval* tasks such as 'retrieving a map layer' and 'selecting a subset of features from a layer'. Type **II** actions are about *analytical tasks* (spatial and/or statistical) such as 'making a buffer around certain features', 'finding spatial clusters', and 'areal aggregation'. Type **III** actions are *cartographic and visualization* tasks such as 'adding/removing layer', 'showing/hiding layer', 'zoom in/zoom out', 'pan', 'highlighting', and 'changing
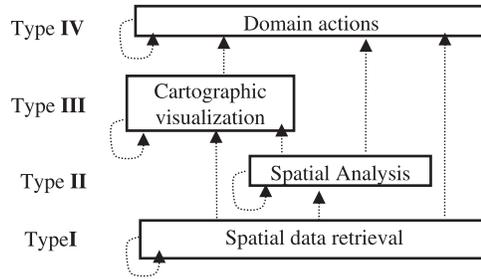
Figure 3   A typology of actions in *GeoDialogue*. Arrow means that one type of actions can be the subactions of another

cartographic symbols'. Type **IV** actions are *domain specific tasks* such as 'planning for evacuation' in a hurricane response domain.

   The ontology of actions in *GeoDialogue* also specifies how complex actions can be decomposed into subactions. This is accomplished by the definition of recipes in *GeoDialogue*'s recipe library. Given the role of *GeoDialogue* as a geographical information assistant to human users, *GeoDialogue*'s recipe definition allows the following patterns:

- A (complex) domain action may subordinate other domain actions, cartographic actions, spatial analysis actions, and/or spatial data retrieval actions.
- A (complex) cartographic action may subordinate other cartographic actions, spatial analysis actions, and/or spatial data retrieval actions.
- A (complex) spatial analysis action may subordinate other spatial analysis actions, and/or spatial data retrieval action.
- A (complex) spatial data retrieval action may subordinate other spatial data retrieval actions only.

These rules are represented in Figure 3 by arrows pointing from one type of actions to another.

   The above task model of geographical information activities serves as the basis for *GeoDialogue* to construct collaborative plans of an ongoing conversation. As a concrete example, Figure 4 shows the PlanGraph representation of a dialogue centered on a nuclear release event. Here the PlanGraph is rooted at a domain action "understanding impacted area". The "ShowMap" action (which is the main cartographic action) is a subordinate action contributing to the domain action at the root. The "Buffer analysis" action (which is a type II action) contributes to the "show map" action by adding a layer (which records the result of the buffer) to the map. Finally, "IdentifyNamedFeature" is a spatial data retrieval action (type I) contributing to the "buffer analysis" action.

3.4 Reasoning in Conversational Grounding and Generation

Each time the system detects a user's event in the form of a multimodal utterance, it will trigger a reasoning process for interpreting the user's message and subsequently generating responses. Following the theories of collaborative discourse (Lochbaum 1998) and conversational grounding (Clark and Brennan 1991, Brennan 1998), this reasoning
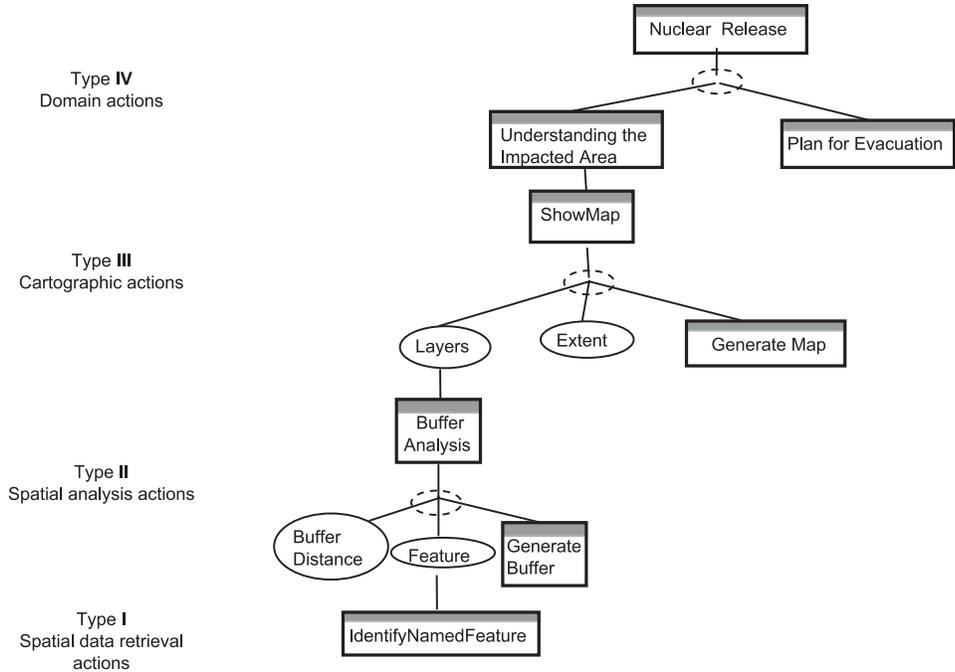
Figure 4   An Example of PlanGraphs

process drives both the interpretation of user's input and generation of response by two objectives: (1) maintaining common ground, and (2) advancing the joint activity. In doing so, *GeoDialogue* draws knowledge from two sources: (1) the PlanGraph representation of the discourse context, and (2) the activity model in the domain of geographical information dialogues.

Conversational grounding involves two steps of analysis: *semantic interpretation* and *explanation*. Semantic interpretation is the process of parsing a user's utterance into meaningful phrases and subsequently assigning a meaning to each phrase. Suppose that an utterance U is composed of a number of phrases $U = \{p_i, i = 0, 1, \ldots, n\}$. The contents of recognizable phrases are defined by a grammar, which specifies what combinations of words, gestures, or even other phrases are legitimate phrases. For each phrase p_i, the system knows one or more senses (or meanings) about it. At the end of this step, the system has a set of candidate meanings, $\varphi(p_i)$, for each phrase p_i. However, the system has not yet decided which meaning from $\varphi(p_i)$ is what the user intended. This is the purpose of the second step.

Explanation is the step where the system attempts to explain how the meanings of the new input relate to each other and to previous conversations. The purpose is to infer what the intended meaning of the user's utterance is. An explanation Ψ of utterance U exists under the current discourse context DC and a task model T if:

1. There exists $\Psi = \{\varphi_1, \varphi_2, \ldots, \varphi_i, \ldots \varphi_n\}$, where $\varphi_i =$ NULL or $\varphi_i \in \varphi(p_i)$;
2. Ψ can be "meaningfully merged" with DC with the task domain specified by T. The result is a new discourse context DC_{new}.

In this case, the system believes that the intended meaning of utterance U is Ψ. The term "meaningfully merged" has a specific meaning here. The intended meaning φ of phrase p is said to be meaningfully merged with the current discourse context DC if one of the following conditions is present: (1) φ provides a piece of information that helps the system to assign a value to a parameter in the current PlanGraph; (2) φ provides more details on the implementation of a complex action in the PlanGraph; or (3) φ helps establishing necessary mental states for a collaborative act.

If explanation is successful, *GeoDialogue* will merge the meaning of utterance U into the discourse context and update the PlanGraph accordingly. If some phrases of U can not be successfully explained, they will not be dropped immediately. Instead, un-explained phrases are pushed into a stack of "unexplained phrases", and will be further explained together with subsequent inputs. This mechanism is critical for *GeoDialogue* to deal with over-specified input which is a common phenomena in spoken dialogues (McTear 2002). An over-specified input includes information that is not immediately useful but the speaker anticipates the need of such information in subsequent steps.

The grounding process is followed by another reasoning stage, called *elaboration*. The main goal in this stage is to advance the human-GIS joint activity through the cooperative acts from the system side. The system can elaborate the current PlanGraph in several ways: (1) contributing a recipe to achieve the current Action-in-Focus; (2) identifying one or more parameters in the Action-in-Focus; or (3) executing any subplans that have been fully developed. This elaboration process also attempts to discover further obstacles for advancing the plan (e.g. conflicting beliefs, impossible goals, missing details, or ambiguous choices). A data structure called *Agenda* is used to collect all the detected 'obstacles' into action items. An action item includes a description of the nature of the problem and what needs to be done. This process allows the system to gather all the potential issues first, and then deal with them in the order of their priorities. The elaboration process ends when no more parts of the PlanGraph can be further elaborated without the intervention of the user. The elaboration process may also generate partial results (such as a map or an answer to a question).

For response generation, *GeoDialogue* make decisions on two things: (1) what the system intends to communicate back to the user; and (2) how. *First*, the system will derive its communicative intention by selecting an action item from the *Agenda*. When there are many items in the *Agenda*, the item with the highest priority will be chosen. *Second*, the system will decide on a strategy to communicate back to the user. There can be many strategies to accomplish the same communicative goal. For example, if a critical parameter is missing and the system intends to get it from the user, the system can either generate a verbal question to ask the user explicitly, or propose a best guess while expecting corrections from the user. In addition, a conceptual message can be communicated to the user using different modalities and media. For example, drawing a user's attention to an object on the screen can be accomplished by visual or iconic labeling, textual (spoken or screen text) labeling, gestures (animation and flashing), or using a combination of the above. Currently, the selection of communicative strategies and media in *GeoDialogue* is relatively naïve, where a small number of situation types are considered. This problem is being investigated separately.
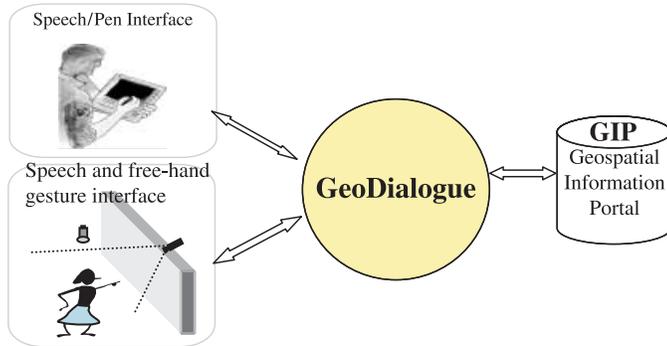
Figure 5   *GeoDialogue* as a plug-in component in a conversational Interface for GIS

3.5 Architecture and Implementation

GeoDialogue is designed as a plug-in software component that functions as a 'middle-agent' between a multimodal interface system (MMIS) and a geographical information portal (GIP) (Figure 5). The multimodal interface system can use any type of gesture input devices (pen-based or free-hand gestures), microphones and speech recognition engines. Communications between *GeoDialogue* and the multimodal interface system are through a simple message protocol, DAVE_GXML, that can be transmitted by common Internet message protocols (such as HTTP XML or SOAP(W3C 2000)). A DAVE_GXML message can be either a REQUEST or a RESPONSE. A REQUEST message is sent from the MMIS to *GeoDialogue*, and should include a speaker identifier, followed by a list of recognized elements (words, phrases and gestures). A RESPONSE message is always paired with a REQUEST message, and includes three types of elements: (1) the type of response (which can be 'map', 'confirmation', or 'question'); (2) the URL of a map image to be displayed on the user's screen; and (3) the textual message, if any. Some examples of such DAVE_GXML messages are given in Table 2.

The communication between *GeoDialogue* and the GIP is also accomplished through an XML message protocol that is mutually understood by both parties. To ensure interoperability of *GeoDialogue* with various geographical information portal

Table 2   Examples of DAVE_GXML messages

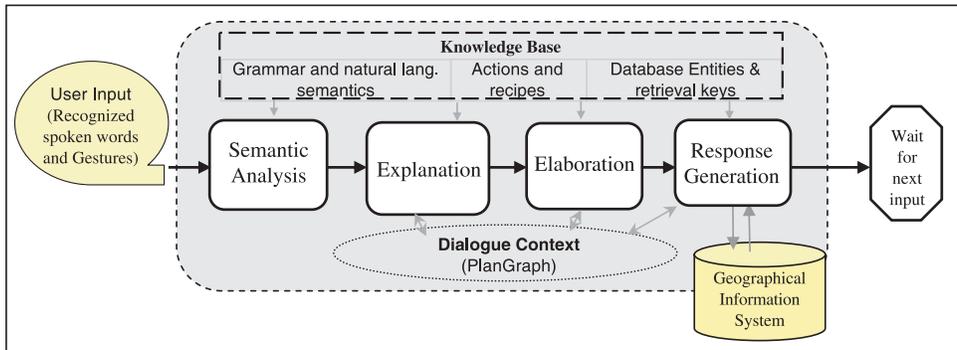| REQUEST: "Highlight this[gesture] HAZMAT facility"	RESPONSE: "Here is the map of Florida"

```<?xml version="1.0"? encoding="UTF-8"?>```
  ```<DAVE_GXML version="1.0.0">```
    ```<REQUEST speaker="Matt">```
      ```<UTTERANCE time="4/15/2003 14:30:15">```
        ```<PHRASE> <WORD>Highlight</WORD> </PHRASE>```
        ```<PHRASE> <WORD>This</WORD> </PHRASE>```
        ```<PHRASE> <WORD>HAZMAT</WORD>```
          ```<WORD>facility</WORD> </PHRASE>```
      ```<GESTURE type='pointing'> <POINT x=210 Y=75>```
      ```</GESTURE> </UTTERANCE> </REQUEST>```
  ```</DAVE_GXML>```
	```<?xml version="1.0"? encoding="UTF-8"?>```
	  ```<DAVE_GXML version="1.0.0">```
	    ```<RESPONSE type="map" >```
	      ```<MAP_IMAGE```
	```URL="http://spatial-srv1/output/Florida2102209.png"/>```
	      ```<TEXT>Here is the map of Florida.</TEXT>```
	    ```</RESPONSE>```
	  ```</DAVE_GXML>```

Figure 6 Architecture of *GeoDialogue*

technologies, this message protocol should be defined as close as possible to the existing standards such as the Web Mapping Service specification published by the Open Geospatial Consortium (OGC at www.opengeospatial.org). In our current implementation, the GIP is running on ESRI's ArcIMS, and its communication with *GeoDialogue* is through a subset of ArcXML language (ESRI 2002).

Internally, *GeoDialogue* is composed of several processing modules (Figure 6). These modules have a close correspondence to the reasoning process as described in section 3.4. *First*, the "*Semantic Analysis*" module iterates through the collection of phrases and searches for candidate meanings to each phrase. This stage of processing requires accessing grammatical and semantic knowledge available from the system's Knowledge Base. The '*Explanation*' module and the "*Elaboration*" module cause the change of the PlanGraph structure as new inputs are explained and new plan nodes are introduced. Both modules consult the knowledge-base when recipes for complex actions are needed. Finally, the "*Response Generation*" module needs to access knowledge about the database queries and mapping services offered by the GIP. The four process modules in *GeoDialogue* (as described above) form a complete processing flow necessary for mediating conversational dialogue between the user and the geographical information portal.

3.6 Functionalities

GeoDialogue, in its current implementation, can handle multimodal (speech/gesture) dialogues with GIS for basic geographical information retrieval and display tasks. Beyond responding to isolated requests related to general GIS commands (such as requests for displaying layers, selection of features on the map, generation of buffer zones, zooming the map and panning the map), *GeoDialogue* is unique in its capability to handle two types of extended dialogues when users interact with geographical information using natural modalities.

First, the system supports human-computer collaboration in constructing one or a series of dynamic maps. Making a map that meets the user's need is a complex task that requires careful planning of many details (what layers to be added, what map extent to be shown, what part of the map should be highlighted, what symbols should be used on each of the map features, etc.), all of which needs to be communicated to the system. *GeoDialogue* does not require the user to provide all details at one time and in a fixed

structure (as a formal command language interface does). Instead, *GeoDialogue* captures similar conceptual schemas as those used by humans to structure their spatial information request naturally in spoken language. These schemas, stored as recipes in our system, are the abstract knowledge structure of the tasks and their relationships described in Figure 3. At the time of interaction, the system follows the user's communicated intentions and activates proper schemas (recipes) to assimilate and organize user's inputs. In addition to attending to user's input, *GeoDialogue* also plays active roles ensuring that the overall goal of composing a map will be successful. First, the system is made responsible for raising issues to the user if the system believes that some user's intended goal is not achievable or parts of user's beliefs are not true or need to be revised. Second, the system actively communicates its belief about the user's request by using the map display as visual feedback. Since such maps afford user's guidance and corrections through gestures and speech input, the system may rely on the user's further reactions to provide clues for improving its model of the user's map request.

The *second* unique capability of *GeoDialogue* is that it is able to handle map requests that involve ambiguous or vague concepts. A concept is *ambiguous* to the system if it finds more than one match in the database or it is subject to several interpretations. For example, the concept "roads" may refer to highways, and/or local freeways, and/or local streets. A concept is *vague* to a computer if the concept corresponds to a fuzzy category or objects of broad boundaries. When ambiguity is detected, *GeoDialogue* will initiate clarification dialogues that involve the user in making the choice. A vague concept is hard to communicate among humans and is even harder to communicate between a human and a machine. As Cai et al. (2003) explained, there are potentially a large number of contexts that may influence the meaning of a vague concept in an unknown fashion. *GeoDialogue* is capable of initiating conversational dialogues when a vague spatial concept is detected in the user's input. This functionality is explained more fully in the hypothetical scenario given below.

3.7 An Application Scenario

We use a sample multimodal human-GIS dialogue to illustrate the functionalities of *GeoDialogue* in more detail. The scenario and the dialogue script are provided in Figure 7. The map responses generated during this dialogue are presented in Figure 8. Figure 9 shows the PlanGraph of this dialogue, where the labels (U1–U5) along the plan nodes indicate which user input step caused the addition of the plan node to the Plan-Graph. For example, a plan for the action 'find extent near a feature' was added to the PlanGraph by the user input step (U4).

The application context of this dialogue is emergency response to hurricane events. The map that the user is trying to make is the one labeled as G6 in Figure 8. However, in the beginning, the user has a very vague idea about what to include in the map. Thus, the user starts with an initial request (U1): 'show me a map of Florida'. This is an extremely underspecified request because it provides no indication about what themes to include in this map. The system recognizes this problem from reasoning on the ShowMap action of the PlanGraph, and decides to ask a question back to the user in order to get this critical piece of information. The strategy that the system uses to ask this question is to make a guess and ask the user to criticize. It generates a map response, G1 (in Figure 8), by putting in two basic contextual map layers: state boundaries and highways. The map is not meant to be precise, but merely to provide the user with a

G1 G2 G3 G4 G5 G6

Note: Each map is numbered by the dialogue step in Figure 7

G4 covers 28 miles buffer around Jacksonville, FL
G5 covers 4 miles buffer around Jacksonville, FL
G6 covers 8 miles buffer around Jacksonville, FL

Figure 8 Map responses generated by *GeoDialogue*.
These figures appears in colour in the electronic version of the article and in the plate section at the back of the printed journal

The dialogue begins with an empty map on the screen
U: The user;
G: GeoDialogue

U1: Show me a map of Florida.
G1: {*show a map with basic contextual layers*} What themes do you want to see?
U2: Show me major cities and hurricane surge areas.
G2: {*show a map with major cities*} There are three map layers related to "hurricane surge areas": category 1 surge zones; category 2 surge zones; and category 3 surge zones. Which one did you mean?
U3: Category 3.
G3: {*Show a map*} Here it is.
U4: Show me the area near this[gesture] city.
G4: {*Show a map with estimated map extent*} What are you going to do with this map?
U5: I am planning to evacuate people in the hurricane surge area.
G5: {*Show a new map*} Is this OK for you?
U6: No, I prefer to see more surrounding areas.
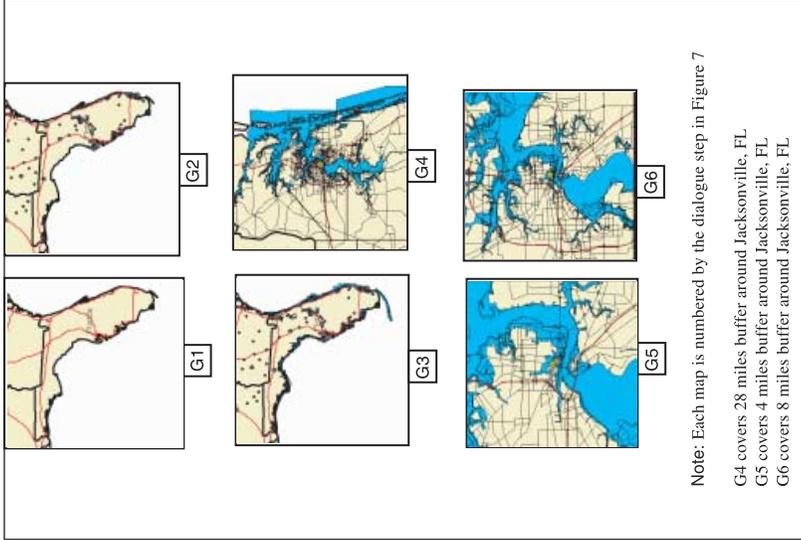G6: {*Show a new map*} Is this OK for you?
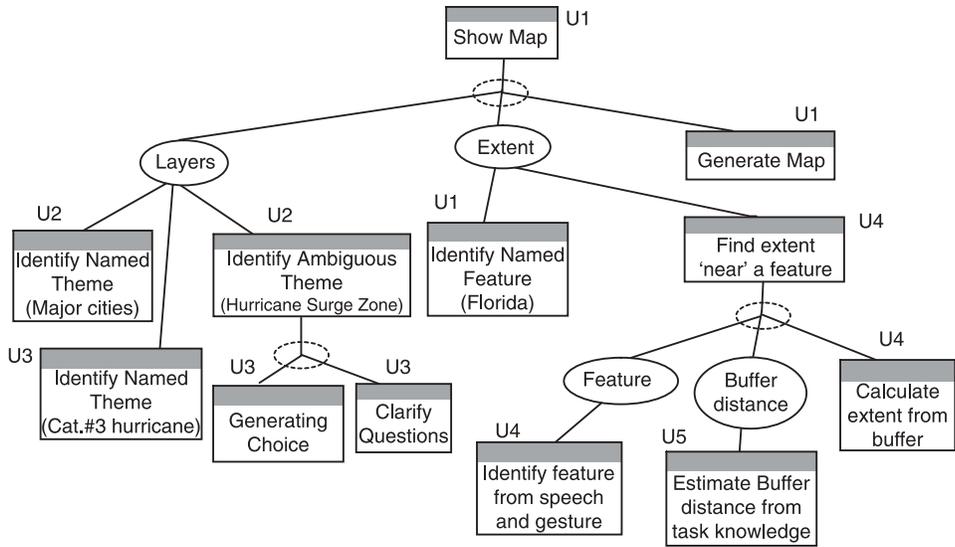U7: Yes, thanks.

Figure 7 A sample dialogue

Figure 9   Structure of a PlanGraph

visual aid for further improving the specification of the map request. A textual message, "what themes do you want to see?" is sent to the user along with the map, prompting the user's attention to this issue. In the spirit of collaboration, the user answers the question by requesting that "major cities" and "hurricane surge areas" be added to the map.

The dialogue segment (U2)–(G3) gives an example of how *GeoDialogue* handles ambiguous concepts. After receiving the request (U2), the system successfully finds one and only one data layer that matches exactly to the concept "major cities". However, the concept "hurricane surge areas" is ambiguous (to the system), since the system finds multiple matches according to its knowledge about the geographical database contents. The user does not realize this problem until the system shares the situation through (G2) and asks for clarification. Such behavior in mixed-initiative interactions makes *GeoDialogue* fundamentally different from traditional types of GIS interfaces.

The dialogue segment from U4 to G5 presents an example of how *GeoDialogue* handles vague concepts when they are part of a map request. The concept '*near*' in U4 is a well-known concept that corresponds to a vague spatial relation. The problem of communicating such a vague concept is that there are borderline cases that may or may not be covered by the meaning of such concepts depending on a variety of contextual factors (such as the task, the spatial scale, spatial arrangement of features, and user-specific preferences). One possible way of dealing with this situation is to generate a question "what do you mean by near?" and expect that the user will answer directly in a form as exact as "use 10 miles radius around this city". Unfortunately, the user may not be able to define the concept in that form, and humans are notoriously poor (and often reluctant) in giving numerical definitions for vague concepts. The approach taken by *GeoDialogue* in this case is to generate shared contextual knowledge between the user and the system, so that the system can adapt its understanding of vague concepts according to contexts. This strategy explains why the system poses a question (in G4)

about the user's task. By learning that the user was trying to plan for evacuation in response to a hurricane, the system was able to adapt its understanding of 'near' from the initial 28-mile radius around the city of Jacksonville (see G4 in Figure 8) to the 4-mile radius (as shown in G5 in Figure 8). More details on this function can be found in Cai et al. (2003).

This scenario illustrates the process through which *GeoDialogue* and the user both actively reason about each other's goals in relation to the overall collaboration. *GeoDialogue* implements a flexible control on dialogue initiatives. The user controls the higher-level dialogue initiatives in terms of what details to focus on, but the system is able to take some temporary initiatives that are consistent with the user's broader initiatives. For example, (G1) uses an explicit question to ask for a missing information piece, and (G3) clarifies a concept that has multiple mappings to database entities.

4 Discussion and Future Work

Conversational interfaces (Zue and Glass 2000, Allen et al. 2001, McTear 2002) have been accepted as a promising alternative to traditional graphical user interfaces, but have made little impact on how people access geographical information and use GIS. We made the first step of exploring the benefit of this technology by presenting a conceptual framework for developing natural conversational interfaces to GIS. This was motivated by the fact that technological progress in capturing human modalities for interacting with computers (and GIS in particular) has not been matched by advances in the semantic management of that interaction. Our computational model follows the human emulation approach to bring human and GIS into collaboration, focusing on the use of human modalities and conversational dialogues. The architecture of *GeoDialogue* mirrors the cognitive architecture for human perception and communication, where the SharedPlan model of dialogue context plays much the same role as the instantiated schemata in human information processing. The *GeoDialogue* prototype demonstrates the feasibility of a new paradigm for interacting with geospatial information. Such technologies, once robust and domain-relevant, have the potential to change fundamentally the way individuals, businesses and government agencies make use of geographical information. Users in these contexts will be able to interact with geographical information directly (instead of mediated by a technician) and naturally (through spoken language and gestures).

Towards intelligent dialogue management for more nature interaction with GIS, we have identified a number of challenging issues that require a sound computational framework if we are to make progress toward their solution. The approach taken by *GeoDialogue* is based on deep analysis of user tasks in natural geographical information seeking behavior, and is grounded in well-established theories in collaborative discourse and multi-agent planning. A difficulty faced in this task is the lack of data on or observations about how people interact with multimodal dialogue systems. We overcome this difficulty by applying a cognitive systems engineering approach to understand map use in real world contexts (Brewer and McNeese 2003). Throughout the development of the *GeoDialogue* prototype, application scenarios and dialogue scripts, generated from our field knowledge about current emergency operation environments, are used to guide the development of human-GIS dialogue behavior. In this sense, we have followed the scenario-based approach (Carroll 2000).

While our focus here has been on human-computer interaction, the interaction paradigm developed has important implications for addressing research challenges associated with both collaborative visualization and group spatial decision support. For both, research challenges recently delineated include making interfaces more natural and supporting shared concepts and collaborative construction of knowledge (MacEachren and Kraak 2001, Muntz et al. 2003). Both geovisualization and spatial-decision support focus on ill-structured problems. Support for group work in either activity will require tools that enable partial specification of a problem and an iterative process of working toward a solution. The collaborative dialogue approach, as outlined above for *GeoDialogue*, addressed these issues directly.

The approach we have taken is to model a dialogue as two agents negotiating meaning and sharing perspectives through a collaborative discourse. Future work will extend this model to provide system support for sharing ideas among multiple human collaborators engaged in the application of collaborative visualization tools to knowledge construction activities. Similarly, potential mechanisms for extending our *GeoDialogue* approach to multi-user environments are suggested by work in collaborative geovisualization on the semiotics of shared representation and the potential of map-based visual display to provide boundary objects through which to connect ontologies from different knowledge domains (MacEachren and Brewer 2004).

In order to assess the progress made by *GeoDialogue* towards creating a user-centered, usable, and robust geographical information environment, we are in the process of conducting a series of usability studies on the new interaction style enabled by *GeoDialogue*. Due to the unique challenge of evaluating agent-based dialogue systems, our usability studies apply several methods in combination. Three types of studies are planned: (1) expert guideline-based evaluations (user interface experts identify potential usability problems using heuristics); (2) formative usability assessments (domain users apply task-based scenarios, discuss and test the prototype); and (3) summative usability assessments (the prototype is tested against other comparable tools) (Nielsen 1993, Rubin 1994, Torres 2002). The qualitative and quantitative results of the usability assessments will support the refinement of *GeoDialogue* and provide a good basis for future developments of user-centered conversational interfaces.

Acknowledgements

References

Albrecht J 1997 Universal analytical GIS operations: A task-oriented systematization of data-structure-independent GIS functionality. In Craglia M and Onsrud H (eds) *Geographic Information Research: Transatlantic Perspectives*. London, Taylor and Francis: 577–91
Allen J, Byron D, Dzikovska M, Ferguson G, Galescu L, and Stent A 2001 Towards conversational human-computer interaction. *AI Magazine* 22(4): 27–37
Armstrong M P 2001 The four way intersection of geospatial information and information technology. In *Proceedings of the NRC/CSTB Workshop on Intersections between Geospatial Information and Information Technology*, Washington, D.C.

Armstrong M P and Densham P J 1995 A conceptual framework for improving human-computer interaction in locational decison-making. In Nyerges T L, Mark D M, Laurini R, and Egenhofer M J (eds) *Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems*. Dordrecht, Kluwer Academic Publishers: 343–54

Bolt R A 1980 Put-that-there: Voice and gesture at the graphics interface. *ACM Computer Graphics* 14: 262–70

Bratman M E 1992 Shared cooperative activity. *Philosophical Review* 101: 327–41

Brennan S E 1998 The grounding problem in conversation with and through computers. In Fussell S R and Kreuz R J (eds) *Social and Cognitive Psychological Approaches to Interpersonal Communication*. Hillsdale, NJ, Lawrence Erlbaum: 201–25

Brewer I and McNeese M D 2003 Using cognitive systems engineering to develop collaborative geospatial military technologies. In *Proceedings of the International Conference on Military Geology and Geography*, West Point, New York

Cai G, Wang H, and MacEachren A M 2003 Communicating vague spatial concepts in human-GIS interactions: A collaborative dialogue approach. In Kuhn W, Worboys M F, and Timpf S (eds) *Spatial Information Theory: Foundations of Geographic Information Science*. Berlin, Springer Lecture Notes in Computer Science No 2825: 287–300

Carroll J M 2000 Five reasons for scenario-based design. *Interacting with Computers* 13: 43–60

Clark H H and Brennan S E 1991 Grounding in communication. In Resnick L B, Levine R M, and Teasley S D (eds) *Perspectives on Socially Shared Cognition*. Washington, DC, APA: 127–49

Cohen P R, Johnston M, McGee D, Oviatt S, Pittman J, Smith I, Chen L, and Clow J 1997 QuickSet: Multimodal interaction for distributed applications. In *Proceedings of the Fifth ACM International Conference on Multimedia*, Seattle, Washington: 31–40

Copas C V and Edmonds E 2000 Intelligent interfaces through interactive planners. *Interacting with Computers* 12: 545–64

Egenhofer M J 1996 Multi-modal spatial querying. In Kraak J M and Molenaar M (eds) *Advances in GIS Research II: Proceedings of The Seventh International Symposium on Spatial Data Handling*. London, Taylor and Francis: 785–99

ESRI 2002 *ArcXML Programmer's Reference Guide – ArcIMS 4.0*. Redlands, CA, Environmental Systems Research Institute

Fischer G 2001 User modeling in human-computer interaction. *User Modeling and User-Adapted Interaction* 11: 1–18

Flanagan J L and Huang T S 2003 Scanning the special issue on human-computer multimodal interfaces. *Proceedings of the IEEE* 91: 1267–70

Florence J, Hornsby K, and Egenhofer M J 1996 The GIS wallboard: Interactions with spatial information on large-scale displays. In Kraak J M and Molenaar M (eds) *Advances in GIS Research II: Proceedings of the Seventh International Symposium on Spatial Data Handling*. London, Taylor and Francis: 449–63

Frank A U and Mark D M 1991 Language issues for GIS. In Maguire D J, Goodchild M F, and Rhind D W (eds) *Geographical Information Systems: Principles and Applications*. London, Longman: 147–63

Glass J, Flammia G, Goodine D, Phillips M, Polifroni J, Sakai S, Seneff S, and Zue V 1995 Multilingual spoken-language understanding in the MIT Voyager system. *Speech Communications* 17: 1–18

Gorin A L, Riccardi G, and Wright J H 1997 How may I help you? *Speech Communications* 23: 113–27

Grosz B J and Kraus S 1996 Collaborative plans for complex group action. *Artificial Intelligence* 86: 269–357

Grosz B J and Sidner C L 1986 Attention, intentions, and the structure of discourse. *Computational Linguistics* 12: 175–204

Grosz B J and Sidner C L 1990 Plans for discourse. In Cohen P R, Morgan J L, and Pollack M E (eds) *Intentions in Communication*. Cambridge, MA, MIT Press: 417–44

Hagen E 1999 An approach to mixed initiative spoken information retrieval dialogue. *User Modeling and User-Adapted Interaction* 9: 167–213

Hobbs J R 1979 Coherence and coreference. *Cognitive Science* 3: 67–89

Jennings N R, Sycara K, and Wooldridge M 1998 A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems* 1: 7–38

Jones R M, Copas C V, and Edmonds E A 1997 GIS support for distributed group-work in regional planning. *International Journal of Geographical Information Science* 11: 53–71

Juang B H and Furui S 2000 Automatic recognition and understanding of spoken language: A first step toward natural human-machine communication. *Proceedings of the IEEE* 88: 1142–65

Kettebekov S and Sharma R 2000 Understanding gestures in a multimodal human computer interaction. *International Journal of Artificial Intelligence Tools* 9: 205–24

Kettebekov S, Krahnstöver N, Leas M, Polat E, Raju H, Schapira E, and Sharma R 2000 i²Map: Crisis management using a multimodal interface. In *Proceedings of the ARL Federate Laboratory 4th Annual Symposium*, College Park, Maryland

Lochbaum K E 1994 Using Collaborative Plans to Model the Intentional Structure of Discourse. Unpublished Ph.D. Dissertation, Harvard University

Lochbaum K E 1998 A collaborative planning model of intentional structure. *Computational Linguistics* 24: 525–72

Lochbaum K E, Grosz B J, and Sidner C L 2000 Discourse structure and intention recognition. In Dale R, Moisl H, and Somers H (eds) *Handbook of Natural Language Processing*. New York, Marcel Dekker: 123–46

Lokuge I and Ishizaki S 1995 Geospace: An interactive visualization system for exploring complex information spaces. In *Proceedings of the International Conference in Human Factors in Computer Systems (CHI'95)*, New York: 409–14

MacEachren A M and Brewer I 2004 Developing a conceptual framework for visually-enabled geocollaboration. *International Journal of Geographical Information Science* 18: 1–34

MacEachren A M, Cai G, Sharma R, Brewer I and Rauschert I 2005 Enabling collaborative geoinformation access and decision-making through a natural, multimodal interface. *International Journal of Geographical Information Science* 19: 1–26

MacEachren A M and Kraak M-J 2001 Research challenges in geovisualization. *Cartography and Geographic Information Science* 28: 3–12

Mann W C and Thompson S A 1987 *Rhetorical Structure Theory: A Theory of Text Organization*. Los Angeles, CA, Information Science Institute (ISI) Report No. RR-87–90

Mark D and Frank A 1992 *User Interfaces for Geographic Information Systems*. Santa Barbara, CA, National Center for Geographic Information and Analysis Technical Report No. 92–3

McTear M F 2002 Spoken dialogue technology: Enabling the conversational user interface. *ACM Computing Surveys* 34: 90–169

Mondschein L G 1994 The role of spatial information systems in environmental emergency management. *Journal of the American Society for Information Science* 45: 678–85

Muntz R R, Barclay T, Dozier J, Faloutsos C, MacEachren A M, Martin J L, Pancake C M, and Satyanarayanan M 2003 *IT Roadmap to a Geospatial Future: Report of the Committee on Intersections Between Geospatial Information and Information Technology*. Washington, DC, National Academy of Sciences Press

Neal J G, Thielman C Y, Dobes Z, Haller S M, and Shapiro S C 1998 Natural language with integrated deictic and graphic gestures. In Maybury M T and Wahlster W (eds) *Readings in Intelligent User Interfaces*. San Francisco, CA, Morgan Kaufman: 38–51

Neal J G, Thielman C Y, Funke D J, and Byoun J S 1989 Multimodal output composition for human-computer dialogues. In *Proceedings of the IEEE AI Systems in Government Conference*, Washington, DC: 250–7

Nielsen J 1993 *Usability Engineering*. Boston, AP Professional

Nyerges T, Barndt M, and Brooks K 1997 Public participation geographic information systems. In *Proceedings of Auto-Carto 13*, Seattle, Washington: 224–33

O'Shaughnessy D 2003 Interacting with computers by voice: Automatic speech recognition and synthesis. *Proceedings of the IEEE* 91: 1272–305

Oviatt S 1996 Multimodal interfaces for dynamic interactive maps. In *Proceedings of the International Conference on Human Factors in Computing Systems (CHI'96)*: 95–102

Oviatt S and Cohen P 2000 Multimodal interfaces that process what comes naturally. *Communications of the ACM* 43(3): 45–53

Oviatt S, Cohen P, Wu L, Vergo J, Duncan L, Suhm B, Bers J, Holzman T, Winograd T, Landay J, Larson J, and Ferro D 2000 Designing the user interface for multimodal speech and pen-based gesture applications: State-of-the-art systems and future research directions. *Human-Computer Interaction* 15: 263–322

Pollack M E 1990 Plans as complex mental attitude. In Cohen P R and Morgan J L (eds) *Intentions in Communication*. Cambridge, MA, MIT Press: 77–103

Rauschert I, Agrawal P, Fuhrmann S, Brewer I, Sharma R, Cai G, and MacEachren A 2002 Designing a user-centered, multimodal GIS interface to support emergency management. In *Proceedings of the Tenth ACM International Symposium on Advances in Geographic Information Systems*, McLean, Virginia: 119–24

Rubin J 1994 *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. New York, John Wiley and Sons

Shapiro S C, Chalupski C H, and Chou H C 1991 *Linking Arc/INFO with SNACTor*. Santa Barbara, CA, National Center for Geographic Information and Analysis Technical Report No. 91–11

Sharma R, Pavlovic V I, and Huang T S 1998 Toward multimodal human-computer interface. *Proceedings of the IEEE* 86: 853–69

Sharma R, Poddar I, Ozyildiz E, Kettebekov S, Kim H, and Huang T S 1999 Toward interpretation of natural speech/gesture: Spatial planning on a virtual map. In *Proceedings of ARL Advanced Displays Annual Symposium*, Adelphi, Maryland: 35–9

Sharma R, Yeasin M, Krahnstoever N, Rauschert, Cai G, Brewer I, MacEachren A, and Sengupta K 2003 Speech-gesture driven multimodal interfaces for crisis management. *Proceedings of the IEEE* 91: 1327–54

Terveen L G 1995 Overview of human-computer collaboration. *Knowledge-Based Systems* 8: 67–81

Timpf S 2003 Geographic activity models. In Duckham M, Goodchild M, and Worboys M F (eds) *Foundations of Geographic Information Science*. London, Taylor and Francis: 241–54

Torres R J 2002 *Practitioner's Handbook for User Interface Design and Development*. Upper Saddle River, NJ, Prentice Hall

Traynor C and Williams M G 1995 Why are geographic information systems hard to use? In *Proceedings of the International Conference on Human Factors in Computing Systems (CHI'95)*, Denver, Colorado: 288–9

W3C 2000 Simple Object Access Protocol (SOAP) 1.1, World Wide Web Consortium

Wang F 2003 Handling grammatical errors, ambiguity and impreciseness in GIS natural language queries. *Transactions in GIS* 7: 103–21

Wilson A and Bobick A 1999 Parametric hidden Markov models for gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21: 884–900

Zerger A and Smith D I 2003 Impediments to using GIS for real-time disaster decision support. *Computers, Environment and Urban Systems* 27: 123–41

Zue V, Glass J, Goddeau D, Goodine D, Leung H, McCandless M, Phillips M, Polfroni J, Seneff S, and Whitney D 1990 Recent progress on the MIT voyager spoken language system. In *Proceedings of the ICSLP*: 1317–20

Zue V W and Glass J R 2000 Conversational interfaces: advances and challenges. *Proceedings of the IEEE* 88: 1166–80