Language Generation for Multimedia Healthcare Briefings

Kathleen R. McKeown
Shimei Pan and **James Shaw**
Dept. of Computer Science
Columbia University
New York, NY 10027, USA
kathy,pan,shaw@cs.columbia.edu

Desmond A. Jordan*
Barry A. Allen**
Dept. of Anesthesiology* and
Medical Informatics Dept.**
College of Physicians and Surgeons
Columbia University
New York, NY 10032

Abstract

This paper identifies issues for language generation that arose in developing a multimedia interface to healthcare data that includes coordinated speech, text and graphics. In order to produce brief speech for time-pressured caregivers, the system both combines related information into a single sentence and uses abbreviated references in speech when an unambiguous textual reference is also used. Finally, due to the temporal nature of the speech, the language generation module needs to communicate information about the ordering and duration of references to other temporal media, such as graphics, in order to allow for coordination between media.

1 Introduction

In a hospital setting it can be difficult for caregivers to obtain needed information about patients in a timely fashion. In a Cardiac Intensive Care Unit (ICU), communication regarding patient status is critical during the hour immediately following a coronary arterial bypass graft (CABG). It is at this critical point, when care is being transferred from the Operating Room (OR) to the ICU and monitoring is at a minimum, that the patient is most vulnerable to delays in treatment. During this time, there are a number of caregivers who need information about patient status and plans for care, including the ICU nurses who must prepare for patient arrival, the cardiologist who is off-site during the operation, and residents and attendings who will aid in determining post-operative care. The only people who can provide this information are those who were present during surgery and they are often too busy attending to the patient to communicate much detail.

To address this need, we are developing a multimedia briefing system, MAGIC (Multimedia Abstract Generation for Intensive Care), that takes as input online data collected during the surgical operation as well as information stored in the main databases at Columbia Presbyterian Medical Center (Roderer and Clayton, 1992). MAGIC generates a multimedia briefing that integrates speech, text, and animated graphics to provide an update on patient status (Dalal et al., 1996a). In this paper, we describe the issues that arise for language generation in this context:

- Conciseness: The generation process must make coordinated use of speech and text to produce an overview that is short enough for time pressured caregivers to follow, but unambiguous in meaning.
- Media specific tailoring: Generation must take into account that one output medium is speech, as opposed to the more usual written language, producing wording and sentence structure appropriate for spoken language.
- Coordination with other media: The language generation process must produce enough information so that speech and text can be coordinated with the accompanying graphics.

In the following sections, we first provide an overview of the full MAGIC architecture and then describe the specific language generation issues that we address. We close with a discussion of our current directions.

2 System Overview

MAGIC's architecture is shown in Figure 1. MAGIC exploits the extensive online data avail-

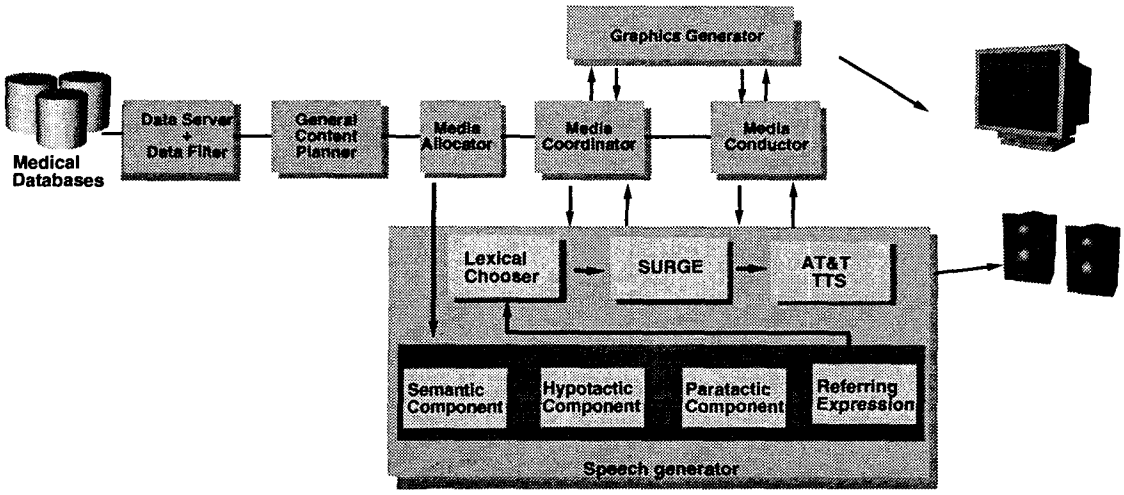


Figure 1: MAGIC system architecture.

able through Columbia Presbyterian Medical Center (CPMC) as its source of content for its briefing. Operative events during surgery are monitored through the LifeLog database system (Modular Instruments Inc.), which polls medical devices (ventilators, pressure monitors and alike) every minute from the start of the case to the end recording information such as vital signs. In addition, physicians (anesthesiologist and anesthesia residents) enter data throughout the course of the patient's surgery, including start of cardiopulmonary bypass and end of bypass as well as subjective clinical factors such as heart sounds and breath sounds that cannot be retrieved by medical devices. In addition, CPMC main databases provide information from the online patient record (e.g., medical history).

From this large body of information, the data filter selects information that is relevant to the bypass surgery and patient care in the ICU. MAGIC's content planner then uses a multimedia plan to select and partially order information for the presentation, taking into account the caregiver the briefing is intended for (nurse or physician). The media allocator allocates content to media, and finally, the media specific generators realize content in their own specific media (see (Zhou and Feiner, 1997) for details on the graphics generator). A media coordinator is responsible for ensuring that spoken output and animated graphics are temporally coordinated.

Within this context, the speech generator receives as input a partially ordered conceptual representation of information to be communicated. The generator includes a micro-planner, which is responsible for ordering and grouping information into sentences. Our approach to micro-planning integrates a variety of different types of operators for aggregation information within a single sentence. Aggregation using semantic operators is enabled through access to the underlying domain hierarchy, while aggregation using linguistic operators (e.g., hypotactic operators, which add information using modifiers such as adjectives, and paratactic operators which create, for example, conjunctions) is enabled through lookahead to the lexicon used during realization.

The speech generator also includes a realization component, implemented using the FUF/SURGE sentence generator (Elhadad, 1992; Robin, 1994), which produces the actual language to be spoken as well as textual descriptions that are used as labels in the visual presentation. It performs lexical choice and syntactic realization. Our version of the FUF/SURGE sentence generator produces sentences annotated with prosodic information and pause durations. This output is sent to a speech synthesizer in order to produce final speech. (Currently, we are using AT&T Bell Laboratories' Text To Speech System).

Our use of speech as an output medium provides an eyes-free environment that allows caregivers the opportunity to turn away from the display and continue carrying out tasks involving patient care. Speech can also clarify graphical conventions without requiring the user to look away from the graphics to read an associated text. Currently, communication between OR caregivers and

ICU caregivers is carried out orally in the ICU when the patient is brought in. Thus, the use of speech within MAGIC models current practice. Future planned evaluations will examine caregiver satisfaction with the spoken medium versus text.

3 Issues for Language Generation

In the early stages of system development, a primary constraint on the language generation process was identified during an informal evaluation with ICU nurses and residents (Dalal et al., 1996a). Due to time constraints in carrying out tasks, nurses, in particular, noted that speech takes time and therefore, spoken language output should be brief and to the point, while text, which is used to annotate the graphical illustration, may provide unambiguous references to the equipment and drugs being used. In the following sections, we show how we meet this constraint both in the speech content planner, which organizes the content as sentences, and in the speech sentence generator, which produces actual language.

In all of the language generation components, the fact that spoken language is the output medium and not written language, influences how generation is carried out. We note this influence on the generation process throughout the section.

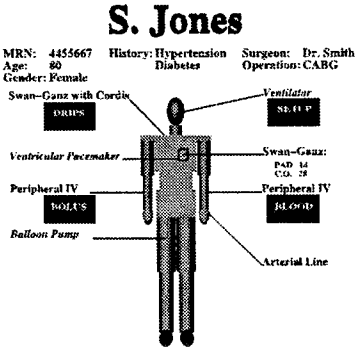
An example showing the spoken output for a given patient and a screen shot at a single point in the briefing is shown in Figure 3.

In actual output, sentences are coordinated with the corresponding part of the graphical illustration using highlighting and other graphical actions. In the paper, we show the kinds of modifications that it was necessary to make to the language generator in order to allow the media coordinator to synchronize speech with changing graphics.

3.1 Speech Micro-Planner

The speech micro-planner is given as input a set of information that must be conveyed. In order to ensure that speech is brief and yet still conveys the necessary information, the speech micro-planner attempts to fit more information into individual sentences, thereby using fewer words.

Out of the set of propositions given as input, the micro-planner selects one proposition to start with. It attempts to include as many other propositions as it can as adjectives or other modifiers of information already included. To do this, from the remaining propositions, it selects a proposition which is related to one of the propositions already selected via its arguments. It then checks whether it can be lexicalized as a modifier by looking ahead



Voice: Ms. Jones is an 80 year old, hypertensive, diabetic, female patient of Dr. Smith undergoing CABG. Presently, she is 30 minutes post-bypass and will arrive in the unit shortly. The existing infusion lines are two IVs, an arterial line, and a Swan-Ganz with Cordis. The patient has received massive vasotonic therapy, massive cardiotonic therapy, and massive-volume blood-replacement therapy. Drips in protocol concentrations are nitroglycerin, levophed, dobutamine, epinephrine, and inocor...

Figure 2: Multimedia presentation generated by MAGIC

to the lexicon used by the lexical chooser to determine if such a choice exists. The syntactic constraint is recorded in the intermediate form, but the lexical chooser may later decide to realize the proposition by any word of the same syntactic category or transform a modifier and a noun into a semantic equivalent noun or noun phrase.

The micro-planner uses information from the lexicon to determine how to combine the propositions together while satisfying grammatical and lexical constraints. Semantic aggregation is the first category of operators applied to the set of related propositions in order to produce concise expressions, as shown in lower portion of Fig. 1. Using ontological and lexical information, it can reduce the number of propositions by replacing them with fewer propositions with equivalent meanings. While carrying out hypotactic aggregation operators, a current central proposition is selected and the system searches through the un-aggregated propositions to find those that can be realized as adjectives, prepositional phrases and relative clauses, and merges them in. After hypotactic aggregation, the un-aggregated propositions are then combined using paratactic operators, such as appositions or coordinations.

```
X is a patient.
X has property last name = Jones.
X has property age = 80 years old.
X has property history = hypertension property.
X has property history = diabetes property.
X has property gender = female.
X has property surgery = CABG.
X has  property doctor = Y.
Y has property last name = Smith.
```

Figure 3: propositions for the first sentence

In the first sentence of the example output, the micro-planner has combined the 9 input propositions shown above in Figure 3 into a single sentence: Ms Jones is an 80 year old hypertensive, diabetic female patient of Dr. Smith undergoing CABG. In this example this is possible, in part because the patient's medical history (diabetes and hypertension) can be realized as adjectives. In another example, "Mr. Smith is a 60 year old male patient of Dr. Jordan undergoing CABG. He has a medical history of transient ischemic attacks, pulmonary hypertension, and peptic ulcers.", the medical history can only be realized as noun phrases, thus requiring a second sentence and necessarily, more words.

3.2 Speech Sentence Generator

The speech sentence generator also contributes to the goal of keeping spoken output brief, but informative. In particular, through its lexical choice component, it selects references to medical concepts that are shorter and more colloquial than the text counterpart. As long as the text label on the screen is generated using the full, unambiguous reference, speech can use an abbreviated expression. For example, when referring to the devices which have been implanted, speech can use the term "pacemaker" so long as the textual label specifies it as "ventricular pacemaker". Similarly, MAGIC uses "balloon pump" in speech instead of "intra-aortic balloon pump", which is already shown on the screen.

In order to do this, lexical choice in both media must be coordinated. Lexical choice for text always selects the full reference, but lexical choice for speech must check what expression the text generator is using. Basically, the speech lexical chooser must check what attributes the text generator includes in its reference and omit those.

Finally, we suspect that the syntactic structure of sentences generated for spoken output should be simpler than that generated for written language. This hypothesis is in conflict with our criteria for generating as few sentences as possible, which often results in more complex sentences. This is in part acceptable due to the fact that MAGIC's output is closer to formal speech, such as one might find in a radio show, as opposed to informal conversation. It is, after all, a planned one-way presentation. In order to make the generated sentences more comprehensible, however, we have modified the lexical chooser and syntactic generator to produce pauses at complex constitutions to increase intelligibility of the output. Currently, we are using a pause prediction algorithm which utilizes the sentence's semantic structure, syntactic structure as well as the linear phrase length constraint to predict the pause position and relative strength. Our current work involves modifying the FUF/SURGE language generation package so that it can produce prosodic and pause information needed as input to a speech synthesizer, to produce a generic spoken language sentence generator.

3.3 Producing Information for Media Coordination

Language generation in MAGIC is also affected by the fact that language is used in the context of other media as well. While there are specific modules in MAGIC whose task is concerned with utilizing multiple media, media coordination affects the language generation process also. In particular, in order to produce a coordinated presentation, MAGIC must temporally coordinate spoken language with animated graphics, both temporal media. This means that spoken references must be coordinated with graphical references to the same information. Graphical references may include highlighting of the portion of the illustration which refers to the same information as speech or appearance of new information on the screen. Temporal coordination involves two problems: ensuring that ordering of spoken references to information is compatible with spatial ordering of the graphical actions and synchronizing the duration of spoken and graphical references (Dalal et al., 1996b).

In order to achieve this, language generation must provide a partial ordering of spoken references at a fairly early point in the generation process. This ordering indicates its preference for how spoken references are to be ordered in the output linear speech in accordance with both graphical and presentation constraints. For example, in the first sentence of the example shown in Figure 3, the speech components have a preference for medical history (i.e., "hypertensive, diabetic") to be

presented *before* information about the surgeon, as this allows for more concise output. It would be possible for medical history to be presented after all other information in the sentence by generating a separate sentence (e.g., "She has a history of hypertension and diabetes.") but this is less preferable from the language point of view. In our work, we have modified the structure of the lexical chooser so that it can record its decisions about ordering, using partial ordering for any grammatical variation that may happen later when the final syntactic structure of the sentence is generated. These are then sent to the media coordinator for negotiating with graphics an ordering that is compatible to both. Details on the implementation of this negotiation are presented in (Dalal et al., 1996b) and (Pan and McKeown, 1996).

In order to synchronize duration of the spoken and graphical references, the lexical chooser invokes the speech synthesizer to calculate the duration of each lexical phrase that it generates. By maintaining a correspondence between the referential string generated and the concepts that those referential actions refer to, negotiation with graphics has a common basis for communication. In order to provide for more flexible synchronization, the speech sentence generator includes facilities for modifying pauses if conflicts with graphics durations arise (see (Pan and McKeown, 1996) for details).

4 Related Work

There is considerable interest in producing fluent and concise sentences. EPICURE (Dale, 1992), PLANDOC(Kukich et al., 1994; Shaw, 1995), and systems developed by Dalianis and Hovy (Dalianis and Hovy, 1993) all use various forms of conjunction and ellipsis to generate more concise sentences. In (Horacek, 1992) aggregation is performed at text-structure level. In addition to conjoining VP and NPs, FLOWDOC(Passonneau et al., 1996) uses ontological generalization to combine descriptions of a set of objects into a more general description. Based on a corpus analysis in the basketball domain, (Robin, 1994) catalogued a set of revision operators such as adjoin and nominalization in his system STREAK. Unlike STREAK, MAGIC does not use revision to combine information in a sentence.

Generating spoken language from meanings or concepts (Meaning to Speech, MTS) is a new topic and only a few such systems were developed in recent years. In (Prevost, 1995) and (Steedman, 1996), they explore a way to generate spoken language with accurate contrastive stress based on information structure and carefully modeled domain knowledge. In (Davis and Hirschberg, 1988), spoken directions are generated with richer intonation features. Both of these systems took advantage of the richer and more precise semantic information that is available during the process of Meaning to Speech production.

5 Conclusions and Current Directions

The context of multimedia briefings for access to healthcare data places new demands on the language generation process. Language generation in MAGIC addresses its user's needs for a brief, yet unambiguous, briefing by coordinating spoken language with the accompanying textual references in the graphical illustration and by combining information into fewer sentences. It also must explicitly represent its decisions as it generates a sentence in order to provide information to the media coordinator for negotiation with graphics.

Our development of MAGIC is very much an ongoing research project. We are continuing to work on improved coordination of media, use of the syntactic and semantic structure of generated language to improve the quality of the synthesized speech, and analysis of a corpus of radio speech to identify characteristics of formal, spoken language.

6 Acknowledgments

MAGIC is a joint project which involves the Natural Language Processing group (the authors), the Graphics and User Interface group (Steve Feiner, Michelle Zhou and Tobias Hollerer), the Knowledge Representation group (Mukesh Dalal and Yong Feng) in the Department of Computer Science of Columbia University and Dr. Desmond Jordan and Prof. Barry Allen at the Columbia College of Physicians and Surgeons (authors). This work is supported by DARPA Contract DAAL01-94-K-0119, the Columbia University Center for Advanced Technology in High Performance Computing and Communications in Healthcare (funded by the New York State Science and Technology Foundation) and NSF Grants GER-90-2406.

References

M. Dalal, S. Feiner, K. McKeown, D. Jordan, B. Allen, and Y. alSafadi. 1996a. Magic: An experimental system for generating multimedia briefings about post-bypass patient status. In *Proceedings of American Medical Informatics Association 1996 Fall.*

M. Dalal, S. Feiner, K. McKeown, S. Pan, M. Zhou, T. Hollerer, J. Shaw, Y. Feng, and J. Fromer. 1996b. Negotiation for automated generation of temporal multimedia presentations. In *Proceedings of ACM Multimedia '96.*

R. Dale. 1992. *Generating Referring Expressions: Constructing Descriptions in a Domain of Objects and Processes.* MIT Press, Cambridge, MA.

H. Dalianis and E. Hovy. 1993. Aggregation in natural language generation. In *Proceedings of the Fourth European Workshop on Natural Language Generation*, pages 67–78, Pisa, Italy.

J. Davis and J. Hirschberg. 1988. Assigning intonational features in synthesized spoken discourse. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, pages 187–193, Buffalo, New York.

M. Elhadad. 1992. *Using argumentation to control lexical choice: A functional unification-based approach.* Ph.D. thesis, Computer Science Department, Columbia University.

H. Horacek. 1992. An integrated view of text planning. In *Aspects of Automated Natural Language Generation*, pages 29–44. Springer-Verlag.

K. Kukich, K. McKeown, and J. Shaw. 1994. Practical issues in automatic documentation generation. In *Proceedings of the 4th ACL Conference on Applied Natural Language Processing*, pages 7–14, Stuttgart.

S. Pan and K. McKeown. 1996. Spoken language generation in a multimedia system. In *Proceedings of ICSLP 96*, volume 1, pages 374–377, Philadelphia, PA.

R. Passonneau, K. Kukich, V. Hatzivassiloglou, L. Lefkowitz, and H. Jing. 1996. Generating summaries of work flow diagrams. In *Proceedings of the International Conference on Natural Language Processing and Industrial Applications*, pages 204–210, New Brunswick, Canada, June. Univeristy of Moncton.

S. Prevost. 1995. *A Semantics of Contrast and Informaiton Structure for Specifying Intonation in Spoken Language Generation.* Ph.D. thesis, University of Pennsylvania.

J. Robin. 1994. *Revision-Based Generation of Natural Language Summaries Providing Historical Background.* Ph.D. thesis, Computer Science Department, Columbia University.

N. Roderer and P. Clayton. 1992. Iaims at columbia presbyterian medical center: Accomplishments and challenges. In *Bull. Am. Med. Lib. Assoc.*, pages 253–262.

J. Shaw. 1995. Conciseness through aggregation in text generation. In *Proceedings of the 33rd ACL (Student Session)*, pages 329–331.

M. Steedman. 1996. Representing discourse informationn for spoken dialogue generation. In *Proceedings of ISSD 96*, pages 89–92, Philadelphia, PA.

M. Zhou and S. Feiner. 1997. Top-down hierarchical planning of coherent visual discourse. In *Proc. IUI '97 (1997 Int. Conf. on Intelligent User Interfaces)*, Orlando, FL, January 6–9.