Integrating Communicative Goals For Real-time Clinical Decision Support[1]

Sandra Carberry
Dept. of Computer Science
University of Delaware
Newark, DE 19716

Terrence Harvey
Dept. of Computer Science
University of Delaware
Newark, Delaware 19716

John R. Clarke M.D.
Dept. of Surgery
Allegheny Univ. of the Health Sciences
Philadelphia, PA 19129

A critiquing system evaluating a physician's management plan may produce a set of individual comments that, taken together, appear repetitious or incoherent. This paper presents TraumaGEN, a system for integrating sets of possibly inter-related communicative goals into one or more coherent messages. TraumaGEN takes account of the purpose of the messages, the situation in which the messages will be received, and the social role of the system. Preliminary evaluation of TraumaGEN indicates that it produces coherent integrated messages.

INTRODUCTION

To deliver real-time decision support effectively in trauma management, a system must not only know what is appropriate for patient care, but also be able to communicate such support concisely and coherently to the physician. We take the unit of communication to be a *communicative goal* and its realization in language (or a combination of language and graphics). In much of trauma management, a system may have several different communicative goals at the same time. It therefore requires a message generator that can take an arbitrary and often inter-related set of communicative goals and produce text that realizes the entire set in a concise and coherent form.

In this paper, we describe the kind of decision-support that TraumAID, a system for emergency center trauma care, and its critiquing interface TraumaTIQ have been designed to provide. We then present our solution to the problem of producing concise and coherent texts that convey this decision-support. Throughout the paper we present examples from our implemented system, and the paper includes the results of an evaluation of the resultant messages.

REAL-TIME DECISION SUPPORT

TraumAID[1] is a decision support system for addressing the initial definitive management of multiple trauma. TraumAID consists of two components: 1) a rule-based reasoner that draws conclusions from the available data and posts appropriate diagnostic and therapeutic goals, and 2) a planner that constructs a plan for addressing these goals. One important feature of TraumAID's planner is that it recognizes where actions chosen to satisfy one goal can be used in satisfying other ones. Thus, as one might expect in a domain where time is at a premium, actions are frequently *overloaded* to satisfy multiple goals. Initial evaluation and validation studies indicate that TraumAID produces high-quality diagnostic and therapeutic plans for managing patient care in both simple and complex trauma cases [2].

To *use* such plans for real-time clinical decision support, an interface was developed, TraumaTIQ [2, 3], that uses TraumAID's plans to produce critiques in response to physician orders. Because orders involve future actions, they can be modified or cancelled by the physician. Thus TraumaTIQ's critiques can lead to changes in what is actually done, by focusing solely on clinically relevant differences between the physician's orders and TraumAID's plan.

TraumaTIQ hypothesizes the physician's plan based on his or her orders and actions and the current status of the patient, identifies differences between that plan and TraumAID's current plan, and then notifies the physician of any discrepancies that could seriously impact patient care. By inferring the physician's plan, TraumaTIQ can take into account *why* the physician may be performing a particular action when deciding whether to produce a critique and, if so, what information to include in the critique. Thus not only can it avoid criticizing relatively minor differences in plans, but it can also couch its recommendations in terms of the goals that the physician *is pursuing* or that the system believes *should be addressed.*

TraumaTIQ has been designed to recognize four types of differences between plans:

1. Errors of Omission - where an action needed to achieve a management goal has not been ordered within a few minutes of the goal having been posted;
2. Errors of Commission - where an action has been ordered but is either unmotivated by a known goal or is subsumed by some other action;
3. Procedure Choice Errors - where a less costly or less invasive action would be preferred;
4. Scheduling Errors - where another action should be scheduled before the ordered action due to the relative urgency of the actions, constraints on the temporal order of actions in a procedure, or dependencies between the actions.

[1]This work was supported by the National Library of Medicine under grant R01-LM-05764-01. We wish to thank Bonnie Webber for her comments and suggestions on this paper.

TraumaTIQ's critiques are conveyed using natural language sentences generated by filling in sentence schemata.

TraumaTIQ has been compared with human judges evaluating trauma care [2, 3] and found to have several behavioral similarities with them vis-a-vis the production of critiques. The problem we address here is that, while in isolation each of TraumaTIQ's critiques may effectively warn a physician about a problem in their plan, in most cases where the physician's plan differs from TraumAID's, several problems are detected and thus multiple critiques are produced. We found that some critiques detracted from other ones, that some critiques would make more sense if they took explicit account of those appearing earlier, and that there was informational overlap among the critiques. We felt that a text planner could help in producing more coherent and concise messages that achieve multiple communicative goals.

Although natural language generation has been used in other health care systems [4, 5], these applications have not required that several independent comments be combined into a single integrated message. While the HealthDoc system (currently under development) [6] produces for a patient a simplified version of a physician's "master text", it is focused on surface-level changes to the extracted text that enhance coherence (e.g. inserting pronouns, deleting extraneous references, etc.) As future systems make use of independent modules with different sources of expertise, such as a Bayesian reasoner, a rule-based inference engine, etc., they also will need to integrate possibly disparate recommendations into a coherent message and will need to adopt techniques similar to ours.

CONSTRUCTING INTEGRATED MESSAGES

TraumaGEN starts with several communicative goals and a means for achieving each goal in isolation. It uses a set of transformational rules that transform these into coherent message units that achieve the overall set of goals. These message units are then translated into natural language using sentence schemata.

The nature of trauma management suggested several objectives that should influence the generation process and which are reflected in TraumaGEN:

- Purpose: Since the purpose of the messages is to support decision-making, the system's recommendations should continue to be organized in terms of relevant domain goals.
- Situation: Since the emergency center requires rapid decision-making in a chaotic setting, the messages must be succinct, unambiguous, and easily assimilated.
- Social role: Since the system's social role on the medical team is that of an expert consultant to the physician who retains responsibility for the quality of patient care, it must recognize that the physician can ignore its recommendations.

Integrating Critiques into Message Units

We analyzed 5361 individual critiques comprising 753 critique sets produced by TraumaTIQ, where a critique set represents the critiques resulting from a single order. This revealed 22 common patterns of inter-related critiques, each pattern covering a subpart of a critique set. We designed a set of nine rules for transforming such inter-related critiques into integrated messages.

Interacting Critiques. Our analysis of TraumaTIQ's original output showed that one critique could detract from another critique, although each was both justified and coherent in isolation. Consider, for example, the following two critiques:

> **Performing local visual exploration of all abdominal wounds is preferred over doing a peritoneal lavage for ruling out a suspicious abdominal wall injury.*

> **Please remember to check for laparotomy scars before you do a peritoneal lavage.*

While individually, each critique is coherent, together they are not: the first is meant to dissuade the physician from doing a peritoneal lavage, while the second assumes one will be done. In some situations, such as tutoring, the system might discard the second critique; however, in real-time decision-support, discarding the second critique would be inappropriate since the system's advice about the preferred procedure may not be followed. Thus the system should produce a coherent message that integrates the two critiques.

Our solution to this problem is to allow revision rules that are triggered when two critiques have an apparent conflict. In such cases, the two critiques are merged into a single message, and the conflicting critique is revised. For example, our rule *Revise-Conflict* is triggered whenever a critique whose goal is that an ordered action be properly scheduled occurs with a critique whose intention is that the action be replaced with one more highly preferred. *Revise-Conflict* merges the two critiques into a single message that ends by *conceding* that the ordered action might still be done and giving the scheduling critique in case it is. Thus, the previous two conflicting critiques would be revised and merged into the following message:

> *Performing local visual exploration of all abdominal wounds is preferred over doing a peritoneal lavage for ruling out a suspicious abdominal wall injury. However, if you do a peritoneal lavage, then remember to first check for laparotomy scars.*

Note that the new message still recommends the better procedure, but leaves the final choice in the hands of the physician who is responsible for quality patient care.

Another rule, *Revise-Interactions*, is triggered when a critique whose goal is to postpone a dependent action occurs in conjunction with a critique whose goal is execution of the action on which the dependency is based. For example, consider the following two individual critiques produced by TraumaTIQ:

> **Caution: do a peritoneal lavage immediately as part of ruling out abdominal bleeding.*
>
> **Do not reassess the patient in 6 to 24 hours until after doing a peritoneal lavage. The outcome of the latter may affect the need to do the former.*

While technically the two critiques do not conflict, TraumaGEN's revision rules exploit the relation between their communicative goals to produce the following more concise and coherent message:

> **Do a peritoneal lavage immediately as part of ruling out abdominal bleeding. Use the results of the peritoneal lavage to decide whether or not to reassess the patient in 6 to 24 hours.*

Informational Overlap. One prevalent problem in TraumaTIQ's output is *informational overlap* — actions often appear in several different warnings and thus the message as a whole appears repetitious. The top part of Figure 1 presents three such overlapping critiques produced by TraumaTIQ. In some cases we found actions were repeated over as many as four critiques.

Our approach for merging such critiques is motivated by four criteria: 1) group by relevant treatment goals; 2) avoid repeated mention of the same action, since this may erroneously imply that it be done several times; 3) produce concise messages; and 4) produce few, rather than many, individual messages. Our approach draws upon work in rhetorical structure theory[7] which posits that a coherent text plan consists of segments related to one another by rhetorical relations. So another transformational rule, *Combine-Similar-Intentions*, finds critiques with different goals but a subset of similar actions and considers merging them into a single message consisting of segments related by a *Sequence* relation. This is illustrated in the bottom part of Figure 1.

Since the above criteria may sometimes be in conflict, TraumaGEN evaluates the possible ways in which overlapping critiques might be combined into a merged message with a metric that weighs 1) the reduction in repetition of actions; 2) whether goals must be repeated; 3) the number of individual critiques that are merged; and 4) the number of segments comprising a message. The first three measure how well a merge achieves concise, unambiguous, and easily assimilated messages (as required by the *situation* in which the messages will be received). Reducing the number of repeated action specifications contributes both to concise messages and less ambiguity; however, achieving this often requires repeated goal specifications which detracts from conciseness. We hypothesize that a few coherent messages will be more easily assimilated than many individual messages and thus our metric takes into account the number of individual critiques that are merged into the resultant message. The last term in our metric takes into account the number of sequentially related segments comprising a merged message, since actions related to a single treatment goal will now be distributed over the segments. We arbitrarily limit combined messages to three segments in order to maintain a goal-oriented organization, as dictated by the *purpose* of our messages. Figure 2 presents a message that results from applying several rules, including both *Combine-Similar-Intentions* and *Revise-Interactions*.

Trailing Comments. When several critiques are merged into a single message, the message may refer to actions that are also part of critiques that did not participate in the merge. Once those actions have been introduced in the merged message, focusing heuristics[8] suggest that the other critiques referencing these actions be included in the merged message as well. However, rather than restructure

<u>Original Critiques with Informational Overlap:</u>

**Caution: get a post chest tube x-ray immediately to evaluate the chest.*

**Caution: insert a right chest tube and get a post chest tube x-ray immediately to treat the simple right pneumothorax.*

**Caution: insert a right chest tube, close the right chest wound, cover all chest wounds with occlusive dressings, and get a post chest tube x-ray immediately to treat the right open sucking chest wound.*

<u>Merged Message:</u>

**Caution: insert a right chest tube to treat the simple right pneumothorax and the right open sucking chest wound. Next close the right chest wound and cover all chest wounds with occlusive dressings to further treat the right open sucking chest wound. Then get a post chest tube x-ray to complete both goals and also to evaluate the chest.*

Figure 1: An Example of a Merged Message

Original Critiques Produced by TraumaTIQ:

**Caution: get a chest x-ray immediately to rule out a simple right pneumothorax.*

**Caution: get a chest x-ray immediately to rule out a simple right hemothorax.*

**Do not perform local visual exploration of all abdominal wounds until after getting a chest x-ray. The outcome of the latter may affect the need to do the former.*

**Please get a chest x-ray before performing local visual exploration of all abdominal wounds because it has a higher priority.*

Merged Message:

**Caution: get a chest x-ray to rule out a simple right pneumothorax and a simple right hemothorax, and use the results of the chest x-ray to decide whether or not to perform local visual exploration of all abdominal wounds.*

Figure 2: A Merged Message Using Several Rules

the result of our merge transformation, we append them to the end of the message. Thus we refer to them as *trailing comments*.

Unfortunately, trailing comments have the potential to erroneously suggest new instances of actions. Our solution to this problem is to (1) make the focused action the subject of the sentence, reflecting its *given* status in the discourse, and (2) utilize clue words to call additional attention to its occurrence earlier in the message and to the new information being conveyed. Thus the first trailing comment is introduced with the clue word *moreover* since it implies more being said about something already discussed. The clue word *also* is used to introduce the additional information. In one such example constructed by TraumaGEN, a critique that is not included in a merged message involves an action (checking for muffled heart sounds) that is part of the merged message. Thus this critique is realized as the following trailing comment at the end of the merged message:

Moreover, checking for muffled heart sounds is also indicated to assess the possibility of a pericardial injury.

A trailing comment may need to refer to other actions in addition to the one previously focused on. We accomplish this by subordinating those actions in a phrase introduced by the clue words *along with*, in a sentence in which the previously focused action is the subject, as in the following trailing comment about doing a laparotomy:

Moreover, doing the laparotomy is also indicated, along with repairing the left diaphragm, to treat the lacerated left diaphragm.

Other Considerations. Communication is best understood when it reflects *shared knowledge.* In the emergency center, shared knowledge can be equated with the current state of the case, as entered into the computer-based medical record (CBMR). When a procedure is ordered, it thus becomes part of this shared knowledge. Consequently, we use definite noun phrases to refer both to procedures and actions already introduced into the treatment plan by one of the system's messages and to entities introduced via data entry into the CBMR. For example, even though a peritoneal lavage does not appear in any of the system's earlier messages, TraumaGEN produces the following message about a related scheduling precondition:

Please remember to check for laparotomy scars before you do the peritoneal lavage.

Shared knowledge is also relevant when the system disagrees with the physician about whether a procedure is appropriate. Since the use of a definite noun phrase suggests an action's acceptance into the treatment plan, an indefinite noun phrase is used to refer to a procedure about which there is conflict. For example, if the physician has ordered a peritoneal lavage and the system believes that it may not be needed, depending on the results of a chest x-ray, TraumaGEN would generate the message

Do not do a peritoneal lavage until after getting a chest x-ray since the outcome of the latter may affect the need to do the former.

RESULTS

While TraumaGEN's effectiveness can only be measured by deploying it in a trauma bay and evaluating the degree to which its messages affect the physician's plan, preliminary evaluation can be used to determine its limitations and to identify where further work is needed.

We ran TraumaGEN on 48 collected cases of actual trauma care under a scenario in which critiques were produced after each physician order. We extracted a set of critiques from the middle of each of the 48 cases and used them in our analysis.[2]

[2]We used a set of critiques from the middle of each case since there is nothing to critique at the very beginning of a case and little to critique at the end. It is generally the middle of a case where critiques appear in sufficient number to consider the effect of TraumaGEN.

We compared the critiques generated by TraumaTIQ alone with the messages produced when it was augmented with TraumaGEN, using four criteria:

1. The number of individual messages produced;
2. The conciseness of the messages;
3. The number of focus shifts required to assimilate the information;
4. The coherence and subjective quality of the messages.

The number of individual critiques was reduced by 18% in the 48 critique sets examined. The results for individual sets ranged from no reduction in cases where the critiques were independent of one another, to 60% in critique sets that were heavily interrelated. More concise messages resulted from a 12% reduction in the number of references to diagnostic or therapeutic actions and goals. Many references were replaced by pronouns, making the critiques shorter and more natural. So the number of actual repetitions of an action or goal decreased by an additional 7%, for a total reduction of 19%.

TraumaGEN kept the same procedures in focus for a longer time. Twenty-three instances of unnecessary topic change were eliminated from the critique sets out of a possible 79 (29%). Further improvement in comprehensibility was gained by signaling an impending focus change six times in the new critiques.

To evaluate coherence and quality of the messages, we asked a human subject not affiliated with our project to compare the new messages to the original ones. The subject was given the messages produced by TraumaTIQ and TraumaGEN for a dozen cases, including the first eleven cases in which the message sets produced by TraumaTIQ and TraumaGEN were distinct from each other and from previous cases. A later case was included because we wanted our test set to include at least one conflicting critique. The old and new messages were presented as pairs, sometimes TraumaTIQ's messages first and sometimes TraumaGEN's messages first. The written instructions given to the subject requested that he note whether one set of messages was more comprehensible than another and, if so, his judgment as to why that was the case.

In ten of the twelve cases, the subject preferred TraumaGEN's messages; in eight of these cases, the preference was very strong while in the other two cases it was moderate. In one case, the subject preferred the original TraumaTIQ messages, and in another case the subject had no preference. Where the subject preferred the original messages produced by TraumaTIQ, the preference was based on the English translation of two goals: the subject found the phrasing confusing when the messages were combined (since the two goals had very similar translations) but not confusing when the messages were separated. This problem is easily remedied. The subject's comments indicated that his preferences for TraumaGEN's messages were generally based on reduction of repetition, merging of related messages, and elimination of conflict.

DISCUSSION

Our message planner, TraumaGEN, draws on previous work in discourse theory to produce integrated messages from individual critiques each of which is designed to achieve its own communicative goal. The need to construct coherent text from multiple individual text plans is a problem that will increasingly face clinical decision support systems as sophisticated systems distribute their processing across individual modules, each of which may produce a recommendation or have some other need for communicating with the user. In determining how to transform individual critiques into coherent integrated messages, TraumaGEN takes into account knowledge about the *purpose* of the messages, the *situation* in which the messages will be received, and the *social role* of the system. Preliminary evaluation of TraumaGEN indicates that it successfully constructs concise, coherent messages from an arbitrary and often inter-related set of critiques.

References

[1] B. L. Webber, R. Rymon, and J. R. Clarke. Flexible support for trauma management through goal-directed reasoning and planning. *A. I. in Medicine*, 4:145–163, 1992.

[2] A. Gertner, B. L. Webber, J. R. Clarke, C. Hayward, T. Santora, and D. Wagner. On-line quality assurance in the initial definitive management of multiple trauma: Evaluating system potential. *A. I. in Medicine*, to appear.

[3] A. Gertner and B. L. Webber. A Bias Towards Relevance: Recognizing Plans Where Goal Minimization Fails. *Proc. of the National Conference on Artificial Intelligence*, 1133-1138, 1996.

[4] B. Buchanan, J. Moore, D. Forsythe, S. Ohlsson, and G. Banks. Involving patients in health care: Using medical informatics for explanation in the clinical setting. *Proc. of the 16th Symposium on Computer Applications in Medical Care*, 510–514, 1992.

[5] K. Binsted, A. Cawsey, and R.B. Jones. Generating personalized information using the medical record. *Proc. of AIME*, 29–41, 1995.

[6] G. Hirst and C. DiMarco. Automatic customization of health-education brochures for individual patients. *Proc. of the Conf. on Information Technology in Community Health*, 1996.

[7] W. C. Mann and S. A. Thompson. Rhetorical Structure Theory: A Theory of Text Organization. Technical Report ISI/RS-87-190, ISI/USC, June 1987.

[8] K. R. McKeown. *Text Generation*. Cambridge University Press, Cambridge, 1985.