Speech and Natural Language

Proceedings of a Workshop
Held at
Cape Cod, Massachusetts
October 15-18, 1989

Sponsored by:

Defense Advanced Research Projects Agency
Information Science and Technology Office

This document contains copies of reports prepared for the DARPA Speech and Natural Language Workshop. Included are reports from DARPA/ISTO sponsored programs and other materials prepared for use at the workshop.

Distributed by

Morgan Kaufmann, Publishers, Inc.

2929 Campus Drive

San Mateo, California 94403

ISBN 1-55860-112-0

Printed in the United States of America

Natural Language with Integrated Deictic and Graphic Gestures [1]

J.G. Neal[2], C.Y. Thielman[2], Z. Dobes[3]
S.M. Haller[3], S.C. Shapiro[3]
Calspan-UB Research Center (CUBRC)
P.O. Box 400, 4455 Genesee Street
Buffalo, NY 14225

ABSTRACT

People frequently and effectively integrate deictic and graphic gestures with their natural language (NL) when conducting human-to-human dialogue. Similar multi-modal communication can facilitate human interaction with modern sophisticated information processing and decision-aiding computer systems. As part of the CUBRICON project, we are developing NL processing technology that incorporates deictic and graphic gestures with simultaneous coordinated NL for both user inputs and system-generated outputs. Such multi-modal language should be natural and efficient for human-computer dialogue, particularly for presenting or requesting information about objects that are visible, or can be presented visibly, on a graphics display. This paper discusses unique interface capabilities that the CUBRICON system provides including the ability to: (1) accept and understand multi-media input such that references to entities in (spoken or typed) natural language sentences can include coordinated simultaneous pointing to the respective entities on a graphics display; use simultaneous pointing and NL references to disambiguate one another when appropriate; infer the intended referent of a point gesture which is inconsistent with the accompanying NL; (2) dynamically compose and generate multi-modal language that combines NL with deictic gestures and graphic expressions; synchronously present the spoken natural language and coordinated pointing gestures and graphic expressions; discriminate between spoken and written NL.

1 INTRODUCTION

One of the strong arguments in favor of using Natural Language (NL) processing systems as front-ends to sophisticated application systems is that if human-computer communication is conducted in an NL that most users know, then the cost of training a user to use the system

[1]This research was supported, in part, by the Defense Advanced Research Projects Agency and monitored by the Rome Air Development Center under Contract No. F30603-87-C-0136 and the National Science Foundation grant No. SES-88-10917 to The National Center for Geographic Information and Analysis

[2]Calspan Corporation

[3]State University of New York at Buffalo

should be greatly reduced. Human-computer communication can be made even more natural and effective for the user if deictic gestures and drawing expressions are incorporated into the language, since people very commonly and effectively augment their NL with deictic gestures, drawing, and other modes of communication when engaged in human-human dialogue. As part of the CUBRICON project, we are developing NL processing technology that incorporates deictic and graphic gestures with simultaneous coordinated NL for both user inputs and system-generated outputs.

The CUBRICON project [Neal88a, Neal88b, Neal89] is devoted to the development of knowledge-based interface technology that integrates speech input, speech output, natural language text, geographic maps, tables, graphics, and pointing gestures for interactive dialogues between human and computer. The objective is to provide both the user and system with modes of expression that can be combined and used in a natural and efficient manner, particularly when presenting or requesting information about objects that are visible, or can be presented visibly, on a graphics display. The goal of the project is to develop interface technology that uses its media/modalities intelligently in a flexible, highly integrated manner modelled after the manner in which humans converse in simultaneous coordinated multiple modalities.

The interface technology developed as part of this project has been implemented in the form of a prototype system, called CUBRICON (the CUBRC Intelligent CONversationalist). Although the application domain used to drive the research for the CUBRICON project is that of tactical Air Force mission planning, the interface technology incorporated in CUBRICON is applicable to domains with similar communication characteristics and requirements.

This paper discusses the research effort within the CUBRICON project that has focused on integrating NL with deictic and graphic gestures for user inputs and system-generated outputs. The unique interface capabilities that have been developed and implemented in the CUBRICON system include the ability to: (1) accept and understand multi-media input such that references to entities in (spoken or typed) natural language sentences can include coordinated simultaneous pointing to the respective entities on a graphics display; use simultaneous pointing and NL references to disambiguate one another when appropriate; infer the intended referent of a point gesture which is inconsistent with the accompanying NL; (2) dynamically compose and generate multi-modal language that combines NL with deictic gestures and graphic expressions; synchronously present the spoken natural language and coordinated pointing gestures and graphic expressions; discriminate between spoken and written NL.

2 SYSTEM OVERVIEW

The CUBRICON design provides for the use of a unified multi-media language, by both the user and system, for communication in a dialogue setting. Input and output streams

are treated as compound streams with components corresponding to different media. This approach is intended to imitate, to a certain extent, the ability of humans to simultaneously accept input from different sensory devices (such as eyes and ears), and to simultaneously produce output in different media (such as voice, pointing motions, and drawings).

An overview of the CUBRICON software system and hardware I/O devices is presented in Figure 1. CUBRICON accepts input from three input devices: speech input device, keyboard, and mouse. CUBRICON produces output for three output devices: high-resolution color-graphics display, high-resolution monochrome display, and speech production device. The primary path that the input data follows is indicated by the modules that are numbered in the figure: (1) Input Coordinator, (2) Multi-Media Parser Interpreter, (3) Executor/Communicator to Target System, (4) Multi-Media Output Planner, and (5) the Coordinated Output Generator. The Input Coordinator module accepts input from the three input devices and fuses the input streams into a single compound stream, maintaining the temporal order of tokens in the original streams. The Multi-Media Parser/Interpreter is a generalized augmented transition network (GATN) that has been extended to accept the compound stream produced by the Input Coordinator and produce an interpretation of this compound stream. Appropriate action is then taken by the Executor module. This action may be a command to the mission planning system, a database query, or an action that entails participation of the interface system only. An expression of the results of the action is then planned by the Multi-Media Output Planner for communication to the user. The Output Planner uses a GATN that produces a multi-media output stream representation with components targeted for the different output devices. This output representation is translated into visual/auditory output by the Output Generator module. This module is responsible for producing the multi-media output in a coordinated manner in real time (e.g., the Planner module can specify that a certain icon on the color-graphics display must be highlighted when the entity represented by the icon is mentioned in the simultaneous natural language output).

The CUBRICON system includes several knowledge sources that are used for both understanding input and composing output. The knowledge sources include: a lexicon, a grammar defining the multi-modal language used by the system for input and output, a discourse model, a user model, and a knowledge base of task domain and interface information. The latter knowledge sources are discussed briefly in the following paragraphs.

The *knowledge base* consists of information about the task domain of tactical Air Force mission planning. This knowledge base includes information about concepts such as SAMs, air bases, radars, and missions as well as related HCI concepts such as verbal/graphical expressions for the domain concepts.

The *discourse model* is a representation of the attentional focus space [Grosz86] of the dialogue carried out in multi-modal language. It consists of (1) a main focus list that includes those entities and propositions that have been explicitly expressed (by the user or by CUBRI-

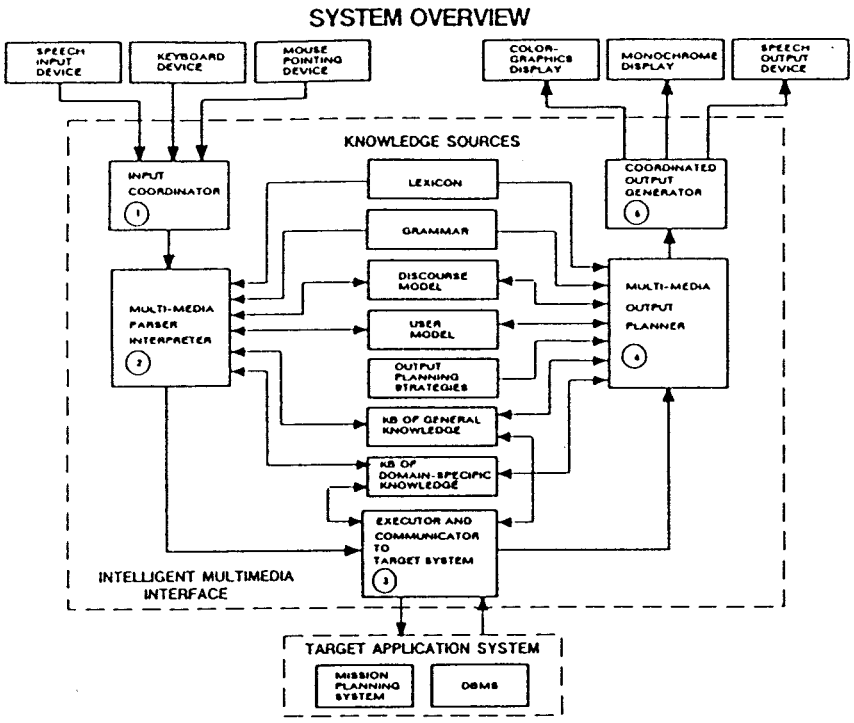


Figure 1: System Overview

CON) via natural language and/or graphic/pointing gestures and (2) a display model that includes a representation of all the objects (windows and their contents) that are "in focus" because they are visible on one of the two CRT screens.

The *user model* [Kobsa88] consists of an *entity rating module* that includes a task-dependent representation of the relative importance of all the entity types known to the system and an algorithm for modifying these ratings depending on task and dialogue activity.

Key features of the CUBRICON design, discussed in this paper, include the integration of NL and graphics in a unified language that is defined by a multi-modal grammar and the generation of synchronized speech and graphics in real time. The integration of NL and graphics in a unified language distinguishes this research from other approaches to multi-modal interface technology [Sullivan88, Arens89]. The Integrated Interface system [Arens88] and the XTRA system [Kobsa86, Allgayer89] are two of the most relevant. The Integrated Interface system is a multi-modal system in that it uses both maps and NL for the presentation of information to the user. The system provides information about the status and movements of naval platforms and groups in the Pacific Ocean. The system displays NL in text boxes positioned on a map display near the relevant objects. The system does not use a multi-modal language, however. The language generated is purely NL with no integrated graphics. The XTRA system is a multi-modal interface system which accepts and generates NL with accompanying point gestures for input and output, respectively. In contrast to the XTRA system, however, CUBRICON supports a greater number of different

types of pointing gestures and does not restrict the user to pointing at form slots alone, but enables the user to point at a variety of objects such as windows, table entries, icons on maps, and geometric points. In added contrast to XTRA, CUBRICON provides for multiple point gestures per NL phrase and multiple point-accompanied phrases per sentence during both user input and system-generated output. CUBRICON also includes graphic gestures (i.e., certain types of simple drawing) as part of its multi-modal language, in addition to pointing gestures. Furthermore, CUBRICON addresses the problem of coordinating NL (speech) and graphic gestures during both input and output.

CUBRICON software is implemented on a Symbolics Lisp Machine using the SNePS semantic network processing system [Shapiro79, Shapiro87], a GATN parser-generator [Shapiro82], and Common Lisp. Speech recognition is handled by a Dragon Systems VoiceScribe 1000. Speech output is produced by a DECtalk speech production system.

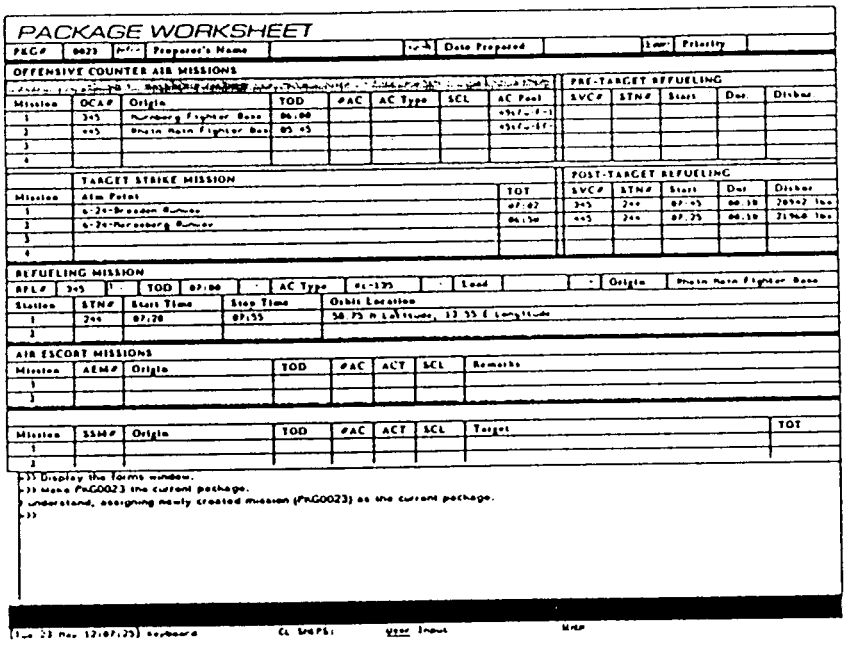
As stated previously, CUBRICON is a multi-modal system that integrates the following modalities: geographic maps, tables, forms, printed text, and NL with graphic and deictic gestures. Subsequent sections of this paper present example sentences that include simultaneous coordinated pointing gestures to objects on the graphics displays. Figure 2 shows example CUBRICON displays containing a form, geographic map, table, part-whole decomposition.

The following sections discuss CUBRICON's input understanding and output composition processes and their use of the knowledge sources discussed above.

3 MULTI-MODAL LANGUAGE UNDERSTANDING

People commonly and naturally use coordinated simultaneous natural language and graphic gestures when working at graphic displays. These modes of communication combine synergistically to form an efficient language for expressing definite references and locative adverbials. One of the benefits of this multi-modal language is that it eliminates the need for the lengthy definite descriptions that would be necessary for unnamed objects if only natural language were used. Instead, a terse reference such as "this SAM" (surface-to-air missile system) accompanied by a point to an entity on the display can be used. CUBRICON accepts such NL accompanied by simultaneous coordinated pointing gestures. The NL can be input via the keyboard, the speech recognition system, or a mixture of both. CUBRICON provides

- variety in the object types that can be targets of point gestures; these object types include windows, form slots, table entries, icons, and geometric points;
- variety in the number of point gestures allowed per phrase; each noun phrase can be accompanied by zero or more point gestures; such a phrase may contain no words, just the pointing gestures;



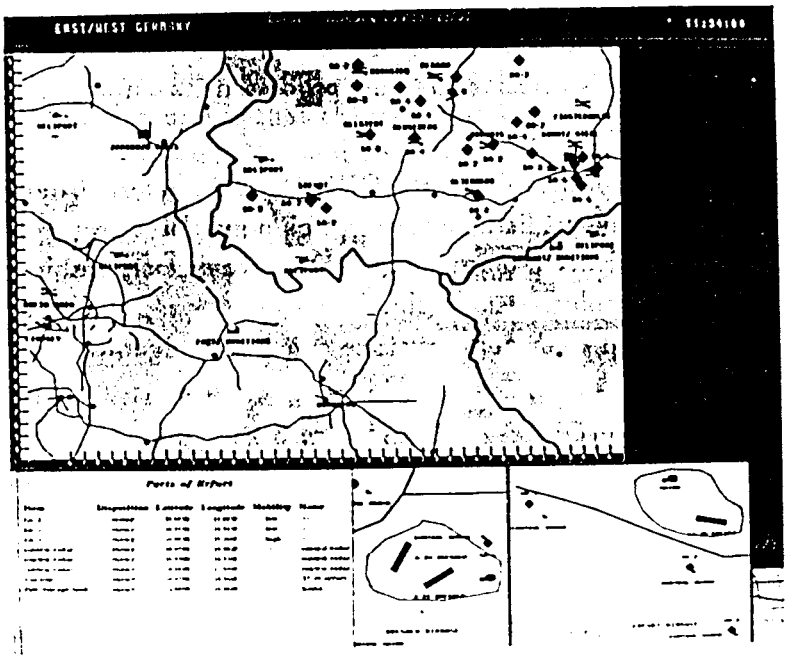


Figure 2: Example CUBRICON Displays

- variety in the number of multi-modal phrases allowed per sentence; deictic gestures can accompany more than one phrase per sentence.

Just as natural language used alone has shortcomings, so also does the use of pointing gestures alone. Pointing used alone has the following problems: (1) a point gesture can be ambiguous if the point touches the area where two or more graphical figures or icons overlap or (2) the user may inadvertently miss the object at which he intended to point. To handle these pointing problems, some systems use default techniques such as having a point handler return the entity represented by (a) the "top" or "foremost" icon where the system has a data structure it uses to remember the order in which icons are "painted" on the display (i.e., which are further in the background and which are foremost in the foreground) or (b) the icon whose "center" is closest to the location on the screen/window touched by the point. A serious disadvantage of such default point-interpretation techniques is that it is difficult, if not impossible, for certain icons to be selected via a point reference.

CUBRICON's acceptance of dual-media input (NL accompanied by coordinated pointing gestures) overcomes the limitations of the above weak default techniques and provides an efficient expressive referencing capability. The CUBRICON methodology for handling dual-media input is a decision-making process that depends on a variety of factors such as the *types* of candidate objects being referenced, their *properties*, the *sentential context*, and the *constraints on the participants* or fillers of the semantic case frame for the verb of any given sentence. CUBRICON's decision-making process draws upon it's knowledge sources discussed briefly in Section 2.

We present a few brief examples to illustrate CUBRICON's referent determination process. This process handles the problems listed above: ambiguous point gestures and point gestures that are inconsistent with the accompanying NL. First we discuss ambiguous point gestures. In each of the following examples, assume that the <point> represents a point gesture with a device such as a mouse and each point gesture can be ambiguous (i.e., it can touch more than one icon).

Example 1: USER: "What is the status of this <point> airbase?"

From the icons touched by the point, the display model is searched for the semantic representation of the objects which were graphically represented by the touched icons. From the hierarchy of the knowledge base, the system determines which of the objects selected by the point gesture are of the type mentioned in the accompanying verbal phrase ("airbase" in the example sentence) and discards the others.

Example 2: USER: "What is the mobility of these $<\text{point}>_1$ $<\text{point}>_2$ $<\text{point}>_3$?"

Example 2 illustrates that CUBRICON enables the user to use more than one point gesture per phrase. Also, in contrast to Example 1, no object type is mentioned in the noun phrase corresponding to the point gestures. In this case, CUBRICON can use a mentioned

property (e.g., mobility) to select from among the candidate referents of the point gesture. CUBRICON accesses the display model to retrieve the semantic representations of the objects touched by each of the user's point gestures, and then determines which of these objects have property "mobility" using the knowledge base of application information.

Example 3: USER: "Enter this <point-map-icon> here <point-form-slot>."

Example 3 illustrates that CUBRICON enables the user to use point gestures in conjunction with more than just one phrase of a sentence and that the point gestures may access different types of windows, even on different CRTs. In Example 3, the user's first point gesture touches an object on a map display on the color-graphics CRT and the second selects a slot of the mission planning form on the monochrome CRT. Two of CUBRICON's features are critical to its ability to process the sentence of Example 3: First, the display model contains semantic representations of all the objects displayed visually in each of the windows of each CRT, and second, all objects and concepts in the CUBRICON system are represented in a single knowledge representation language, namely the formalism of the SNePS knowledge base. This knowledge base is shared by all the modules of the CUBRICON system. Suppose that the <point-map-icon> selects the Nuernberg airbase on the map and the <point-form-slot> touches the "origin airbase" slot on the mission planning form. CUBRICON's response to the input of Example 3 would be to build the knowledge base structure which represents the assertion that Nuernberg is the airbase from which the particular mission will be flown.

As mentioned previously in this section, in addition to being ambiguous, another problem that can arise with point gestures is that the user may inadvertently miss the object at which he intended to point. In this case, the point gesture will be inconsistent with the accompanying natural language phrase, meaning that the natural language part of the expression and the accompanying point cannot be interpreted as referring to the same object(s) (e.g., the user says "this airbase" and points to a factory or points at nothing, missing all the icons). CUBRICON includes methodology to infer the intended referent in this case. CUBRICON uses the information from the sentence, parsed and interpreted thus far, as filtering criteria for candidate objects. The system performs a bounded incremental search around the location of the user's point to find the closest object(s) that satisfy the filtering criteria. If one is found, then the system responds to the user's input (e.g., command or request) and also issues an advisory statement concerning the inconsistency. In the event that no qualified object is found in the vicinity of the user's point, then a response is made to the user to this effect.

4 MULTI-MODAL LANGUAGE GENERATION

Just as CUBRICON accepts NL accompanied by deictic and graphic gestures during input, CUBRICON can generate multi-modal language output that combines NL with *deictic gestures* and *graphic expressions*. An important feature of the CUBRICON design is that NL

and graphics are incorporated in a single language generator providing a unified multi-modal language with speech and graphics synchronized in real time.

Another important aspect of the CUBRICON system is that it distinguishes between spoken and written (to a CRT display) NL. CUBRICON uses graphic and deictic gestures with *spoken NL* only (not with written NL), since a pointing or graphic gesture needs to be temporally synchronized with the corresponding verbal phrase, allowing for multiple graphic gestures within any individual sentence. The coordination between a graphic gesture and its co-referring verbal phrase is lost if printed text is used instead of speech. As mentioned in Section 3, a pointing gesture can be used very effectively with a terse NL phrase (e.g., "this SAM") to reference an object that is visible on one of the displays (by the system as well as the user). When CUBRICON generates *written NL*, however, deictic/graphic expressions are not used, but, instead, definite descriptions are generated as noun phrases with sufficient specificity to hopefully avoid ambiguous references. CUBRICON's use of deictic gestures and graphic expressions are discussed in the following paragraphs.

Deictic gestures are combined with appropriate NL during output to guide the user's visual focus of attention. During language generation, in order to compose a reference for an object,

1. if the object is represented by an icon on the display, then CUBRICON generates a NL expression for the object and a simultaneous coordinated graphic gesture that points to its icon.

   If the object has an individual name or identifier, then CUBRICON uses its name or identifier (e.g., "the Merseberg airbase") as the NL expression

   else CUBRICON generates an expression consisting of a demonstrative pronoun followed by the name of an appropriate class to which the object belongs (e.g., "this SAM", "these SAMs") as the NL expression.

2. if the object (call it X) is not represented by an icon on the display, but is a component of such a visible object (call it Y), then CUBRICON generates a phrase that expresses object X as a component of object Y and uses a combined deictic-verbal expression for object Y as described in the above case. For example, if CUBRICON is generating a reference for the runway of an airbase called Merseberg and an icon for the airbase is visible on the map (the airbase as a whole is represented visibly, but not its parts), then CUBRICON generates the phrase "the runway of the Merseberg Airbase" with a simultaneous point gesture that is directed at the Merseberg airbase icon on the map.

It is frequently the case that an object to which CUBRICON wants to point has a visible representation in more than one window on the CRTs. Therefore the system must select the visual representation(s) of the object (e.g., an icon, table entry, form slot entry) that it will use in its point gesture(s) from among the several candidates. The current CUBRICON

methodology is to point out all the object's visible representations, but to use a strong pointing gesture (e.g., blink the icon to attract the user's attention and add a pointing text-box) for the most significant or relevant representations and weak non-distracting gestures (e.g., just highlight the visible representation) for the less significant ones. In order to select the most relevant visible representations from among all the candidates, CUBRICON:

1. selects all the windows which contain a visible representation of the object.
2. filters out any windows which are not active or not exposed.
3. if there are exposed windows containing a visible representation of the object, then CUBRICON uses all of these representations as objects of weak diectic gestures and selects the visible representation in the most important or salient window [Neal89b] as the target of a strong diectic gesture.
4. if there are no exposed windows displaying the object's visible representation, then CUBRICON determines the most important active de-exposed window [Neal89b] displaying the object. CUBRICON exposes this window and uses the representation of the object in this window in a strong deictic gesture.

CUBRICON combines *graphic expressions* with NL output when the information to be expressed is, at least partially, amenable to graphic presentation. In the current CUBRICON implementation, the type of information that falls in this category includes (1) locative information and (2) path traversal information. We discuss only the locative case in this paper.

When generating locative information about some object (call it the figure object [Herskovits85]), CUBRICON selects an appropriate landmark as the ground object [Herskovits85], determines a spatial relationship between the figure and ground object, and generates a multi-modal expression for the locative information including the spatial relationship. When selecting the ground object, CUBRICON selects a landmark such as a city, border, or region, that is within the current map display (i.e., does not require a map transformation). If possible, CUBRICON uses a landmark that is in focus by virtue of its having been already used recently as a ground object. CUBRICON's discourse model, discussed briefly in Section 2, includes a representation of the attentional focus space of the dialogue, including a main focus list of entities and propositions that have been expressed by CUBRICON or by the user via multi-modal language. If a new landmark must be used as a ground object, then CUBRICON selects the landmark that is nearest the figure object. CUBRICON derives a spatial relation between the ground object and figure object that it represents in its knowledge base. This relation includes (1) the direction from the ground object to the figure object and (2) the distance if the distance is greater than 0.04 of the window width. If the distance is less than 0.04 of the window width, then the figure object appears to be

right next to the ground object. This criterion for deciding whether to include distance as part of the relation reflects the tendency for people to omit a distance measure when the distance is small relative to the geographic area under discussion and to say something like "just northeast of" instead of stating a distance explicitly.

As an illustrative example, the user may ask about the location of a particular object, such as the Fritz Steel plant. The system then uses the steel plant as the figure object, selects a ground object, and derives a spatial relation between ground object and figure object as discussed above. The multi-modal response is given below.

USER: "Where is the Fritz Steel plant?"

CUBRICON: "The Fritz Steel plant is located here <point>, 45 miles southwest of Dresden <graphic-expression>."

The <point> consists of a gesture that points out the Fritz Steel plant icon to the user via a gesture that uses a combination of blinking, highlighting, circling the icon and the attachment of a pointing label-box that identifies the icon. The <graphic-expression> is a visual presentation of the spatial relation between the figure object (Fritz steel plant) and the ground object (Dresden city), consisting of an arrow drawn from the Dresden city icon to the steel plant icon, a label stating the distance, and a label identifying the city (the steel plant should already be labeled).

CUBRICON's multi-modal language generation is also discussed in [Neal89].

5 FUTURE DIRECTIONS

There are numerous worthwhile areas and ideas to be investigated and developed to advance this research. We briefly discuss two of these areas:

CUBRICON is currently being extended so that it accepts a larger vocabulary of graphic drawing gestures as part of the user's multi-modal input. An integrated language consisting of both verbal and graphic "tokens" can be used for both referencing objects that the system already knows about as well as explaining and defining new concepts to the system. Such a multi-modal input language should be especially useful for the definition and explanation of geographical and spatial concepts to a system that would then use the concepts for geographical applications. We are currently focusing on adding polylines to the set of graphic gestures that CUBRICON accepts. Polylines can be used to approximate free-hand drawing and thereby give the user great expressive power.

We are also planning to conduct a research program to investigate the problem of user gestures that are not synchronized with their corresponding NL phrases. We are interested in the characteristics of the phenomenon: to what degree are gestures of different types not synchronized with their corresponding NL phrase, how frequently does the phenomenon

occur, is there a correlation between characteristics of the phenomenon and characteristics of the corresponding natural language? We also plan to investigate methods that would enable the system to decide which phrase of the accompanying natural language input is the co-referring phrase for any pointing gesture that is not synchronized with its co-referring phrase.

6 SUMMARY

People frequently augment their NL with deictic gestures, drawing, and other modes of communication when engaged in human-human dialogue. The CUBRICON project is devoted to the development of knowledge-based interface technology that integrates speech input, speech output, natural language text, geographic maps, tables, graphics, and pointing gestures for interactive communication between human and computer. The objective is to provide both the user and system with modes of expression that are combined and used in a natural and efficient manner, particularly when presenting or requesting information about objects that are visible, or can be presented visibly, on a graphics display.

As part of the CUBRICON project, we are developing NL processing technology that integrates deictic and graphic gestures with simultaneous coordinated NL to form a multi-modal language for human-computer dialogues. CUBRICON's main I/O processing modules have access to several knowledge sources or data structures, including one modeling each of (1) the application domain, (2) the discourse, and (3) the user.

This paper discussed the unique interface capabilities that the CUBRICON system provides including the ability to: (1) accept and understand multi-media input such that references to entities in (spoken or typed) natural language sentences can include coordinated simultaneous pointing to the respective entities on a graphics display; use simultaneous pointing and NL references to disambiguate one another when appropriate; infer the intended referent of a point gesture which is inconsistent with the accompanying NL; (2) dynamically compose and generate multi-modal language that combines NL with deictic gestures and graphic expressions; synchronously present the spoken natural language and coordinated pointing gestures and graphic expressions; discriminate between spoken and written NL.

7 REFERENCES

[Allgayer89] Allgayer, J., Jansen-Winkeln, R., Reddig, C., & Reithinger, N. 1989. Bidirectional Use of Knowledge in the Multi-Modal NL Access System XTRA. *Proc. of IJCAI-89*, Detroit, MI, pp. 1492-1497.

[Arens88] Arens, Y., Miller, L., & Sondheimer, N.K. 1988. Presentation Planning Using an Integrated Knowledge Base, in *Architectures for Intelligent Interfaces: Elements and Prototypes*, J.W. Sullivan & S.W. Tyler (eds.), Addison-Wesley, pp. 93-108.

[Arens89] Arens, Y., Feiner, S., Hollan, J., & Neches, R. (eds.) 1989. *A New Generation of Intelligent Interfaces, IJCAI-89 Workshop*, Detroit, MI.

[Grosz86] Grosz, B.J. 1986. The Representation and Use of Focus in a System for Understanding Dialogs, in *Readings in Natural Language Processing*, B.J. Grosz, K.S. Jones, & B.L. Webber (eds.), Morgan Kaufmann Pub., pp. 353-362.

[Haller89] Haller, S.M. 1989. Technical Report: Spatial Relations and Locative Phrase Generation in a Map Context. Computer Science Department, State University of New York at Buffalo.

[Herskovits85] Herskovits, A. 1985. Semantics and Pragmatics of Locative Expressions. *Cognitive Science*, 9:341-378.

[Kobsa86] Kobsa, A., Allgayer, J., Reddig, C., Reithinger, N., Schmauks, D., Harbusch, K., & Wahlster, W. 1986. Combining Deictic Gestures and Natural Language for Referent Identification, *Proc. of the 11th International Conf. on Computational Linguistics*, Bonn, FR Germany.

[Kobsa88] Kobsa, A. & Wahlster, W. (eds.), 1988. *Computational Linguistics*, Special Issue on User Modeling, MIT Press.

[Neal88a] Neal, J.G. & Shapiro, S.C. 1988. Intelligent Multi-Media Interface Technology, in *Architectures for Intelligent Interfaces: Elements and Prototypes*, J.W. Sullivan & S.W. Tyler (eds.), Addison-Wesley, pp. 69-91.

[Neal88b] Neal, J.G., Dobes, Z., Bettinger, K.E., & Byoun, J.S. 1988. Multi-Modal References in Human-Computer Dialogue, *Proc. AAAI-88*, St. Paul, MN, pp. 819-823.

[Neal89a] Neal, J.G., Thielman, C.Y., Funke, D.J., & Byoun, J.S. 1989. Multi-Modal Output Composition for Human-Computer Dialogues. *Proc. of the AI Systems in Government Conference*, George Washington Univ., Wash. D.C., pp. 250-257.

[Neal89b] Neal, J.G. et. al. 1989. The CUBRICON Multi-Modal Interface System. (Journal paper in preparation).

[Shapiro79] Shapiro, S.C. 1979. The SNePS Semantic Network Processing System, in *Associative Networks - The Representation and Use of Knowledge by Computers*, N. Findler (ed.), Academic Press, pp. 179-203.

[Shapiro82] Shapiro, S.C. 1982. Generalized Augmented Transition Network Grammars for Generation from Semantic Networks. *AJCL*, Vol. 8, No. 1, pp. 12-25.

[Shapiro87] Shapiro, S.C. & Rapaport, W.J. 1987. SNePS Considered as a Fully Intensional Propositional Semantic Network, in *The Knowledge Frontier, Essays in the Representation of Knowledge*, N. Cercone & G. McCalla (eds.), Springer-Verlag, pp. 263-315.

[Sullivan88] Sullivan, J.W. & Sherman, W.T. (eds.) 1988. *Architectures for Intelligent Interfaces: Elements and Prototypes*, Addison-Wesley Pub. Co.