Automatic Generation of Multi-Modal Dialogue from Text Based on Discourse Structure Analysis

Helmut Prendinger
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku
Tokyo 101-8430, Japan
helmut@nii.ac.jp

Paul Piwek
Centre for Research in Computing
The Open University
Walton Hall, Milton Keynes MK7 6AA, UK
p.piwek@open.ac.uk

Mitsuru Ishizuka
Graduate School of Information Science and Technology
University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
ishizuka@i.u-tokyo.ac.jp

Abstract

In this paper, we propose a novel method for generating engaging multi-modal content automatically from text. Rhetorical Structure Theory (RST) is used to decompose text into discourse units and to identify rhetorical discourse relations between them. Rhetorical relations are then mapped to question–answer pairs in an information preserving way, i.e., the original text and the resulting dialogue convey essentially the same meaning. Finally, the dialogue is "acted out" by two virtual agents. The network of dialogue structures automatically built up during this process, called DialogueNet*, can be reused for other purposes, such as personalization or question–answering.*

1 Introduction

There is a huge demand for content that is easy-to-access, easy-to-understand, tailored to people's needs and background knowledge, trustworthy, and presented in an attractive, engaging, user-friendly way. However, the effort to create such content is often tremendous, and current technologies offer only insufficient or partial solutions [1]. Therefore, the problem this paper aims to address is the development of an effective technology to automatically create high-quality content, so that anyone can produce professional multi-modal content easily.

The importance of and need for our proposed technology can be best seen in the areas of E-healthcare and E-learning. E-healthcare is an important future application of our method. The vision includes that elderly persons or patients can receive professional care over the internet [15]. Currently, however, little progress is made towards creating content that is tailored to caretakers' needs. As pointed out in [30], medical information is difficult to convey to patients because of its complexity, inherent emotional sensitivity, and the wide use of technical terms.

The work described in [21] pioneered the use of dialogue scripts to present (medical) information in a more intuitive way. The authors suggest to transform e.g. the sentence "*To take a tablet, you should first remove it from the foil and then swallow it with water.*" (from a pill prescription) into a dialogue (script) of the form:

Patient:	How should I take the tablet?
Pharmacist:	Remove it from the foil. Then swallow it with water.

Dialogue scripts offer new opportunities to present information in a more natural and engaging way. However, the dialogues presented in [21] had to be created manually. Instead, our proposed technology will provide a means to generate dialogues automatically. E-healthcare will serve as an application domain. E-learning is another application domain that would benefit from the proposed technology. Here, textbook knowledge could be transformed into a dialogue between an investigative student asking questions, and an expert teacher providing the answers.

An obvious alternative to presenting a written dialogue script is to assign dialogue contributions to virtual agents that can perform (or "act out") the script. Recently, life-like characters (or virtual interface agents) have gained increas-

ing popularity as an effective means to present content in an easy-to-understand and also enjoyable way [24]. Relying on their anthropomorphic appearance, these character agents can deliver multi-modal contents naturally through speech, gestures, and emotional expression. Currently, however, multi-modal content is still sparse, because of the significant preparatory effort required for the scripting and animation of character agents, as well as the provision of content in appropriate formats. In order to support multi-modal content creation, we already developed the Multimodal Presentation Markup Language (MPML), which allows anyone to control and synchronize multi-modal contents with little effort [13]. The usability of MPML has been demonstrated in a variety of environments, including web browser, mobile phone, and the physical world (using a humanoid robot). Our latest version, MPML3D, is able to control highly realistic three-dimensional (3D) agents [20].

Our method produces a knowledge structure called *DialogueNet*, which is automatically created from the rhetorical structure of text. The use of DialogueNet is three-fold:

1. *Presentation*: DialogueNet is converted to MPML3D and acted out by virtual characters.
2. *Re-Generation*: DialogueNet is transformed or regenerated for particular purposes, such as personalization, summarization (dialogue compression), emphasis (dialogue amplification), etc.
3. *Semantic Authoring*: DialogueNet provides the basis for a wiki-type semantic authoring interface [12], and question–answering.

In this paper, we will focus on the automatic generation of DialogueNet and its presentation by virtual agents. The next section discusses related work on multi-modal dialogue generation and the use of rhetorical relations in semantic authoring. Section 3 is the core of the paper. First, a brief introduction to Rhetorical Structure Theory (RST) is provided. Next, we explain the formal underpinnings of the mapping from rhetorical discourse relations to question–answer pairs. Finally, an example of a multi-modal dialogue using virtual agents is presented. Section 4 briefly discusses dialogue personalization and integration of ontologies. Conclusions are drawn in Section 5.

2 Related Work

There are a number of studies that deal with the problem of automatically generating multi-modal dialogues between life-like animated agents. However, they differ in the type of input they require and the techniques that are employed to map the input to multi-modal dialogue (see Section 2.1). In Section 2.2, we turn to work on semantic authoring using discourse relations that is related to our approach.

2.1 Automatic Dialogue Generation

In Intelligent Multimedia Presentation (IMMP) systems the authoring process is automated by employing methods from artificial intelligence, knowledge representation, and planning (see [1] for an overview). An IMMP system assumes a so-called "presentation goal" and uses planning methods to generate a sequence of presentation acts. The generation of a presentation is based on dedicated information sources that encode information about presentation content and objects [2]. The difference to our approach is that we do not require the formulation of planning operators, which assumes a background in artificial intelligence. Our approach is solely based on existing material (currently text and, in future also associated graphics), and thus easy-to-use by non-experts and not suffering from the knowledge representation bottleneck.

Recently developed related systems include Web2TV and Web2Talkshow [19], and e-Hon [28]. Web2TV uses two animated characters to readout a given text in a TV-style environment. Web2Talkshow transforms a (summary) of text from the web into a humorous dialogue between character agents. e-Hon transforms text into an easy-to-understand dialogue based on rephrasing content, and enriching it with animations.

Web2Talkshow and e-Hon on the one hand, and our system on the other, are similar in that they both aim to generate dialogues automatically from text. The differences lie in the details of the mapping from text to dialogue. First, Web2Talkshow and e-Hon analyze the *information structure* of single sentences as the basis of the generated dialogue, rather than the *discourse structure* (rhetorical structure) of text content as the basis of the dialogue. This gives rise to differences both in the way individual sentences are dealt with and at the level of larger spans of text. Information structure relates to theme/rheme [9] or topic/comment [10] distinctions within sentences, whereas our approach can also reflect how textual units (e.g. sentences in text) relate to each other in a meaningful way at the discourse level rather than on the sentence level. Second, whereas our aim is to faithfully render the content of the input text as a dialogue, a feature of Web2Talkshow is that it generates humorous dialogues, exploiting distortions and exaggerations of what is actually said in the input text.

The "text-to-presentation" system [26] generates summary slides from a given text automatically. Like ours, this approach is based on the analysis of the discourse structure of text, and similar to [21], the dependency between text units is reflected by (slide) layout, such as itemizing and indentation. Due to space constraints on the presentation slides, [26] prune text by leaving out non-essential parts of a sentence. Our approach, on the other hand, aims to preserve the information of the input text.

The investigations on generating scripted dialogues described in [23, 21] provided some of the foundations for the current work. Those researches also investigated the combination of information from sources other than text. In one scenario [22], the principal information is an electronic health record, and supplementary information is drawn from thesauri, wikis, and ontologies.

2.2 Semantic Discourse Knowledge Base

Semantic authoring is the activity of annotating content in a structured way, typically based on some ontology [11]. The *Semantic Authoring* interface suggested in [12] uses a standardized set of rhetorical relationships,[1] or ontology, to connect sentences and phrases in a graphical representation. Applications include text composition, discussion-supporting groupware, collaborative authoring, and so on.

The knowledge representation format of *Semantic Authoring* [12] and our DialogueNet are comparable, as both are based on rhetorical relationships. However, DialogueNet focuses on the discourse relationships between dialogue contributions rather than sentences. Certainly, the most crucial difference is that DialogueNet is automatically generated from text, whereas the network of rhetorical relations in *Semantic Authoring* is hand-crafted. In fact, one of the functions of system described in [12] is to generate (monological) text from the semantic discourse representation. Whereas text in DialogueNet is the input, the *Semantic Authoring* system has it as output.

3 From Text to Multi-Modal Dialogue

Our system implements a modular architecture that follows an input-to-output pipeline approach. It is written in Java and uses XML format to exchange representations between modules. An overview of the pipelined architecture is shown in Figure 1.

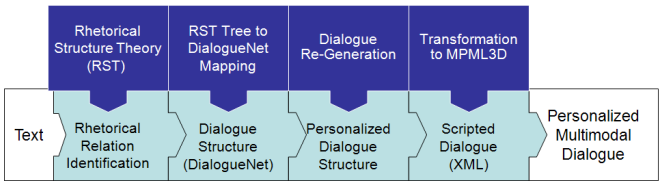

Figure 1. Text-to-dialogue pipeline.

3.1 Rhetorical Structure Theory

As input, the system accepts text, e.g. from the web (possibly including pictures). We are using sentences about medical pill prescriptions from the PIL ("Patient Information Leaflets") corpus, comprising 465 leaflets.[2]

The first step is to extract the discourse structure from text. We apply Rhetorical Structure Theory (RST), a descriptive theory of text organization [17]. RST allows us

- to segment text into non-overlapping, semantically independent units (i.e. clauses or sentences), and
- to identify rhetorical relations between text segments, which indicate functional relations between them, such as MOTIVATION, ELABORATION, CONTRAST, CONDITION, CONCESSION, SEQUENCE, MEANS, QUESTION-ANSWER, etc.

Here are example sentences from the PIL corpus.

> *Do not take Klaricid tablets if you are allergic to clarithromycin or other macrolide antibiotics such as erythromycin or azithromycin. If you have any liver or kidney problems consult your doctor before taking these tablets. Klaricid does not interact with oral contraceptives.*

Text segmentation produces the units shown in Table 1.

Table 1. Discourse units of input text.

[Do not take Klaricid tablets]$_a$
[You are allergic to clarithromycin or other macrolide antibiotics such as erythromycin or azithromycin.]$_b$
[You have any liver or kidney problems]$_c$
[Consult your doctor before taking these tablets.]$_d$
[Klaricid does not interact with oral contraceptives.]$_e$

Let us look at the discourse units a and b. RST suggests that "*Do not take Klaricid tablets*" is a consequence of the fulfillment of the condition "*You are allergic to clarithromycin or other macrolide antibiotics such as erythromycin or azithromycin*" [4]. Furthermore, RST determines that "*Klaricid does not interact with oral contraceptives*" presents additional information to the information presented in units a–d.

RST does not only identify rhetorical relations within text, it also determines which text information is more essential, called the 'nucleus' of the relation, and which information is secondary, called the 'satellite' of the relation. In c–d, for instance, the core information is captured by the information that one should consult the doctor before taking the Klaricid tablets, whereas the conditional part ("*You*

[1] ISO/TC37/SC4/TDG3

[2] http://mcs.open.ac.uk/nlg/old_projects/pills/corpus/PIL/

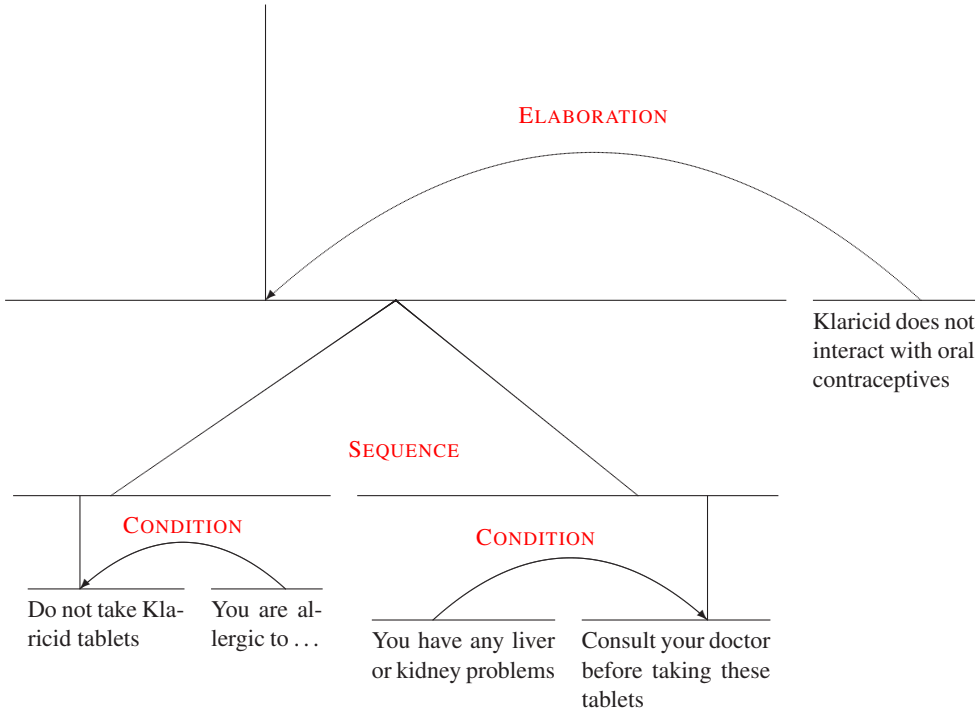

Figure 2. RST tree.

have any liver or kidney problems") is regarded as subordinate information with respect to the nucleus.

The result is displayed in Figure 2. Arrows denote links from the satellite to the nucleus of the rhetorical relationship. The names associated with arrows stand for the relationship that holds between the discourse units. Horizontal lines identify text spans, and vertical lines indicate nuclei. Tilted lines connect multiple nuclei of multi-nuclear relations such as SEQUENCE. (SEQUENCE simply states temporal order and is often used when no other, more specific relation is applicable.)

In order to generate RST trees automatically, we employ the Discourse Analysis System (DAS) [14], a state-of-the-art syntactical parsing system based on cue phrases and various textual coherence structures. An alternative machine learning based approach relying on an ensemble of support vector classifiers is presented in [25]. The detection of rhetorical relations from text is a non-trivial and error-prone task, even for human annotators. Reasonably accurate results have been achieved for the simplified task of sentence-level (rather than text-level) discourse parsing [27], or for a restricted set of rhetorical relations. E.g., [18] achieve high accuracy when the set of discourse relations is confined to only four relations (CONTRAST, EXPLANATION-EVIDENCE, CONDITION, ELABORATION). Systematic studies regarding the robustness of the DAS discourse parser are underway.

3.2 Building the DialogueNet

In order to generate a dialogue structure (DialogueNet) automatically from the discourse structure of text, we implemented "abstract mappings" from rhetorical relations to question–answer pairs, the elementary building blocks of a dialogue. A core requirement for the mapping is that it is information preserving, i.e., both the declarative sentence and the question–answer pair should convey the (essentially) same message. Our method, which is based on lambda abstraction, will be introduced in the next section. Thereafter, we will show some examples of mapping rules.

3.2.1 Information Preserving Question Formation

As a starting point for the explanation of question formation, take the following RST structure (bold face indicates the nucleus)

$$\text{CONDITION}(\mathbf{P}, \text{Q}) \quad (1)$$

which can represent CONDITION("*Consult your doctor before taking these tablets*", "*You have any liver or kidney problems*"), as in Figure 2. A conceivable question–answer pair carrying the same information (or message) as (1) is

$$\text{QUESTION-ANSWER}(\text{Under what circumstances } \mathbf{P}?, \text{Q}) \quad (2)$$

It is important to note that QUESTION-ANSWER is just another rhetorical relationship. In [4, p. 64], it is defined as "In

a question–answer relation, one textual span poses a question (not necessarily realized as an interrogative sentence), and the other text span answers the question."

However, when inspecting (1) and (2), it is not obvious why they should be equivalent in terms of conveying the same information. We address this issue by using a device from mathematical logic, called λ-abstraction [6], which allows us to explicate the fact that (2) has an underlying CONDITION relation.

Hence, instead of (2), we use the representation

$$\text{QUESTION-ANSWER}(\lambda x.\ \text{CONDITION}(\mathbf{P},\ x),\ \text{Q}) \qquad (3)$$

where "Under what circumstances **P**?" is replaced by a formal representation, λx.CONDITION(**P**, x). The open variable x indicates that we are dealing with a question. With this method, a question can be analyzed as λ-abstraction over one of the two arguments of the CONDITION relation.

An operation related to λ-abstraction is function application. If we apply a lambda expression $(\lambda x.M)$ to another expression N, the result is defined as follows: $(\lambda x.M)N \mapsto M[x := N]$. Our formal interpretation of the QUESTION-ANSWER relation is application, and consequently, λx.CONDITION(**P**, x)Q can be related to CONDITION(**P**, Q). In this way, the information equivalence between (1) and (3) can be demonstrated. Relating declarative sentences to question–answer pairs based on abstraction and application was independently proposed by other researchers [3].

We have seen how λ-abstraction allows us to state information equivalence on RST structures. Abstraction also provides a generic tool for generating question–answer pairs from declarative sentences. Question formation over some subexpression E of S can be expressed by the general formula

$$S \longmapsto \lambda x.S'\ E,$$

where $S' = S[E := x]$. By this formula, different types of question–answer pairs can be generated by abstracting over different parts of the input. Here is a list of examples.

- *Question formation over the first argument of a relation:* Input: $S = R(P, Q)$. Output: $\lambda x.R(x, Q)P$.

  Example: see above.

- *Question formation over the second argument:* Input: $S = R(P, Q)$. Output: $\lambda x.R(P, x)Q$.

  Example: If you have any liver or kidney problems consult your doctor before taking these tablets. $\mapsto$ What if I have any liver or kidney problems? Then consult the doctor before taking these tablets.

- *Question formation over a subexpression of a simple proposition:* Input: S, whereby $E \sqsubseteq S$. Output: $\lambda x.S'E$, with $S' = S[E := x]$.

  Example: Klaricid does not interact with oral contraceptives. $\mapsto$ With what does Klaricid not interact? With oral contraceptives.

Observe that the last mentioned type of question formation covers the approach described in [19], which is based on information rather than on discourse structure.

3.2.2 Mapping Rules

Having described the general method of mapping an RST tree to an equivalent RST tree (DialogueNet), we now turn to some examples of mapping rules. Specifically, we will discuss mapping rules for the CONDITION and ELABORATION (rhetorical) relations (see Figure 3).

(A) Mapping rules for CONDITION

(i) Nucleus in Imperative Form

CONDITION(**P**, Q) & imperative(**P**) $\Longrightarrow$
Layman: Under what circumstances should I **P***?
Expert: If Q.

CONDITION(**P**, Q) & neg-imperative(**P**) $\Longrightarrow$
Layman: Under what circumstances should I not **P***?
Expert: If Q.

(ii) Nucleus in Declarative Form with Modal Auxiliary

CONDITION(**P**, Q) & declarative-modal-aux(**P**) $\Longrightarrow$
Layman: Under what circumstances *flip*(**P***)?
Expert: If Q.

P* is **P**[I:=you, you:=I, my:=your, your:=my, mine:=yours, yours:=mine]; *flip(X)* is a function that performs the "interrogative flip" inversing subject and auxiliary

(iii) Example of alternate mapping

CONDITION(**P**, Q) $\Longrightarrow$
Layman: What if Q*?
Expert: Then **P**.

Q* is Q[I:=you, you:=I, my:=your, your:=my, mine:=yours, yours:=mine]

(B) Mapping rule for ELABORATION

ELABORATION(**P**, Q) $\Longrightarrow$
Expert: **P**.
Expert: Should I tell you more?
Layman: Yes, please.
Expert: Q.

Figure 3. Mapping rules for discourse relations "condition" and "elaboration".

We inspected conditionals occurring in the PIL corpus (4212 instances), and derived the rules depending on the syntactic realization of their nucleus. Machinese Syntax

[16] was used to parse discourse units at the syntax level. Let us look at an example of a conditional (from Figure 2) where the nucleus is in negative imperative form. Here, CONDITION(**P**, Q) is instantiated to CONDITION("*Do not take Klaricid tablets*", "*You are allergic to clarithromycin or other macrolide antibiotics such as erythromycin or azithromycin*"). Application of the second mapping rule of (A.i) in Figure 3 yields:

Layman: Under what circumstances should I not take Klaricid tablets?
Expert: If you are allergic to clarithromycin or other macrolide antibiotics such as erythromycin or azithromycin.

An example of a conditional with declarative nucleus containing a modal auxiliary ("should") is "*It should not produce any undesirable effects if you (or somebody) accidentally swallows the cream*". The dialogue resulting from application of rule (A.ii) in Figure 3 is:

Layman: Under what circumstances should it not produce any undesirable effects?
Expert: If you (or somebody) accidentally swallows the cream.

Since applying the same templates for some relation over and over will produce a monotonous dialogue, alternate mappings are provided to introduce *variation* into dialogue scripts. (A.iii) in Figure 3 is one example of an alternate mapping rules for the conditional. Applied to the previous example, we obtain the following alternate dialogue:

Layman: What if I (or somebody) accidentally swallows the cream?
Expert: Then it should not produce any undesirable effects.

An example for the ELABORATION relation will be given in the next section. We have already implemented mapping rules for the most frequently occurring relations, and extend the set continuously.

3.3 Multi-Modal Dialogue

The DialogueNet corresponding to some input text can be "acted out" through the performance of 3D virtual agents. Our agents were created by a professional Japanese character designer for 'digital idols' and are controlled by MPML3D [20]. They are endowed with the following features: (i) Conversational and iconic gesture behavior (around thirty gestures); (ii) facial emotion expression (happy, sad, surprised); and (iii) synthetic voice with proper lip-synchronization.

The agents are shown in Figure 4. The role of the male agent is 'expert' (e.g. pharmacist), and the female agent performs the role of the 'layman' (e.g. a patient). The 'speech balloons' are displayed only for convenience, and are not used in our implemented system.

Figure 4. Dialogue between virtual agents.

Below is the (whole) dialogue that results from the sample text presented in Section 3.1, utilizing the intermediate RST tree shown in Figure 2.

(1) Layman: Under what circumstances should I not
(2) take Klaricid tablets?
(3) Expert: If you are allergic to clarithromycin or
(4) other macrolide antibiotics such as
(5) erythromycin or azithromycin.
(6) Layman: What if I have any liver or kidney
(7) problems?
(8) Expert: Then consult your doctor before taking
(9) these tablets.
(10) Expert: Should I tell you more?
(11) Layman: Yes, please.
(12) Expert: Klaricid does not interact with oral
(13) contraceptives.

Our system can produce this dialogue script and have it performed by our life-like agents. We believe that the multi-modal presentation of (medical) information in dialogue format can benefit users (patients) in various ways. For example, [7] found that dialogue presentations can be a more effective means for communicating information than monologue as it enables vicarious learning. [30] argue that a dialogue may support the patients' preparation of asking

medical questions related to their case. The preparation includes learning the adequate use of medical terms.

However, even in our sample dialogue based on the PIL corpus, many technical terms occur that might not be easily understood by most patients. Therefore, in the next section, we will discuss *personalization* as one possible improvements of the dialogues generated by our system.

4 Personalizing Dialogues

A personalized dialogue is tailored to the context or skill level of the user. (On tailoring medical information, see e.g. [8]). For instance, few users will understand the medical term "clarithromycin". In order to accommodate for a user's level of expertise in medical terminology, we might simply replace the technical term by a more common term, such as "antibiotic drug". This approach is adopted in [28], where stories are adapted to the knowledge level of children. Another approach is to *re-generate* the existing dialogue script by inserting an explanatory (sub)dialogue. Taking our example (Section 3.3), we could insert the following explanatory (sub)dialogue after line (5).

Layman: What is clarithromycin?
Expert: An antibiotic drug used to treat infection.

Observe that the (sub)dialogue draws information from an external source, i.e., the information about clarithromycin is not contained in the PIL corpus. In this case, we took the definition from a medical ontology, the Unified Medical Language System (UMLS) [29].

Dialogue re-generation by adding (sub)dialogues derived from external information providers is a form of information 'merging'. In order to preserve the quality (trustworthiness, reliability) of the dialogue, the (re-generated) DialogueNet is decorated with labels $S_1, \ldots, S_n$ that indicate the information source. If the reliability of the some source S_i is questionable or should be made explicit for some reasons, it will be reflected by the prefix "According to S_i, [...]" in the dialogue.

Another type of dialogue personalization relates to replacing some nouns by nouns that name a specific person or location. In line (8) of the dialogue above, "your doctor" can be replaced by a proper noun denoting an existing doctor. Similarly, the description "your nearest hospital" can be resolved by an existing location.

5 Conclusions

In this paper, we described a new technique for generating engaging multi-modal contents automatically – assuming just (monological) text as input. Our implemented system first extracts the discourse (rhetorical) structure of text, and then maps the resulting RST tree into a corresponding dialogue structure, the DialogueNet, by applying information preserving operations. Finally, the dialogue contributions are assigned to animated agents that use multiple modalities (speech, gesture, facial expression) to "act out" the dialogue script.

By watching the layman agent ask questions to an expert agent and hearing the answers, the agent performance can support users' vicarious learning [7]. Our conversational life-like agents are designed to deliver information in a more user-friendly and engaging way. We focussed on the E-health domain (medical pill prescriptions), but in principle, our method applies to any textual information.

In this paper, one core motivation was to propose a method that allows anyone to create professional multimodal contents with little effort. That is, even non-experts (in content creation) should be able to produce attractive content easily from available sources. We want to emphasize that the automatically generated DialogueNet may serve for other purposes as well. By applying dedicated regeneration operations, dialogue scripts can be personalized, summarized, or amplified, in order to adapt the dialogue to the user's information need and context. Among others, dialogue scripts will be re-generated by 'merging' them with already available ontologies.

The systematic treatment of dialogue re-generation operations is left for future research. We are also interested in investigating the usability of DialogueNet as a basis for semantic authoring [12]. We envision that the visualization of DialogueNet can serve as an interface for the editing of additional answer nodes to given question nodes, thereby facilitating collaborative knowledge creation.

Furthermore, we want to explore in what way DialogueNet can contribute to advanced question–answering systems. Research on "Context Question Answering" [5] argues that questions are hardly asked in isolation. Instead, users ask a series of related questions about some topic. DialogueNet might provide a suitable representation for the question context, which is given by the set discourse roles in a question and discourse transitions between questions. We will report on our efforts to exploit the many ways in which a discourse structure based dialogue knowledge representation can be applied in the future.

Acknowledgements

We would like to thank Hugo Hernault for his contribution to the implementation. The research was supported by the Research Grant (FY1999–FY2003) for the Future Program of the Japan Society for the Promotion of Science (JSPS), by a JSPS Encouragement of Young Scientists Grant (FY2005–FY2007), an NII Joint Research Grant with the Univ. of Tokyo (FY2006), and a Memorandum of Understanding with the Open Univ., UK.

References

[1] E. André. The generation of multimedia presentations. In R. Dale, H. Moisl, and H. Somers, editors, *Handbook of Natural Language Processing*, pages 305–327. Marcel Dekker, Inc, 2000.

[2] E. André, T. Rist, S. van Mulken, M. Klesen, and S. Baldes. The automated design of believable dialogue for animated presentation teams. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 220–255. The MIT Press, Cambridge, MA, 2000.

[3] R. Bäuerle and T. Zimmermann. Fragesätze (Interrogation sentences). In A. von Stechow and D. Wunderlich, editors, *Semantics. An International Handbook of Contemporary Research*, pages 333–348. Mouton de Gruyter, Berlin/New York, 1991.

[4] L. Carlson and D. Marcu. Discourse tagging reference manual. Technical Report ISI-TR-545, ISI, September 2001.

[5] J. Y. Chai and R. Jin. Discourse structure for context question answering. In *Proceedings of HLT-NAACL 2004 Workshop on Pragmatics in Question Answering*, pages 23–30. ACL, 2004.

[6] A. Church. *The Calculi of Lambda-Conversion*. Princeton University Press, Princeton, 1941.

[7] R. Cox, J. McKendree, R. Tobin, J. Lee, and T. Mayes. Vicarious learning from dialogue and discourse: A controlled comparison. *Instructional Science*, 27:431–458, 1999.

[8] N. Elhadad, M.-Y. Kan, J. Klavans, and K. McKeown. Customization in a unified framework for summarizing medical literature. *Artificial Intelligence in Medicine*, 33(2):179–198, 2005.

[9] J. Firbas. On the interplay of prosodic and non-prosodic means of functional sentence perspective. In V. Fried, editor, *The Prague School of Linguistics and Language Teaching*, pages 77–94. Oxford University Press, London, 1972.

[10] J. Gundel. Stress, pronominalization and the given-new distinction. In *University of Hawaii Working Papers in Linguistics 10/2*, pages 1–13, 1978.

[11] S. Handschuh, S. Staab, and A. Maedche. CREAM – creating relational metadata with a component-based, ontology-driven annotation framework. In *Proceedings of 1st International Conference on Knowledge Capture (K-Cap 2001)*, pages 76–83. ACM Press, 2001.

[12] K. Hasida. Semantic authoring and semantic computing. In *New Frontiers in Artificial Intelligence: Joint Proceedings of the 17th and 18th Annual Conferences of the Japanese Society for Artificial Intelligence*. Springer, 2005.

[13] M. Ishizuka and H. Prendinger. Describing and generating multimodal contents featuring affective lifelike agents with MPML. *New Generation Computing*, 24:97–128, 2006.

[14] H. T. Le and G. Abeysinghe. A study to improve the efficiency of a discourse parsing system. In *Proceedings 4th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-03)*, LNCS 2588, pages 101–114. Springer, 2003.

[15] C. Lisetti, F. Nasoz, C. LeRouge, O. Ozyer, and K. Alvarez. Developing multimodal intelligent affective interfaces for tele-home health care. *International Journal of Human-Computer Studies*, 59(1–2):245–255, 2003.

[16] Machinese Syntax, 2006. URL: `http://www.connexor.com/`.

[17] W. C. Mann and S. A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281, 1988.

[18] D. Marcu and A. Echihabi. An unsupervised approach to recognizing discourse relations. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics (ACL-02)*, pages 368–375, 2002.

[19] A. Nadamoto and K. Tanaka. Complementing your TV-viewing by web content automatically-transformed into TV-program-type content. In *Proceedings 13th Annual ACM International Conference on Multimedia*, pages 41–50. ACM Press, 2005.

[20] M. Nischt, H. Prendinger, E. André, and M. Ishizuka. MPML3D: a reactive framework for the Multimodal Presentation Markup Language. In *Proceedings 6th International Conference on Intelligent Virtual Agents (IVA-06)*, Springer LNAI 4133, pages 218–229, 2006.

[21] P. Piwek, R. Power, D. Scott, and K. van Deemter. Generating multimedia presentations. From plain text to screenplay. In O. Stock and M. Zancanaro, editors, *Multimodal Intelligent Information Presentation*, Text, Speech, and Language Technology, pages 203–225. Springer, 2005.

[22] P. Piwek, R. Power, and S. Williams. Generating scripts for personalized medical dialogues for patients. Technical Report 2006/06, Department of Computing, Faculty of Mathematics and Computing, The Open University, UK, 2006.

[23] P. Piwek and K. van Deemter. Towards automated generation of scripted dialogue: some time-honoured strategies. In *Proceedings 6th Workshop on the Semantics and Pragmatics of Dialogue (EDILOG-02)*, pages 141–148, 2002.

[24] H. Prendinger and M. Ishizuka, editors. *Life-Like Characters. Tools, Affective Functions, and Applications*. Cognitive Technologies. Springer Verlag, Berlin Heidelberg, 2004.

[25] D. Reitter. Simple signals for complex rhetorics: On rhetorical analysis with rich-feature support vector models. volume 18, pages 38–52, 2003.

[26] T. Shibata and S. Kurohashi. Automatic slide generation based on discourse structure analysis. In *Proceedings 2nd International Joint Conference on Natural Language Processing (IJCNLP-05)*, pages 754–766. Springer LNAI 3651, 2005.

[27] R. Soricut and D. Marcu. Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of HLT-NAACL 2003*, pages 149–156, 2003.

[28] K. Sumi and K. Tanaka. Transforming E-contents into a storybook world with animations and dialogues using semantic tags. In *Online Proceedings of WWW-05 Workshop on the Semantic Computing Initiative (SeC-05)*, 2005. URL: `http://www.instsec.org/2005ws/`.

[29] Unified Medical Language System, 2007. URL: `http://www.nlm.nih.gov/research/umls/`.

[30] S. Williams, P. Piwek, and R. Power. Generating monologue and dialogue to present personalised medical information to patients. In *Proceedings 11th European Workshop on Natural Language Generation (ENLG-07)*, pages 167–170, 2007.