Fully Automated Generation of Question-Answer Pairs for Scripted Virtual Instruction

Pascal Kuyten[1], Timothy Bickmore[2], Svetlana Stoyanchev[3], Paul Piwek[4], Helmut Prendinger[5], Mitsuru Ishizuka[1]

(1) Graduate School of Information Science & Technology
The University of Tokyo, Japan
pascal@mi.ci.i.u-tokyo.ac.jp
ishizuka@i.u-tokyo.ac.jp

(2) College of Computer and Information Science
Northeastern University, Boston, Massachusetts, USA
bickmore@ccs.neu.edu

(3) Spoken Language Processing Group,
Department of Computer Science
Columbia University, New York, USA
sstoyanchev@cs.columbia.edu

(4) NLG Group, Centre for Research in Computing
The Open University, Walton Hall, Milton Keynes, UK
p.piwek@open.ac.uk

(5) National Institute of Informatics
Tokyo, Japan
helmut@nii.ac.jp

Abstract. We introduce a novel approach for automatically generating a virtual instructor from textual input only. Our fully implemented system first analyzes the rhetorical structure of the input text and then creates various question-answer pairs using patterns. These patterns have been derived from correlations found between rhetorical structure of monologue texts and question-answer pairs in the corresponding dialogues. A selection of the candidate pairs is verbalized into a diverse collection of question-answer pairs. Finally the system compiles the collection of question-answer pairs into scripts for a virtual instructor. Our end-to-end system presents questions in pre-fixed order and the agent answers them. Our system was evaluated with a group of twenty-four subjects. The evaluation was conducted using three informed consent documents of clinical trials from the domain of colon cancer. Each of the documents was explained by a virtual instructor using 1) text, 2) text and agent monologue, and 3) text and agent performing question-answering. Results show that an agent explaining an informed consent document did not provide significantly better comprehension scores, but did score higher on satisfaction, compared to two control conditions.

Keywords: Dialogue Generation, Rhetorical Structure Theory, Medical Documents

adfa, p. 1, 2011.

1 Introduction and Motivation

Systems for the automatic generation of dialogue scripts have been used primarily to allow teams of computer-animated dialogue agents to present information to an audience [1-3]. In contrast, we use automatically generated dialogue scripts to drive the conversation between a user and a single virtual agent. Our aim is to evaluate this mode of presentation (following up on [4], which evaluated the use of dialogue script generation for presentation by non-interactive teams of agents).

We propose a system which is capable of creating virtual instruction from textual input only, extending previous work [1], into fully automated generation of agent animation scripts from text. In this section we will use text (as in **Table 1**) from informed consent documents for clinical trials [23] to illustrate the system. First text is translated into rhetorical structure theory (RST) trees (as in **Fig. 1**), by annotating discourse relations using high-level discourse analysis. RST trees are then translated into question-answer pairs (as in **Table 1**), by matching patters on the relations and structure of RST trees. Answers are compiled into an animated virtual instructor, using animation scripts. Users are asked to read the question; click an ask-button; and watch the animation (See **Fig. 2** for a screenshot of the virtual instructor answering a question).

The paper is organized as follows. This Section continues with an introduction to the theory of text organization. In Section 2, we describe related work; Section 3 is dedicated to our system design, In Section 4 we discuss some design considerations, in Section 5 we describe our evaluation study. In Section 6 we discuss future work and Section 7 contains the conclusions.

Theory of text organization. Text can be segmented into non-overlapping, semantically independent units (EDUs) [11]. Between EDUs rhetorical (discourse) relations describe how the more important part (nucleus) and less important part (satellite) relate (e.g. CONTRAST). Text organization can be represented using rhetorical structure theory (RST) trees (as in **Fig. 1**), leaves in RST trees represent EDUs, arrows in the RST tree point from satellite to nucleus, and arrows are labeled with a discourse relation.

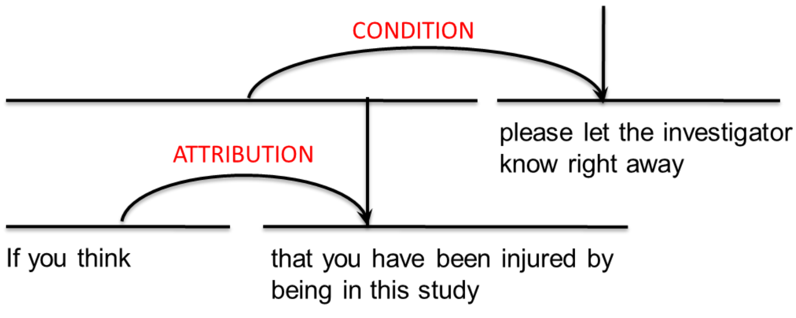

Fig. 1. RST tree, representing the rhetorical structure of text, leaves represent elementary discourse units (EDUs), arrows point from satellite to nucleus, and labels above arrows represents discourse relations.

Table 1. Text from an informed consent document for clinical trials [23] and the corresponding question-answer pair generated by our system

<table>
<tr><td rowspan="2">Text
If you think that you have been injured by being in this study, please let the investigator know right away.</td><td>Question
What if I think that I have been injured by being in this study?</td></tr>
<tr><td>Answer
Please let the investigator know right away.</td></tr>
</table>

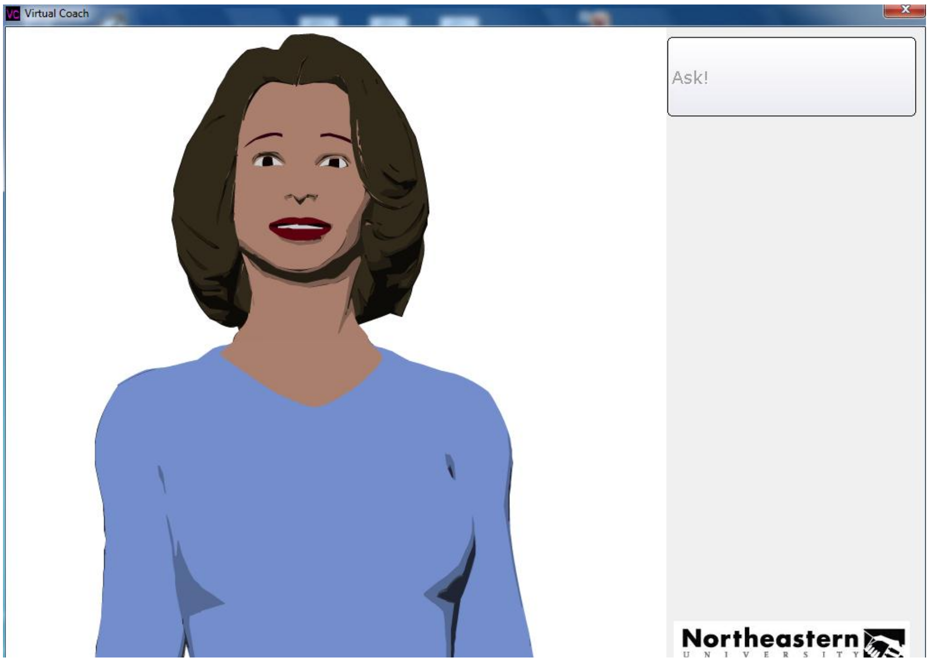

Fig. 2. Screenshot of our virtual instructor answering a question.

2 Related Work

The system designed at university of Pennsylvania [5] is similar to our work in that both aim to generate questions from text, using rhetorical analysis. While they use semantic role labeling for analyzing the meaning of the text, our approach is based on support vector machine classifiers for analyzing the discourse structure of the text [6]. When considering question generation at paragraph level the discourse structure of the text becomes important [10].

The aim of the tutor in the project LISTEN is to improve reading comprehension of children [7]. Although both works aim at improving comprehension of the text, their approach is applying semantic role labeling [5] for generating questions instead of discourse analysis and dialogue generation. Further, their generated questions are used as a tool for classification of children self-questioning responses, whereas our generated question-answer pairs are used as input for the virtual instructor.

Cloze question generation is based on syntactical analysis [8], and takes a similar approach as our work. Trees are constructed, patterns are matched and questions are generated. Different from our work, questions are generated by identifying definition phrases. A part of these phrases are replaced with answer-blanks. Users are asked to fill in the answer-blanks by choosing from the removed answer phrase and distractors. Whereas our system is aiming at automatically generating virtual instruction, cloze question generation is aiming at helping second and third grade students to learn new vocabulary.

The twins Ada and Grace are two virtual characters guiding visitors at the museum of Science in Boston [9, 24]. While in our system users get questions presented, in their work visitors can ask the twins questions. Questions asked by visitors are mapped to nearest known questions from a knowledge base containing question-answer pairs. Answers belonging to found questions are presented by the twins. Question-answer pairs from this knowledge base are acquired by a question-answer generator called Question Transducer [10]. Question Transducer identifies factual questions from text by matching patters on the syntactical structure of sentences or paragraphs in the text. Unlike the question-answer pairs of the Question Transducer, our question-answer pairs go beyond paragraph boundaries and can cover larger spans of text (up to the entire text).

A prototype which aims at providing authors of medical texts feedback about their writing style links two systems G-DEE and Greta using XSLT transformations [25]. G-DEE is a document analysis tool capable of automatically detecting importance of recommendation in clinical guidelines uses shallow natural language processing techniques. And Greta, an agent platform supporting detailed non-verbal expressions linked to a TTS.

3 System Design

Our system (illustrated by **Fig. 3**) generates RST trees from text using high-level discourse analysis. Based on this analysis, question-answer pairs are generated, by translating the RST tree into coherent dialogue. Question-answer pairs are then translated into an agent scripting language. In the final step, scripts are compiled into a run-time agent system (See **Fig. 2** for a screenshot of our system).

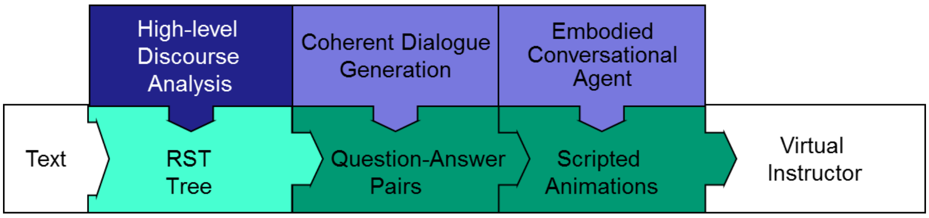

Fig. 3. Setup of the system which generates a virtual instructor based on text, fully automated.

Data between each module is sequenced using XML-files. Besides some minor annotation of the input text, the overall process is fully automated. Text is annotated for guidance of EDU segmentation during the high-level discourse analysis. Annotation of bulleted lists is manual; annotation of sentence- and paragraph-boundaries is scripted.

High-level Discourse Analysis. RST trees are generated by the system using a high-level discourse analyzer, called HILDA [6]. The discourse analyzer first segments text into EDUs. Then, (typically) binary discourse relations are identified between EDUs. HILDA is using three classifiers: 1) for EDU segmentation, 2) for discourse labeling and 3) for RST tree construction. HILDA first segments text into EDUs (illustrated by **Fig. 4**), and then constructs an RST tree (illustrated by **Fig. 1**). RST trees are constructed in an iterative process: in each step the two most likely adjacent RST sub-trees or EDUs are merged into a new RST sub-tree and labeled with the most likely discourse relation (illustrated by **Fig. 5**).

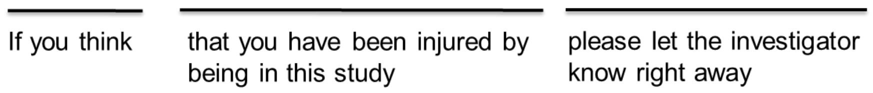

Fig. 4. HILDA segments text into EDUs

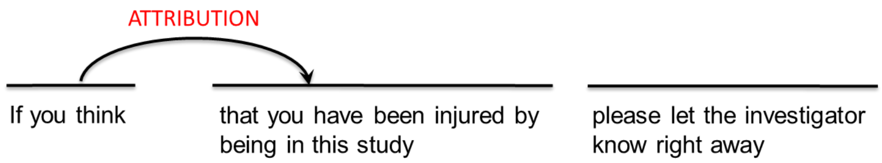

Fig. 5. HILDA merges the most likely adjacent RST sub-trees or EDUs into a new RST sub-tree with the most likely label.

Coherent Dialogue Generation. For mapping from RST structure to a dialogue script we use the approach developed in the CODA project [12]. In CODA, a parallel corpus of annotated monologues and dialogues was constructed, where the dialogues express the same information as the aligned monologues. From this, a mapping was inferred from RST structures in monologue to the dialogue act sequences in dialogue. These mapping are used by the CODA system to map an RST tree (such as the one in **Fig. 1**) to a sequence of dialogue acts (as in **Table 1**). The input for the CODA system is a sequence of one-level RST trees. It maps this to alternative (ranked) sequences of dialogue acts, and verbalizes the top-ranked sequence. The final output is an XML representation of a dialogue act sequence (usually consisting mostly of question-answer pairs).

Embodied Conversational Agent. The user interface for explaining the document to users was based on an embodied conversational agent system developed for health counseling [13]. In this system, dialogue between a single agent and a user is scripted using a custom hierarchical transition network-based scripting language. Agent non-verbal conversational behavior is generated using BEAT [14], and includes beat (baton) hand gestures and eyebrow raises for emphasis, gaze away behavior for signaling turn-taking, and posture shifts to mark topic boundaries, synchronized with synthesized speech. User input is obtained via multiple choice selections of utterances. The system automatically translates XML representation of question-answer pairs into the agent scripting language for compilation into the run-time system.

4 Design considerations

Question-answer pairs of our system go beyond paragraph boundaries and can cover larger spans of text (up to the entire text). HILDA generates a single RST tree for the entire text, CODA then maps at various depths discourse relations in this RST tree to a sequence of dialogue acts. If CODA maps a discourse relation at the root of an RST tree, then the question-answer pairs of these dialogue acts cross paragraph boundaries.

A previously conducted case study indicated structural differences between RST trees generated by HILDA and RST trees used for deriving the rule-base of CODA [15]. Some tail EDUs of sentences were merged with the heads of adjacent sentences, causing misalignments in the RST tree. Some discourse relations in the rule-base of CODA were not identified by HILDA. Changes were made to the initial design and configuration of HILDA and CODA, in order to reduce these differences.

Table 2. Question-answer pairs generated by HILDA

<table>
<tr><td colspan="2">Misaligned question-answer pair, based on the traditional implementation of HILDA</td></tr>
<tr><td>Question
What if I think that I have been injured by being in this study?</td><td>Answer
Please let the investigator know right away. If your part in this study takes place at Bohemia Medical Center.</td></tr>
<tr><td colspan="2">Aligned question answer-pairs, based on the proposed implementation of HILDA</td></tr>
<tr><td>Questions
What if I think that I have been injured by being in this study?

What if my part in this study takes place at Bohemia Medical Center?</td><td>Answers
Please let the investigator know right away.

You can get treatment for the injury at Bohemia Medical Center.</td></tr>
</table>

Effect of RST structure on questions-answer pairs. One of the classifiers of HILDA responsible for the structure of RST trees has been trained with features considering RST sub-trees with a maximum span of three EDUs [6]. Because some sentences in text are segmented into more than three EDUs, we expect some of the structural differences identified [15], are caused by the span limitations of the classifier. Take for example an EDU continuing the text (of **Table 1**): “If your part in this study takes place at Bohemia Medical Center”, here HILDA has several options to construct an RST tree. Traditionally HILDA merges the last EDU of the first sentence with the first EDU of the second sentence (illustrated by **Fig. 6**). Alternatively HILDA could merge EDUs of the first sentence with EDUs of the second sentence (illustrated by **Fig. 7**). CODA generates different question-answer pairs based on the two RST trees (listed in **Table 2**), where the RST tree of the traditional version induces a misalignment. In order to prevent such misalignment we propose a two phase discourse analysis by first merging EDUs within sentences and afterwards merging RST sub-trees.

Effect of discourse relations on patterns of CODA's rule base. Not all discourse relations of CODA's rule base can be identified by HILDA. In order to increase the number of rules which CODA can match on RST trees generated by HILDA, we created new rules for CODA's rule base. When all subclasses of a superclass were listed in the rule-base, the superclass was added as well. For example the rule Explain_Init-Complex-InfoReq_Explain matches, among others, on Elaboration-Additional and Elaboration-Obj-Attribute. We extended this with Elaboration, which can be identified by HILDA.

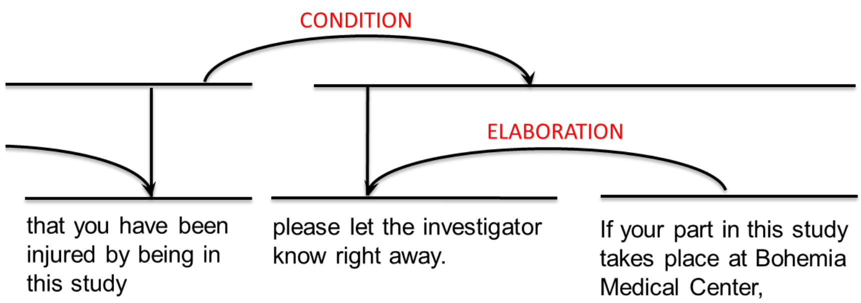

Fig. 6. Merging the last EDU of the first sentence with the first EDU of the second sentence

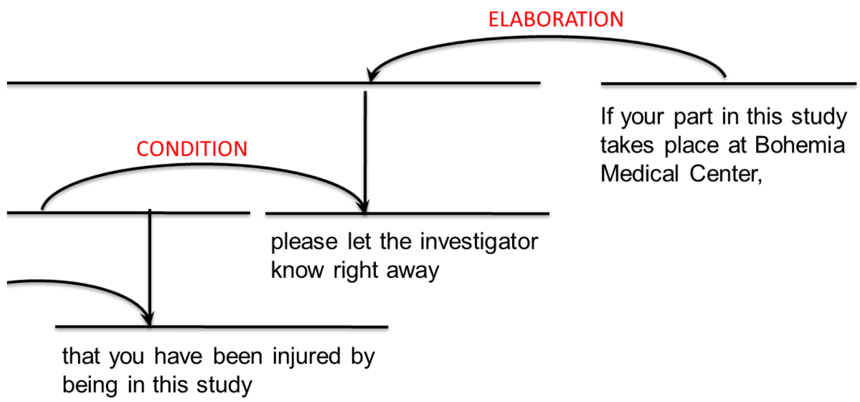

Fig. 7. Merging EDUs of the first sentence with EDUs of the second sentence

5 Evaluation

We conducted an evaluation study to test the effectiveness of our agent-based question-asking system at augmenting explanations of complex text documents. We hypothesized that if a user conducts a question-asking dialogue with an agent about a text, in addition to reading the text, that they will be more cognitively engaged in the material, understand more about it, and be more satisfied with the experience, compared to simply reading the text by itself.

To test this hypothesis, we conducted a 3-arm, counterbalanced, within-subjects experimental study, comparing the question-asking agent (QA) to reading the text (TEXT) and, thirdly a control condition in which the agent read the text (READ), intended to control for exposure time with the agent and hearing the document contents through multiple modalities (text and speech).

The task domain for the experiment is the explanation of research informed consent documents for colonoscopy clinical trials. This domain was selected because the documents contain a wide range of medical and legal terms, facts and concepts that provide a good test for an automated explanation system. Administration of informed consent for clinical trials is often completed without ensuring that participants understand all the terms of the consent agreement, resulting in many potential research subjects signing consent forms that they do not understand [16-18]. In addition, there has been prior work demonstrating some success at having virtual agents explain clinical trial informed consent documents [19, 20]. Colonoscopy is an important area to address: colon cancer is the second leading cause of cancer-related deaths (60,000 deaths each year in the US), and colon screenings have been proven to reduce colon cancer deaths up to 90%, yet compliance with medical recommendations for colonoscopy and other screening is very low. We created three research informed consent documents for this study by taking descriptions of colonoscopy clinical trials [23], adding standard language about research informed consent (from [21] and other sources), and ensuring that the length and complexity was approximately the same across all three.

Our primary hypotheses for the study are:

H1: Users will understand more about documents in the QA condition compared to the TEXT and READ conditions.

H2: Users will be most satisfied with the informed consent process in the QA condition compared to the TEXT and READ conditions.

Measures. Comprehension was assessed by a closed-book knowledge test, consisting of three YES/NO questions (e.g., “Will you be able to choose which of the four bowel preparation medications you will use?”), and three multiple choice questions (e.g., “What risk is associated with ingestion of iodinated oral contrast?”) for each document. Satisfaction was assessed using several single-item, scale response self-report questions, based on the Brief Informed Consent Evaluation Protocol (BICEP) [17], including likelihood to sign the consent document, overall satisfaction with the consent process, and perceived pressure to sign the consent document (**Table 3**). We also asked single-item scale response questions about satisfaction with the agent, desire to continue working with the agent, and the amount of information provided (from "too little" to "too much").

Table 3. Scale Self Report Measures Used

Measure	Question	Anchor 1	Anchor 2
Satisfaction with Agent	How satisfied are you with the instructor?	Not at all	Very satisfied
Desire to Continue with Agent	How much would you like to continue working with the instructor?	Not at all	Very much
Satisfaction with Experience	How satisfied were you?	Extremely unsatisfied	Extremely satisfied
Amount of Information Provided	How much information did you get?	Too little	Too much
Pressure to Sign	How much pressure did you feel?	No pressure	Extreme pressure
Likely to Sign	How likely would you have been to sign it?	Extremely unlikely	Extremely likely

Participants. A convenience sample of twenty-four subjects, 29% female, aged 28-36, participated in the study. Participants were mostly students (58%), well educated (all had some college), and had high levels of computer literacy (58% described themselves as being "experts").

Experimental Protocol. Verbal informed consent was obtained from study participants, after which they completed a brief demographic questionnaire and were randomized into one of six study sequences defining the order of conditions. We randomized the order in which the study conditions were experienced by each participant while holding the order of presentation of the three documents constant, to counterbalance both order effects and the effects of any particular informed consent document. Participants next completed three rounds of document explanation and filling out comprehension and satisfaction questionnaires. Finally, a semi-structured interview was held with them about their experience and preferences among the three conditions, before they were paid and dismissed.

The study was conducted on a standard desktop computer using a mouse and keyboard for input, and all questionnaires were administered via web forms on the same computer. The entire study was conducted within the Embodied Conversational Agent application interface described in Section 3.3. All agent utterances were accompanied by conversational nonverbal behavior generated using BEAT [22].

In the TEXT condition, the agent walked on the screen and said "Hi, I am Karen. I am going to explain an informed consent document to you for a clinical trial." After the user clicked "OK, let's get started!", the first page of the document filled the screen, and the user was allowed to read it until they clicked a “I’m through reading this.” button, at which point the next page of the document was displayed. When the last page of the document had been read, a message was displayed on the screen informing the participant that the session was over.

The READ condition was identical to TEXT, except that after each page of the document was displayed, the agent re-appeared and read the page to the participant in an uninterruptable monologue.

The QA condition was also identical to TEXT, except that after each page of the document was displayed, the agent re-appeared and engaged the user in a question-and-answer dialogue, as generated by the system described in Section 3. Question-and-answer pairs were delivered in sequence. For each, the question was displayed in text on the screen and the user could push an “Ask!” button, after which the agent re-appeared and delivered the answer.

Quantitative Results. We conducted repeated-measures ANOVAs for all self-report measures, knowledge test scores, and session duration, in SPSS. **Table 1** shows descriptive statistics for the outcome measures.

Table 4. Study Results (mean and (SD))

	TEXT	READ	QA	p
Session Duration (seconds)	505 (251)	1081 (249)	1011 (247)	<.001
Comprehension	77% (21%)	69% (28%)	76% (22%)	n.s.
Satisfaction with Agent	3.83 (1.88)	3.96 (1.69)	4.35 (1.75)	n.s.
Desire to Continue with Agent	3.70 (1.82)	3.73 (1.83)	4.30 (1.94)	n.s.
Satisfaction with Experience	4.09 (1.47)	3.83 (1.83)	4.39 (1.92)	n.s.
Amount of Information Provided	4.35 (1.27)	3.38 (1.34)	3.96 (1.07)	.07
Pressure to Sign	2.35 (1.23)	2.52 (1.65)	2.22 (1.28)	n.s.
Likely to Sign	4.13 (1.58)	3.78 (1.70)	4.22 (1.81)	n.s.

There was a significant effect of condition on session duration, with TEXT taking significantly less time $F(2,22)=31.7$, $p<.001$. There was a trend for participants to rate the TEXT condition as providing too much information compared to either the READ or QA conditions, $F(2,44)=2.80$, $p=.07$. No other significant differences were found, although the various satisfaction measures were all trending with QA rated more highly than the other conditions.

The only significant order effect found was that the Amount of Information Provided was rated as increasing session by session, $F(2,42)=3.32$, $p<.05$. This could be due to actual differences in the information content of the three documents, or effects of user fatigue. There were a few significant differences by gender, with females more satisfied with the agent compared to males.

Qualitative Results. Semi-structured interviews with 23 participants were transcribed and coded for common themes. When asked for their overall impressions, several participants volunteered that they liked the concept of an agent explaining a document:

- "It's easier to remember if presented as conversation."
- "The avatar helped me to concentrate."
- "A great way to remember."

Although several others felt uncomfortable with the system:

- "I'm uncomfortable to be explained by an animated character/avatar. I would prefer a human being."
- "I prefer to read instead of listening."

Many participants also volunteered that they liked the question-asking interaction:

- "By asking questions, I am able to get info you need without unnecessary information."
- "The questions did help."
- "Enjoyed question answering, although weird."

Many participants had suggestions for improving the question-asking interaction with the agent. One of the most common suggestions was to make it more interactive, by allowing users to select their questions from a menu:

- "There is no additional values of having an animated character/avatar if there is not much interactivity with avatar, because the avatar was just reading/repeating the content which I read already and the questions were preset."
- "I want to decide my own rhythm."

Other suggestions included grouping the questions by relevance, and displaying or including a question when giving the answer, to provide better context.

When asked which of the three explanation methods was most informative, 15 (65%) expressed a clear preference for the QA condition, with an additional 2 indicating a tie between QA and READ:

- "The [QA] conversation is easier to remember due to highlights."
- "It had interaction, which I liked."
- "Because she was answering questions instead of just reading."
- "[QA] had the best presentation, and [READ] had the best content."

When asked which of the three methods they would like to actually use for informed consent, 12 participants expressed a preference for QA, and an additional three said that either QA or another method would be alright.

- "When I read it I may not know what is important. The questions highlight to me what is important."
- "I felt less pressure to understand the document."
- "I got a little bit lost with the questions, some I would not ask, therefore I lost track."

Discussion. We found generally positive acceptance of the question-asking system by participants. We did not find any support for H1, regarding improved document comprehension with QA compared to the control conditions. However, we did find partial support for H2, with a majority of participants interviewed stating that the QA condition was the most informative compared to the other presentation methods, and all quantitative satisfaction measures trending with QA as the most preferred (although the differences are not significant). This result mirrors results from a previous study in which an agent explaining an informed consent document did not provide significantly better comprehension scores, but did score higher on satisfaction, compared to two control conditions [19].

Lack of support for H1 could be attributed to several factors. Participants were mostly students, who are used to receive information from documents. Further, we presented all questions in pre-fixed order, which may have led to lower engagement. The lower comprehension results for the READ condition, when compared to the TEXT condition, could be attributed to the quality of the TTS or presentation style, e.g. while reading the text to our participants we did not provide any subtitles.

6 Future work

More specific relations. Not all discourse relations of CODA's rule-base can be identified by HILDA, and not all discourse relations identified by HILDA exist in CODA's rule-base. Therefor we are planning to study whether it is possible to improve HILDA's performance for specific domains, in particular for CODA's domain. We are planning to study whether we could improve the output of the overall system, when HILDA has been trained to identify a subset of CODA's discourse relations. Improved output could be measured in terms of more diverse question-answer pairs or increased quality of the question-answer pairs.

Variable length of question-answer pairs. Question-answer pairs generated by our system vary in length, because CODA maps discourse relations at various depths in the RST tree. When discourse relations are matched at the root of RST trees, the generated question-answer pairs will be long. And when discourse relations are matched near the leaves of the RST tree, the generated question-answer pairs will be short. Besides, when RST trees are unbalanced, questions will differ in length from answers. A couple participants of our evaluation study noted the variable sizes of the question-answer pairs and stated they preferred shorter answers. The preferable length of the question and answer may depend on the application of the system. Therefore, we are planning to study whether it is possible to guide the RST tree construction of HILDA, as well as the pattern matching of CODA in order to generate questions and answers of specific length.

Improved comprehension and satisfaction. There are several aspects we can explore to improve comprehension and satisfaction of our users. In a future evaluation study, we could investigate different presentation styles. We could allow users to select questions of their interest and let them intervene during the answering. We could also present the question-answer pairs as a dialogue between two agents. Finally, we could use a different TTS and presentation style of answering questions. We could highlight important aspects of the document while answering a question or add subtitles when the instructor is answering a question.

7 Conclusions

We introduced a novel approach for automatically generating a virtual instructor from textual input only. We described the system design, consisting of high-level discourse analysis, coherent dialogue generation and embodied conversational agent scripting. Furthermore, we discussed some design considerations in order to reduce structural differences found in a previous case study. Finally we conducted an evaluation study to test the effectiveness of our agent-based question-asking system at augmenting explanations of complex text documents. Results show that an agent explaining an informed consent document did not provide significantly better comprehension scores, but did score higher on satisfaction, compared to two control conditions.

Acknowledgements. This work was initiated while Dr. Svetlana Stoyanchev was a visitor of NII and was extended while Pascal Kuyten was a visitor of OU. The experiment was conducted while Dr. Tim Bickmore was a visitor of NII. This work was supported, in part, by a grant from the US National Cancer Institute (NIH grant R01 CA158219), by the NII Grant Challenge Project and the Japanese Government Scholarship program (文部科学省奨学金).

References

1. Piwek, P., Hernault, H., Prendinger, H., Ishizuka, M. T2D: Generating Dialogues Between Virtual Agents Automatically from Text. In: IVA 2007. LNCS, vol. 4722, pp. 161-174, Spinger, Heidelberg (2007), doi: 10.1007/978-3-540-74997-4_16
2. André, E., Rist, T., van Mulken, S., Klesen, M., Baldes, S. The Automated Design of Believable Dialogues for Animated Presentation Teams. In: Embodied Conversational Agents, MIT Presss, Cambridge, 2000
3. Van Deemter, K. Krenn, B., Piwek, P., Klesen, M., Schroeder, M., Baumann, S. Fully Generated Scripted Dialogue for Embodied Agents. In: Artificial Intelligence Journal, vol. 172, no. 10, pp. 1219-1244 (2008)
4. Stoyanchev, S., Piwek, P., Prendinger, H. Comparing Modes of Information Presentation: Text versus ECA and Single versus Two ECAs. In: IVA 2011. LNCS, vol. 6895, pp. 377-383, Springer, Heidelberg (2011), doi: 10.1007/978-3-642-23974-8_41
5. Mannem, P., Prasad, R.P., Joshi, A. Question Generation from Paragraphs at UPenn. In: Proc. of QG2010: The Third Workshop on Question Generation, pp. 84-91, Pitsburg, PA (2010)
6. Hernault, H., Prendinger, H., duVerle, D. and Ishizuka, M. HILDA: A Discourse Parser Using Support Vector Machine Classification. In: Dialogue and Discourse, vol. 1, no. 3, pp. 1-33 (2010)

7. Chen, W., Mostow, J. Aist, G. Using Automatic Question Generation to Evaluate Questions Generated by Children. In: Question Generation: Papers from the 2011 AAAI Fall Symposium (2011)
8. Gates, D., Aist, G., Mostow. J., McKeown. M., Bey, J. How to Generate Cloze Questions from Definitions: A Syntactic Approach. In: Question Generation: Papers from the 2011 AAAI Fall Symposium (2011)
9. Nouri, E., Artstein, R., Leuski, A., Traum, D. Augmenting Conversational Characters with Generated Question-Answer Pairs. In: Question Generation: Papers from the 2011 AAAI Fall Symposium (2011)
10. Heilman, M. Automatic Factual Question Generation from Text. In: Ph.D. Dissertation, Carnegie Mellon University, CMU-LTI-11-004CMU-LTI-09-013 (2011)
11. Mann, W.C., Thompson, S.A. Rhetorical structure theory: Toward a functional theory of text organization. In: Text, vol. 8, no. 3, pp. 243-281 (1988)
12. Piwek, P., Stoyanchev, S. Data-oriented monologue-to-dialogue generation. In: Proc. Of 14th Annual Meeting of ACL, short papers, pp. 242-247, Portland, Oregon (2011)
13. Bickmore, T., Picard, R.: Establishing and Maintaining Long-Term Human-Computer Relationships. ACM Transactions on Computer Human Interaction 12, 293-327 (2005)
14. Cassell, J., Vilhjálmsson, H., Bickmore, T.: BEAT: . In: Conference BEAT: The Behavior Expression Animation Toolkit, pp. 477-486. (2001)
15. Kuyten, P., Hernault, H., Prendinger, H., Ishizuka, M. Evaluating HILDA in the CODA Project: A Case Study in Question Generation Using Automatic Discourse Analysis. In: Question Generation: Papers from the 2011 AAAI Fall Symposium (2011)
16. Wogalter, M.S., Howe, J.E., Sifuentes, A.H., Luginbuhl, J.: On the adequacy of legal documents: factors that influence informed consent. Ergonomics 42, 593-613 (1999)
17. Sugarman, J., Lavori, P.W., Boeger, M., Cain, C., Edson, R., Morrison, V., Yeh, S.S.: Evaluating the quality of informed consent. Clinical Trials 2, 34 (2005)
18. Sudore, R., Landefeld, C., Williams, B., Barnes, D., Lindquist, K., Schillinger, D.: *Use of a modified informed consent process among vulnerable patients: a descriptive study. J Gen Intern Med 21, 867-873 (2006)
19. Bickmore, T., Pfeifer, L., Paasche-Orlow, M.: Using Computer Agents to Explain Medical Documents to Patients with Low Health Literacy. Patient Education and Counseling 75, 315-320 (2009)
20. Fernando, R.: Automated Explanation of Research Informed Consent by Embodied Conversational Agents. College of Computer and Information Science, vol. MS. Northeastern University, Boston (2009)
21. http://ccr.coriell.org/Sections/Support/NIGMS/Model.aspx?PgId=216
22. Cassell, J., Vilhjálmsson, H., Bickmore, T.: BEAT: The Behavior Expression Animation Toolkit. In: Conference BEAT: The Behavior Expression Animation Toolkit, pp. 477-486. (2001)
23. http://nci.nih.gov/clinicaltrials
24. Yao, X., Tosch, E., Chen, G., Nouri, E., Artstein, R., Leuski, A., Sagae, K., Traum, D.: Creating conversational characters using question generation tools. In: Dialogue and Discourse, vol. 3, no. 2, pp. 125-146 (2012)
25. Georg, G., Cavazza, M., Pelachaud, C.: Visualizing the Importance of Medical Recommendations with Conversational Agents. In: IVA 2008. LNCS, vol. 5208, pp. 380-393, Spinger, Heidelberg (2008), doi: 10.1007/978-3-540-85483-8_39