Structuring Two-Medium Dialog for Learning Language and Other Things

Henry Hamburger
George Mason University; Fairfax, VA, USA

Dan Tufis
Research Institute for Informatics; Bucharest, Romania

Raza Hashim
Bridgewater College; Bridgewater, VA, USA

OVERVIEW: Naturalistic two-medium communication with a computational system, using both language and interactive graphics (Cohen, 1991; Maier, 1993; McKeown, 1993), is an important practical complement to studies that involve only language, only graphics and/or only people. Integrative two-medium work should build on insights and findings in the one-medium disciplines of graphical manipulative interfaces (e.g., Hutchins et al., 1986; Sullivan and Tyler, 1991) and natural language discourse (e.g., Grosz and Sidner, 1986; Litman and Allen, 1987; Hovy, 1988; Lambert and Carberry, 1991; Paris, 1991).

In addition to its general use with a variety of systems, two-medium communication provides an essential foundation for a pedagogically important form of foreign language learning experience (Hamburger and Hashim, 1992). Specifically, it permits a supportive dialog practice system for naturalistic acquisition of various language aspects, by combining discourse constraints with independently comprehensible situational contexts. By factoring out and gaining control of the types of dialog interactions, the various situational and object viewpoints, the choice of domain and the intradomain events, we are attempting to provide the tools for a workable tutorial strategy, one that will resolve the competing requirements of conversational continuity and appropriate language challenge (Hamburger and Maney, 1991). Finally, to indicate the potentially broader applicability of this approach, we sketch the implementation of two-medium communication for learning environments with non-language subject matter (Hashim and Hamburger, 1993).

Our second-generation conversational foreign language tutor, FLUENT-2, takes a generative (recombinative) approach to dialog management. We thereby dramatically enhance the linguistic range of the tutor over the earlier, less flexible FLUENT-1, yet continue to impose conversational continuity on the tightly integrated linguistic and spatial communication between student and tutor. The keys to this approach are dialog schemas built from interaction types and situation viewpoints that relate to the objects, states, actions and plan structures of the microworlds that underlie the conversation. The NLP (Felshin, 1993), graphics, microworlds and tutorial reasoning are all in MCL2 Common Lisp with CLOS on a Mac-IIfx with 20M main memory. Dialog structuring work done in Prolog (Hashim and Hamburger, 1992) is being adapted by recoding and use of a Prolog-in-Lisp interpreter (Norvig, 1992).

INTERACTION TYPES FOR DIALOG SCHEMAS: The simplest interaction types consist of one linguistic or spatial move by the student or tutor and then one by the other of them. We have identified nine pedagogically useful types for language learning, listed below, each involving language use in one or both of its two moves. The names, initially a convenient shorthand for the tutor's role, evoke motivationally useful tutor personality traits (Murray, 1987; 1992); in each specification, S = student and T = tutor:

1.	Tourguide	T acts and comments;	S acknowledges
2.	Commander	T makes a command;	S executes it
3.	Narrator	T says something;	S enacts it
4.	Celebrity	T acts;	S tells about it
5.	Quizmaster	T asks a question;	S answers it
6.	Movecaster	S acts;	T describes it
7.	Oracle	S asks a question;	T answers it
8.	Servant	S makes a command;	T executes it
9.	Interpreter	S says something;	T enacts it

More complex types are needed for language errors and communication repair. The dialog schemas composed from these interaction types will differ depending on whether the student is to learn language or the subject matter of the microworld. Examples for the language case appear in Hashim and Hamburger (1992).

Interaction types work jointly with domain-level plans in the determination of the student's intentions. The plans place expectations on the sequence of actions and their resulting states, while the interaction types determine which agent is in control of the visual channel. Recognition of intentions leads in turn to the possibility of detecting the student's misunderstandings or misconceptions and repairing them. Spatial actions are implicitly associated with intentional structures, which the tutor may check for congruency with the currently active plan. For example, picking up an object typically induces in the observer the belief that the picker intends to use it. When the recognized intentions depart from the expectations, the system planner chooses a perspective (a view, see below) from which to deal with the misfit. In such a case, the realization of the tutor's turn may imply rhetorical acts such as drawing attention, arguing, suggesting.

The following dialog (typical of FLUENT-1 and of what we are working on generatively in FLUENT-2) uses this effect to detect and correct a misconception:

TUTOR:	Wash your face.
STUDENT:	<Picks up a box of detergent>
TUTOR:	That's the toothpaste. You can't wash your face with toothpaste. Pick up the soap instead.

The same rhetorical structuring of the tutor's reply might result from violations of microworld constraints (Murray, 1987):

TUTOR:	That's a glass table. You can't put the box on it. The box is too heavy and the glass would break.

From this point of view, this approach presents some similarities with the one taken by Maybury (1993).

The presence of the extra, visual medium in a communication system has significant effects on the use of the language medium, notably in the handling of the discourse phenomenon of reference identification, where it may be unnecessary and even infelicitous to include certain information that is visually available. Consider this sequence:

TUTOR:	Pick up the red book.
STUDENT:	<Picks it up>
TUTOR:	The table is clear now.

Here "the table" identifies the table that the red book was previously on, and which is visible. Suppose that another physically identical table is present, and consider the awkwardness of unnecessarily distinguishing them with a relative clause, thereby replacing the tutor's last comment with:

TUTOR: The table from which you picked up the red book is clear now.

These comments hinge on the visual availability of the relationship between the book and the table, which allows the two tables to be visually distinguished. If this exchange were to take place by telephone the conclusions would be different.

SITUATION VIEWPOINTS: The situation viewpoints referred to above involve the important aspects of the various microworlds: plans, actions and the state of objects. Plans permit non-arbitrary choice of action when initiative falls to the tutor, as it does in types 1-5. Our plan executor can call for language generation at top level, at the intermediate level of subplans, or at the fine-grained level of individual actions. Reference to actions can treat them as independent events or in relation to the (sub)plans containing them. In Movecaster, the student's unconstrained move can be described as an action or in terms of a resulting state or state change. The NPs in these descriptions can be made sensitive to various aspects of focus and specificity, thanks to the NLP software of the Athena Language Learning Project (Felshin, 1993).

We have developed a list of approximately 40 views in eight categories that relate to actions, states, plans, temporality, thematic roles, correction, and, at the level of the noun phrase, reference specification for individuals and for collectivities. Various state views of an action, for example, include (i) the new state of the object acted upon, (ii) the fact that (some property of) its preceding state no longer holds, (iii) an indirect change in a related object (such as a shelf being empty as a result of an object being picked up), (iv) an unchanged attribute of the object acted on. The earlier discussion of intentions is to be implemented in terms of such correction views. Two such views are (i) action disallowed because of unsatisfied precondition and (ii) some preconditions ok, but consequence is unwanted.

NON-LANGUAGE DOMAIN: We have extended our approach by implementing, in Prolog, a natural interface for a tutoring system in a domain outside of natural language (Hashim and Hamburger, 1993). By a natural interface, we mean one that provides readily grasped conventions for interactive graphics along with plan-based natural language explanations. We transport the FLUENT techniques to Coinland, a tutoring and learning environment for constrained exploration of problems in arithmetic. Coinland has a single microworld, one which is somewhat more abstract than those of FLUENT, both visually and conceptually. Like the FLUENT microworlds, Coinland represents an everyday situation, one in which coins can be moved around and exchanged for equal value (Hamburger and Lodgher, 1989). Here, instead of language, the student is to learn the domain, but there is still two-medium redundancy to support learning. We show how the system tutor can talk about the subtraction problem using a view processing mechanism and some of the views listed above. In Coinland, as in FLUENT, the familiarity of the objects and physical actions represented in the microworld, the clarity of the interface, and the use of redundancy combine to make it possible for the student to pick up new knowledge in an active and natural way.

ACKNOWLEDGEMENT: This work is supported by grant number IRI-9020711 from the National Science Foundation.

REFERENCES

Cohen, P. R. (1991) The role of natural language in a multimodal interface. Technical Note 514. Menlo Park, CA: SRI International.

Felshin (1993) A Guide to the Athena Language Learning Project Natural Language Processing System. Copyright, Massachusetts Institute of Technology.

Grosz, B.J. and Sidner, C.L. (1986) Attention, intentions and the structure of discourse. Computational Linguistics, 12, 3, 175-204.

Hamburger, H. and Crain, S. (1987) Plans and semantics in human language processing. Cognitive Science, 11, 1, 101-136.

Hamburger, H. and Hashim, R. (1992) Foreign language tutoring and learning environment. In Swartz, M. and Yazdani, M. (Eds.) Intelligent Tutoring Systems for Foreign Language Learning. New York: Springer-Verlag.

Hamburger, H. and Lodgher, L. (1989) Semantically constrained exploration and heuristic guidance. Machine-Mediated Learning, 3, 81-105.

Hamburger, H. and Maney, T. (1991) Twofold continuity in language learning. Computer-Assisted Language Learning, 4, 2, 81-92.

Hashim, R. and Hamburger, H. (1992) Discourse style and situation viewpoint for a conversational language tutor. Proceedings of the International Conference on Computer-Assisted Learning. New York and Berlin: Springer-Verlag.

Hashim, R. and Hamburger, H. (1993) Natural interfaces for ITS systems. Unpubl., George Mason University.

Hovy, E.H. (1988) Generating Natural Language under Pragmatic Constraints. Hillsdale, NJ: L. Erlbaum Assoc.

Hutchins, E.L., Hollan, J.D. and Norman, D.A. (1986) Direct manipulation interfaces. In D.A. Norman and S.W. Draper (Eds.) User-Centered System Design. Hillsdale, NJ: Lawrence Erlbaum Assoc.

Lambert, L. and Carberry, S. (1991) A tripartite plan-based model of dialogue. Proceedings of the 29th Annual Meeting of the Assn for Computational Linguistics, 47-54.

Litman, D.J. and Allen, J.F. (1987) A plan recognition model for subdialogues in conversations. Cognitive Science, 11, 2, 163-200.

Maney, T. and Hamburger, H. (1993) Pedagogical and cognitive foundations for an immersive foreign language environment. Submitted to the 1993 Conference of the Cognitive Science Society.

Maier, E.A. (1993) The representation of interdependencies between communicative goals and rhetorical relations in the framework of multimedia document generation. In this volume.

Maybury, M (1993) On structure and intention. In this volume.

McKeown, K.R. (1993) Language Generation for Multimedia Explanations. Vancouver, BC, Canada: Pacific Assn. Computational Linguistics Conference.

Murray, J.H. (1987) Humanists in an institute of technology: How foreign languages are reshaping workstation computing at MIT. Academic Computing.

Murray, J.H (1992) Creating scenarios for FLUENT. MIT. Consultant report to the FLUENT project.

Norvig, P. (1992) Paradigms of Artificial Intelligence Programming: Case Studies in Common Lisp. San Mateo, CA: Morgan Kaufmann Publishers.

Paris, C.L. (1991) The role of the user's domain knowledge in generation. Computational Intelligence, 7, 2, 71-93.

Sullivan, J.W. and Tyler, S.W. (1991) Intelligent User Interfaces. New York: Addison-Wesley.