

# Improving Question-Answering With Linking Dialogues

Sudeep Gandhe

Institute for Creative Technologies  
University of Southern California  
13274 Fiji Way  
Marina Del Rey, CA 90292 USA  
+1 310 574 5700

gandhe@ict.usc.edu

Andrew S. Gordon

Institute for Creative Technologies  
University of Southern California  
13274 Fiji Way  
Marina Del Rey, CA 90292 USA  
+1 310 574 5700

gordon@ict.usc.edu

David Traum

Institute for Creative Technologies  
University of Southern California  
13274 Fiji Way  
Marina Del Rey, CA 90292 USA  
+1 310 574 5700

traum@ict.usc.edu

## ABSTRACT

Question-answering dialogue systems have found many applications in interactive learning environments. This paper is concerned with one such application for Army leadership training, where trainees input free-text questions that elicit pre-recorded video responses. Since these responses are already crafted before the question is asked, a certain degree of incoherence exists between the question that is asked and the answer that is given. This paper explores the use of short linking dialogues that stand in between the question and its video response to alleviate the problem of incoherence. We describe a set of experiments with human generated linking dialogues that demonstrate their added value. We then describe our implementation of an automated method for utilizing linking dialogues and show that these have better coherence properties than the original system without linking dialogues.

## Categories and Subject Descriptors

I.2.1 [Applications and Expert Systems]: *Natural language interfaces*

## General Terms

Algorithms

## Keywords

Question-answering dialogue systems, coherence, training

## INTRODUCTION

Natural language interfaces to information-providing systems often use question-answering dialogues. These systems vary in many ways, including the amount of user initiative allowed, the technologies used for the generation of responses, and the production value of the responses. For example, kiosks at museums may have a very constrained interaction using menu-based interfaces, while applications for teaching may give users more freedom with the aim of engaging them in conversations more actively. There is also often a trade-off between production quality and flexibility. Maximum flexibility can be achieved by using text generation, text-to-speech synthesis, and animated body movements, however this will generally not look or sound as good as produced video from entertainment industry professionals. For some applications, the desired levels of

immersiveness and emotional expressivity require produced media for a precise effect, even at the cost of generality. In systems where the users' input is unconstrained and the output is highly constrained, a potential coherence problem exists when no good system response to the user is available to output. There are two sources of this coherence problem; 1) the question asked by the user may not have any appropriate response in the output collection, 2) a response with appropriate content may be available, but it is not phrased in a manner that is appropriate to the user's question. In this paper, we describe a study and implementation of a *mediation* interface, where linking dialogues are used to bridge from the question to the answer in the hopes of increasing the coherence of the system without sacrificing production quality.

## The AXL Application

The Institute for Creative Technologies at the University of Southern California has developed a leadership training application as part of the Army Excellence in Leadership (AXL) project [4]. It follows a case-analysis model of instruction. The trainee is first shown a short movie depicting a challenging Army leadership situation. For example, one scenario depicts a security mission for a food distribution operation, where the leader, CPT Young, makes several mistakes that lead to the failure of the mission. After viewing the film, the trainee then engages in a natural language dialogue with the characters from this fictional story, interviewing them to figure out why things went wrong and what might have been done differently. There is also a virtual character, mentor, who guides the trainee through the analysis of the case. During the interviews, the trainee can type free-text questions and view responses in the form of pre-recorded video clips of the character being interviewed. The system selects the most appropriate response (from an average of 16 responses per character) using a Naive Bayes machine-learning algorithm. With modest amounts of training data (356 examples on average) the classifier achieves an average accuracy of 52% (10-fold cross validation) using word-level features (unigrams and bigrams). The availability of corpus of questions prepared from actual user interactions and the high user initiative, high immersive experience of the AXL application make this an ideal platform to test the idea of mediation using linking dialogues.

## HUMAN-AUTHORED LINKING

In order to investigate the role that linking dialogues could have in improving the coherence of AXL interviews, we started by creating a corpus of human-authored linking dialogues. Five users were presented with a set of 20 example user questions along with five most likely video responses, as determined using the existing classification algorithm. For every question, the

Copyright is held by the author/owner(s).

*IUI'06*, January 29–February 1, 2006, Sydney, Australia.

ACM 1-59593-287-9/06/0001.

raters selected the most appropriate video response and wrote a linking dialogue that would maximize the coherence between the question and the response that they chose. Raters were allowed to write links that would appear before and/or after the selected response. We then analyzed the resulting corpus of human-authored linking dialogues in order to characterize the types of mediations that would have to be generated in order to maximize coherence (for more information about this analysis, see [2]). We identified several functions that these linking dialogues served, including:

**Speaker Introduction:** Several human-authored linking dialogues introduced the character whose video response is chosen for a question, mainly to emphasize the reason why this character is selected.

**Coherence:** This type of a linking dialogue mainly states why the particular video answer is chosen and how it connects to the question being asked. This type of link invariably involves adding more content than that present in question or answer.

**Reformulation:** The response was originally authored with a specific user question in mind, which has some relationship to the question that was actually asked. This type of link tries to explicate this relationship.

**Focus/Summarizing:** This type of link addresses the problem where the response is not phrased appropriately by rephrasing the answer at the beginning or at the end of the response.

**Guidance:** In the course of the interview, the response to a question can influence the next user question. This type of link can be used by the system for guiding the dialogue to achieve particular training goals.

These types of functions for a link are not exclusive. A single human-authored linking dialogue may serve several of these functions. For example, speaker introduction function appears in almost every human-authored linking dialogue.

For evaluating whether human generated linking dialogues add to the coherence of the overall interaction, the coherence of question-link-response (with mediation) was compared with that of question-response (without mediation) interactions. For this evaluation we used a Likert scale for scoring the coherence of question-response pairs, from 1 (bad) to 6 (very good), with assignment instructions listed in Table 1. We also provided raters with examples of two types of interactions for each scoring, one with the mediation link and one without the mediation. We selected 40 Question-link-video triplets from our corpus of human generated linking dialogues. Each of the triple can also be presented as question-video i.e. without showing the mediation

#	Description
1	Response is not related in any way the question
2	Response contains some discussion of people or objects mentioned in the question, but does not really address the question itself.
3	Response partially addresses the question, but little or no coherence between the question and response.
4	Response does mostly address the question, but with major problems in the coherence between question and response; seems like the response is really addressing a different question than the one asked.
5	Response does address the question, but the transition is somewhat awkward.
6	Response answers the question in a perfectly fluent manner.

**Table 1.** Description of evaluation scale

link. We gathered coherence judgments from multiple raters, where no rater saw the same question twice and the presentation order was randomized. Each rater was shown 20 data points (10 Question-link-video triples and 10 Question-video pairs). The with-mediation condition was judged as significantly more coherent than the without-mediation condition, calculated using the Student T-test ( $t= 2.0322$ ,  $DF=118$ ,  $p<0.05$ ).

## AUTOMATED LINKING DIALOGUES

We implemented a prototype system for automatically generating linking dialogues based on our analysis of human-authored linking dialogues. This prototype was developed for use in the AXL system, replacing the original interview functionality for a single AXL character (named CSM Pullman).

After analyzing the corpus of human-authored linking dialogues and the interactions with the AXL system without mediation, we identified two sources of incoherence that we could address in this prototype. First, there is the case where there is no appropriate response available. Incoherence of this type could be addressed by a linking dialogue of the type *Coherence*, as described in the previous section. This type of link would add content directly queried by the user and then gently introduce the most relevant video clip afterwards. Second, there is the case where the response contains an appropriate answer, but it is not the most prominent part of the video clip and can be easily missed. Incoherence of this type could be mitigated with a *Focus/summarizing* link. This type of link helps the user pinpoint the answer to the question he/she asked within the available response. Our prototype system implements each of these two types of links.

To generate coherence-type links, where content outside the fixed responses clips is presented, the system must have some representation of the domain of the dialogue. Graesser et. al. [3] used a conceptual graph representation for question-answering systems for stories. For our prototype, we built on this idea by creating a conceptual graph representation of the fictional situation depicted in the AXL movie. To support the generation of focus/summarizing links, which bring specific parts of the video answer to focus, we created a representation of the content that is conveyed by each response video, and how this content relates to the conceptual graph representation of the movie.

The nodes of this conceptual graph representation consist of textual content that is used directly in generating the linking dialogue. Nodes that describe the content of available video clip responses from characters are called *video-nodes*, and those that describe the scenario content not included in the available responses (but necessary for coherence links) are called *fact-nodes*. Fact nodes were created by examining gaps in available answers to questions in the training corpus.

Links in the conceptual graph denote semantic relationships between the nodes. Since the graph's purpose is to support text generation, links are described as phrases that can link together the contents of two nodes. Thus the links in this graph are templates of phrases used for linking dialogues. The conceptual graph created to support mediation for one AXL interview consisted of 86 nodes (34 fact-nodes and 52 video nodes), and 330 links. Examples of nodes and links appear in Tables 2 and 3, respectively.

Nodes in the conceptual graph	Type
The way of dealing with the warlords is to clear it through the brigade.	Fact
CPT Young did not provide a clear plan for this mission to his subordinates.	Fact
Pullman tried to gather intel on Omar.	Video

Table 2. Examples of nodes in the conceptual graph

Node1	Link	Node2
CPT Young thinks he gave an appropriate level of guidance.	node1, specifically node2	CPT Young thinks he explained clearly to his XO what he wanted.
CPT Young thinks he gave an appropriate level of guidance.	node1, but node2	Lt Perez did not have enough guidance from the CO.
The way of dealing with the warlords is to clear it through the brigade.	node1 because, node2	Brigade CO wanted this food distribution operation to go smoothly without upsetting locals.

Table 3. Examples of links in the conceptual graph

Our strategy for creating coherent linking dialogues is to first select the node with the content that is closest to that of a user's question. For selecting this node, our system uses an implementation of multi-class support vector machine classifier [1], created using training data consisting of 450 user questions categorized into the 86 nodes. Once we have identified the most appropriate node for a user question, there are two possible ways of linking the question to an available answer. If the selected node is video-node, then a focus/summarizing link is generated, and if it is a fact-node, then a coherence link is generated.

**Focus link** when the selected node is a *video-node*, the generated link consists of the node description followed by simple speaker introduction link, e.g.:

*Question* : Has Pullman ever worked with the warlords before?  
*Link* : Pullman was highly experienced and was willing to offer guidance if needed. Let's hear from Pullman,  
*Response* : "I've had 25 years experience in this man's Army, and if I can bring that experience into play, I will. With anyone. If I saw Captain Young overreacting or hesitating or going off on a tangent, it's my duty to speak up. It's my job. And I execute my job well. But I'm not going to pull the OIC aside and tell him, 'Hey, pal, if you need some help, I'm your guy.' He ought to know that. If he doesn't, nothing I can whisper in his ear's going to change things."

**Coherence link** when the selected node is a fact-node, we first decide which response video to play. The candidates are all the video responses that are linked to this fact-node. The ordering among them is decided by a classifier trained to provide the most relevant video response to the question asked. The generated link contains the content of the selected node and content of the selected video clip, joined using the link template that connects them in the conceptual graph, followed by a speaker introduction link (visualized in Figure 1), e.g.:

*Question* : what is the way to deal with warlords?  
*Link* : The way of dealing with local population or warlords is to clear it through the brigade first, that's why Pullman tried to gather intelligence about Omar.

*Video* : "When Captain Young decided we needed to know more about Omar, I pulled out all the stops. There are always people you can push a little to get what you need. But, things can only happen so fast. I'm here. They're there. There's just so much barking you can do over the radio."

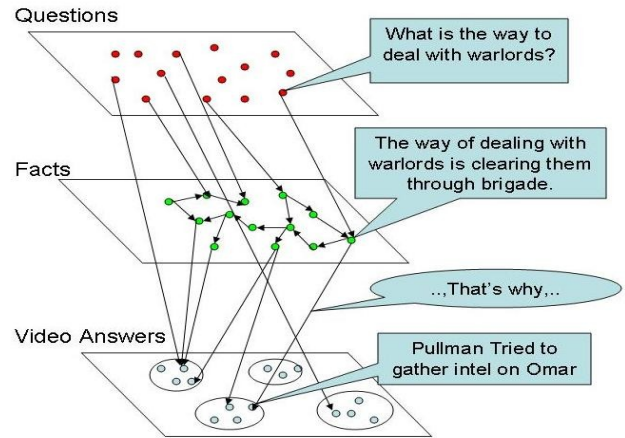


Figure 1. Example of generating a coherence-type link

## EVALUATION

To evaluate whether our implementation of mediated linking dialogues improves coherence of AXL interviews, we generated mediated and non-mediated responses using a test set of user questions. We then followed the same evaluation procedure as for the human-authored linking. To measure the improvement of coherence for given question we calculate the differences between coherence values for with and without mediation judgements (0.675 from the baseline system). This is a significant improvement under 2-way ANOVA, comparing individual questions and presence or absence of links ( $F=22.36, F_{crit}=3.90, p<0.05$ ).

For future work we are looking into improving the performance of the initial node selection step, as well as reducing the effort required to construct the conceptual graph.

## REFERENCES

- [1] Chih-Chung Chang, Chih-Jen Lin, LIBSVM: a library for support vector machines. 2001. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [2] Gandhe, S., Gordon A., Leuski A., Traum D., Oard D., First steps towards linking dialogues: mediating between free-text questions and pre-recorded video answers. *Army Science Conference*, 2004.
- [3] Graesser, A.C., Lang, K. L., and Roberts, R. M. Question answering in the context of stories. *Journal of Experimental Psychology: General*, (1991), 120(3).
- [4] Hill, R. J., Gordon, A., and Kim, J., Learning the lessons of leadership experience: Tools for interactive case method analysis. In *Proceedings of the Twenty-fourth Army Science Conference*, 2004.