



Story Envisioning Framework for Visualized Collective Storytelling in Conversation

Qiang Zhang^{1,2}(✉), Maryam Sadat Mirzaei^{1,2}(✉), Hung-Hsuan Huang²(✉),
and Toyoaki Nishida^{1,2}(✉)

¹ Graduate School of Informatics, Kyoto University, Kyoto, Japan

zhang.qiang.84n@st.kyoto-u.ac.jp,

nishida.toyoaki.4e@kyoto-u.ac.jp

² RIKEN-AIP, Kyoto, Japan

{maryam.mirzaei,hung-hsuan.huang}@riken.jp

Abstract. In this paper, we introduce the method of story envisioning as an efficient approach to collective storytelling in daily conversation and training procedure. The goal of story envisioning is to provide intuitive, concrete and consistent visualization contents as interactive drama for storytellers who are not professional at storytelling. We employ graphic recording to symbolize and visualize the stories written in natural language. We built a partially automated framework TSEiA that allows storytellers to visualize and edit the stories. By traversing through the whole story structure, storytellers are also able to compare and analyze different consequences of branches. Preliminary experiments showed that using visualized interactive drama can help participants have better understanding to the conversation from culture and personality viewpoint. Simultaneously, story envisioning could help the storytellers organize the story structure in a better way and enhance the immersion of their stories. Game platform powered by TSEiA was built to illustrate how story envisioning could be considered as an effective tool for collective storytelling and content generation.

Keywords: Storytelling · Story envisioning · Conversation analysis

1 Introduction

Conversation plays an important role in sharing information, knowledge, thoughts and emotions in human society. Conversation allows people to transmit implied thoughts and experience in their mind and share them with others based on culture and knowledge background [1]. Considered as part of highest-level intelligence of human, conversation skill is frequently studied in the filed of AI. Nishida, etc looked at conversation from five viewpoints: *verbal communication* as central element, *nonverbal communication* as emotion indicators, *social*

discourse as social implications, *narratives and content* as content structure and *cognitive process* as mental processing [2].

From these viewpoints, conversation can be considered as a process of exchanging social signals, contents as story pieces which can be represented as components of the larger story [2]. This property of conversation leads to the irregularities of order and duration which eventually constructs the complicatedly hierarchical discourses structure of conversation.

Serving as an inevitable part of conversational and cultural activities, storytelling is an essential method of sharing experience and knowledge. Storytelling entails the method that mankind organize and package their experience and personal memories as structural stories in daily conversation to help others have a better understanding of the story contents [1]. The method of storytelling has been applied into areas such as education, training, media study, language learning, etc.

In the original idea of storytelling, it also describes the procedure of multiple narrators with different culture background to generate the contents collaboratively. From the studies by Jefferson and Sacks, storytelling in conversation are *locally occasioned, recipient designed* [4] and *co-constructed* [5]. More researches have demonstrated that in conversational storytelling, collaboration among multiple participants plays an important role in content generation. Listeners collaborate with storytellers in substantial ways to exert influence on storytellers [6]. In this procedure, listeners may also provide their own experience and knowledge and add them to the original story. The update of storytelling in daily conversation helps us to create a shared space. This is frequently used to build training system which can train high-level skills containing cultural elements.

However, embedded in human evolution history, storytelling usually demands rich cultural and literal knowledge, which brings difficulties to tell stories. In conversation, inappropriate expressing methods, ambiguous meanings and hierarchical discourses usually complicate the procedures of storytelling. Background knowledge of the story are required to be provided when a third party is willing to join the conversation [7]. These problems are approached by using a method of story visualization. Story is rebuilt in a virtual world in which resources such as characters, animations and objects are used to represent the story contents. Text-to-scene systems inspired the method to generate a static visualized scenario from natural language. Simultaneously, collective storytelling and authoring tools are used to create game-based language training courses. However, using the method of visualized collective storytelling to create and analyze daily conversation remains as blank.

In our research, we claim that story envisioning can serve as both a visualization tool which can assist people to understand conversation better and a collective storytelling tool which can be used for rapid content generation. Story envisioning focuses on facilitating the contents in common ground as interactive drama by treating the conversation as an update of common ground [8]. Rather than directly translating stories into visualized scenarios, story envisioning put more emphasis on rebuilding the whole structure of a conversation to give inspi-

rations on analyzing the high-level interpretations. By creating branches for each interpretation, story envisioning is able to create an visualized shared community in which different people can perceive others' interpretations and leave comments or modifications.

2 Related Work

There have been several systems developed that apply visualized storytelling and collective storytelling. Text-to-scene systems contributed to the initial stage of visualized storytelling. WordsEye [9] represented the approach to creating 3D static scenes by extracting semantic intent of the users. CarSim [10,11] is a program that generates short simple animations from natural language car accident reports. CONFUCIUS [12] is a multimodal systems that takes sentences containing verb as input. More researches by Chang, etc. showed that spatial relationships and objects types in different scenes can be learned from given text knowledge [13,14].

Other researches have focused on developing storytelling systems using various method to represent story. Cassell, etc. created dialogue planner based on topic coherence relations to study user trust on virtual agents [16]. Interactive storytelling system "Say Anything" described the method of providing interactive experience by acquiring large-scale knowledge [17]. Our previous work focusing on a bargaining scenario created storytelling environment in virtual reality to study how misunderstanding are brought to the surface due to culture difference [7]. On top of storytelling, earlier works on collective storytelling and collaborative authoring tools were used in Tactical Iraqi [15], a serious game for learning language and cultural elements. It has been used by a lot of trainees in US military. Recently, visualized collective storytelling and collaborative authoring are frequently embedded in games.

3 Framework TSEiA

To address the goal of story envisioning, we developed system TSEiA (The Story Envisioning Intelligent Assistant). TSEiA provides convenient approaches to story visualization, content management, story traversing and collective storytelling. The framework is designed to solve the problem of information dispersion in daily conversation. The method of using action sequences to represent the story reduced the effort to create repository database. Moreover, experiment shows that using action sequences to represent the story could retain most of the information.

3.1 Framework Overview

TSEiA is a system mainly built on Unity3D game engine. Together with a natural language processing module and a word embedding module, the system can convert stories in natural language into an animated 3D interactive drama. Functionally, the system contains three modules: *Story Visualization Module*, *Story Editing Module* and *Collective Storytelling Module* (Fig. 1).

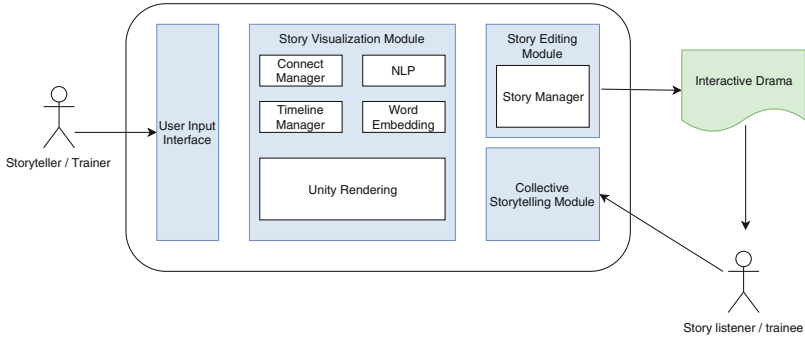


Fig. 1. Architecture of TSEiA

3.2 Story Visualization Module

This module provides services to symbolize and visualize the original story. Story visualization module accepts stories as input and outputs the visualization interactive drama.

Storytelling commonly represents stories as connected events which reveals the information flow [3]. However, on one hand, in daily conversation, the story pieces are highly dispersed throughout the whole conversation. As a result, events are connected inconspicuously and can hardly represent the information flow. On the other hand, from the experience of previous text-to-scene/animation systems, using event to represent story always bring difficulties and limitations for visualization.

In our research, we found that the content of storytelling in daily conversation tends to be more lifelike. Comparing to traditional storytelling, less fictional and legendary elements are being used in storytelling in daily conversation. Instead, storytelling in daily conversation traces a piece of past personal experience, memory or adventure. Individuals put more emphasis on the activities of each subject in their stories. We introduced the method of *action sequence* in which each action represents one activity of a certain object and the whole sequence represents the activity trajectory of this object to stand for stories in daily conversation (Fig. 2).

In TSEiA, original stories are processed by a NLP processor containing rules to extract verbs as “actions”. TimelineManager (Fig. 1) dynamically initialize a Unity TimelineAsset according to the action sequences. Assets are loaded from repository database in which resources that are needed to create scenes are downloaded and stored. In order to reduce the effort to prepare for the database, we use word embedding to categorize actions into a smaller space. With the actions labeled by the word embedding module, the rendering engine can visualize the scenario (Fig. 3).

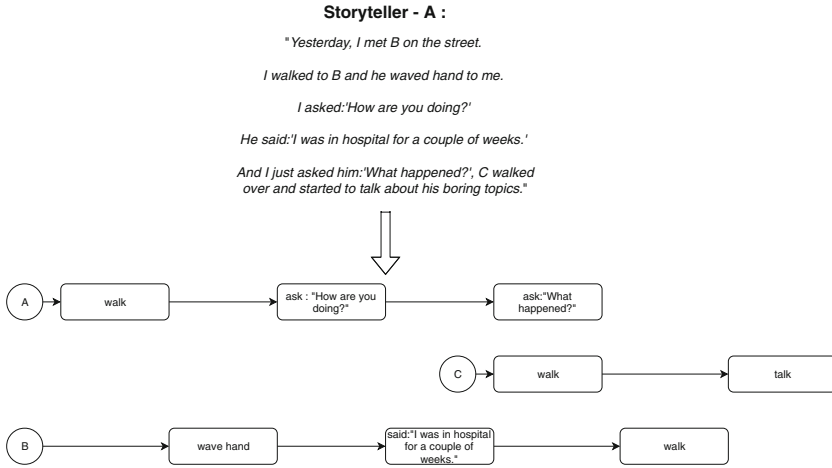


Fig. 2. Action sequence

3.3 Story Editing Module

This module can support user to edit and arrange the story branches. The intelligent story editing module allows storytellers to add, delete, embed, and compare story branches. Each branch contains the visualization of a piece of story. Story-Manager (Fig. 1) is managing a story tree to organize and index the story content created by storytellers. This module also provides the function to traverse through each branch, from which storytellers can inspect mistakes and modify the content (Fig. 3). By traversing through different branches, storytellers are able to locate most important verbal or non-verbal signals that can change the direction of the conversation. TSEiA also provides API for storytellers to edit stories dynamically.

3.4 Collective Storytelling Module

As one of our main contribution, this module allows storytellers to accomplish visualized collective storytelling. All the users are considered as both storytellers and story readers. A typical collaboration flow may comply with the following order: (1) A storyteller creates the domain and topic of a conversation and gives some initial visualized branches. (2) The contents those are generated by this storyteller is synchronized through cloud. (3) Other storytellers read these content and give their modifications. (4) New contents are synchronized through cloud. The contents those are generated by all the storytellers are being stored as file and can be shared inside community. Moreover, learning from the data collected from this module, we are able to build the predictive model that can envision some possible branches.



Fig. 3. An Example UI for Story Editing. This UI contains story input area, branch editing area. Storytellers can use “←” and “→” on the keyboard to traverse through the branch.

4 Experiment

In order to show that story envisioning can be an effective method for visualized collective storytelling, an experiment was conducted to answer the following questions:

- Q1. Can story envisioning help people have a better understanding of the stories in daily conversation?
- Q2. Can story envisioning be helpful to create content for collective storytelling?

We built a game platform powered by TSEiA and conducted experiment with 18 participants. We used a suspense role-playing game platform with the stories that involved multiple characters and human daily relationship.

4.1 Pilot Experiment

Before the real experiment, a pilot experiment was conducted to ensure that our system meets the basic functionality requirements. The pilot experiment contains two phases. In the first phase, some short animated dramas generated by TSEiA are presented to the participants. For each drama, there were five choices containing relevant text to it. The participants need to select the choice from which the drama was generated. In the second phase, participants are requested to create their own drama clips limited to a given repository. By reviewing the final output, participants need to evaluate how much can the system meet their expectations. Participants are also requested to finish the questionnaire after the experiment. Positive results are obtained from the questionnaire.

4.2 Main Experiment

The goal of this experiment is to evaluate how much story envisioning can contribute to content comprehension of storytelling in daily conversation as well as to collective storytelling.

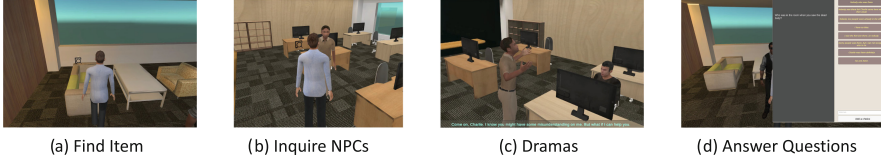


Fig. 4. Experiment Setting. (a) Participant explore the environment to find evidence items. (b) Participant get information by communicating with the NPC. (c) Drama are presented to the participant. (d) Participant converse to the policeman and answer the question.

Experiment Setting. In this experiment, participants are required to play a suspense role-playing game. Participants are divided into 6 groups where each group contains 3 participants. Participants in the same group independently start the game session at same time. Participant plays the role of a chemist (main character) who is working in a medicine company. The main character’s boss died in the office. After investigation, the policeman found out that the boss was poisoned by someone and the poison was actually a new medicine developed by the main character. Other three characters (agents) with unique background are also involved as suspects. Participants are provided with part of background information in text format. In this game setting, the main character needs to explore the environment and search for evidences to show his innocence. The evidences are indicated as special items or agents in the environment (Fig. 4). Once the main character found an evidence, a short animated drama conveying information of the murder will be presented. All dramas are generated by TSEiA from determined stories. For example, one of the evidences in our setting is a video tape. A drama disclosing the content of the video tape is shown to participant once the main character found the tape. There are 6 evidences located at different places in the environment. Participants are not informed the total amount of the evidences prior to the experiment. The more evidences participants can find, the more information they can obtain.

Combining the information obtained from the text and dramas, participants are required to have a conversation with the policeman and answer some questions. For each question, we initially gave 1 to 2 available answers for participants to choose from. However, if participants were not satisfied with the answers being provided, they were able to add their own answers. The answers created by participants in each group were synchronized and available for participants in next group. To be noticed, participants are not informed that the answers are created by other participants prior to the experiment.

We designed two types of question to evaluate both the content comprehension and motivation to create stories of participants. The first type of question is related to the objective content of the story in game. Participants are able to answer this type of question by exploring the game environment and collecting evidence dramas provided in the game. Factual answers are required for this type of question. Half of the first type questions are related to information provided in text format as control group to show the effectiveness of visualization (Q1). The second type of question is open question without clear or determined answers. Participants can give arbitrary based on their own understanding. The quantity and quality of the answers created by participants can reflect participants' motivation to create new stories (Q2). Typical examples for these two types of questions are as following:

First type: Who was in the room when you saw the dead body?

Second type: Tell me about your family, do you have any issues in your family recently?

18 questions including 12 first type questions and 6 second type questions are required to be answered. After the session, participants are requested to finish a questionnaire. We evaluate the integrated performance based on the questionnaire.

4.3 Result and Analysis

Table 1 compares the result of questions related to text information (text-related question) versus those related to visualization information (visualization-related question) and suggests that the average score of visualization information (MEAN=0.71) is significantly higher than that of text information (MEAN=0.47). We evaluate answers given by participants based on the the research on question test by Hensen and Johnston. The score of an answer is decided by three attributes: *Comprehension*, *Accuracy* and *Consistency*. Figure 5 shows the average score of answers to text-related questions and visualization-related questions on each attribute. Answers those are more relevant to fact obtain higher score (0.0–1.0). As a result, it can be inferred that the visualization dramas are more effectively and accurately in terms of conveying information in this experiment. In another word, the visualization dramas can help participants have a better understanding the stories (Q1). This is also reflected on the questionnaire (Fig. 6). Most of the participants agreed that visualization dramas helped them understand the content better.

Table 2 demonstrates some answers to second type questions given by participants. Number increment of answers quantitatively suggests that participants are more motivated to create their own content than choose from available answers (Q2). Simultaneously, available answers provided baseline, comparison and inspired participants to create new content. New answers were given somehow related to previous answers as inferences, deductions or reasons. This transfers the signal that visualized collective storytelling can support participants to

Table 1. Relevance score

	Question	Number	Mean	SD
Score	Text	6	0.47	0.08
	Visualization	6	0.71	0.06

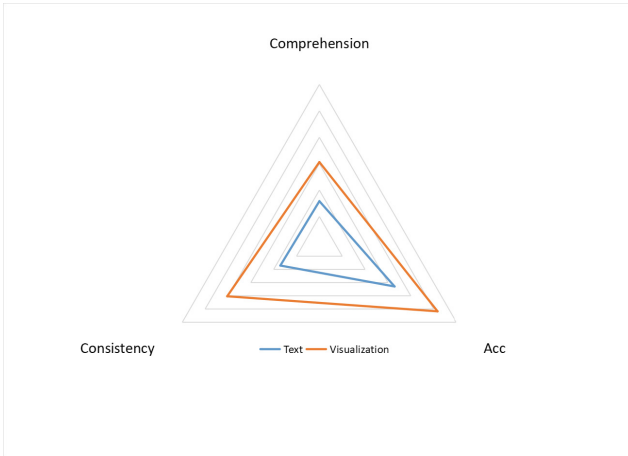


Fig. 5. Evaluation

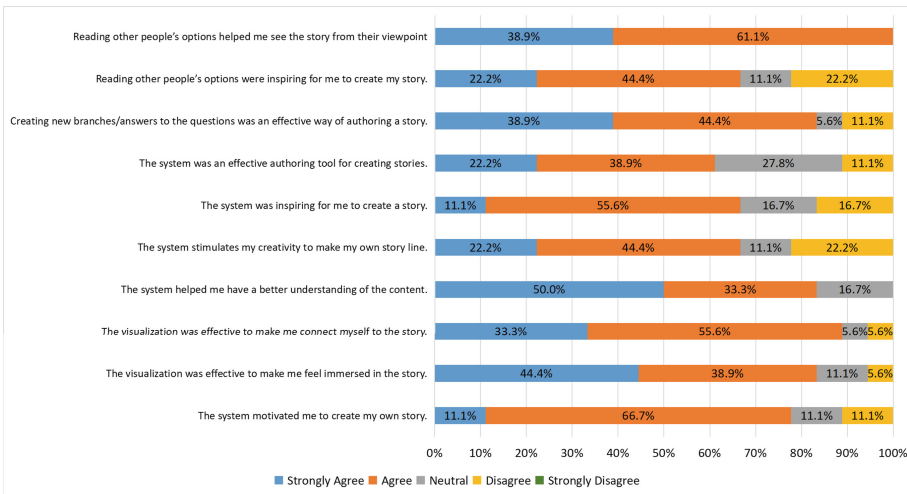


Fig. 6. Questionnaire result

analyze the situation and contribute to content generation. The questionnaire also suggests that story envisioning can stimulate, inspire participants to create more contents (Fig. 6). Moreover, contents created are derived from cultural or

Table 2. Contents created by participants

Question	Number of answers	Sample answer 1	Sample answer 2
Where were you at the time of the murder?	(1) 12	“I was walking a dog”	“I was working on my medicine in the lab.”
Did you personally see anyone attempting to poison the boss?	(1) 10	“I did not. But I know there were some tensions in the office.”	“Maybe Lury uses poison by himself.”
Do you have any issues in your family recently?	(1) 9	“I am personally okay but I cannot say that about everyone.”	“I have no problem but my son get hurt in his soccer game.”
Who is your closest friend in the company and who is the least close person to you in the company?	(1) 11	“Charlie is a nice person I think, but I don’t like Simons.”	“Charlie has been reliable, but any people didn’t trust Simons.”
Explain more about your boss’s family.	(1) 15	“I may have taken the ice cream from his son and threw it on the ground to show him what is mean.”	“I bumped his wife to the street and got hit by a bus, lost her voice. Maybe that’s why he likes me.”
Do you know who has the access to the lab where the new medicine is stored?	(1) 16	“The lock’s broken since 2 weeks ago, we pretended to swipe our cards until they fixed it.”	“Anyone can because the door is always open.”

personality elements of the participants, which indicated that the contents can be potentially used for cross-culture conversation analysis.

5 Conclusion

In this paper, we proposed the method of story envisioning to assist people envision their stories in daily conversation. The method of story envisioning considers human daily conversation as an update of common ground and uses graphic recording theory to visualize the common ground. Story envisioning describes the method of using interactive dramas to represent story pieces in order to involve not only story contents but also prior knowledge, background and other shared cultural information. We state that story envisioning could help people have a better understanding of the stories in daily conversation and could be an effective method for visualized collective storytelling. System TSEiA was built to provide story envisioning services. Experiment is conducted on a role-playing game platform powered by TSEiA to evaluate the usage and effectiveness

of story envisioning. The results showed that story envisioning helped participants understand the stories more concretely and accurately. Simultaneously, the experiment showed that story envisioning inspired and stimulated participants to collaboratively generate more conversational contents. Experiment showed that the system could be used as an effective tool to train communication skills and collect cross-culture data.

The system is still under development and it has limitations. For example, the system has problem with handling stories with object interaction and descriptive information. Although the method of action sequence greatly reduced the difficulty of visualization, it cannot handle complicated input with detailed descriptions. In this sense, relationships between two or more objects in the story should also be extracted. Moreover, to create more contents for cross-culture studies, we will prepare the system for crowd sourcing. Using the data corpus created, we will be able to create predictive model that can automatically give story branches.

References

1. Nishida, T., Nakazawa, A., Ohmoto, Y., Mohammad, Y.: Artificial intelligence and conversational intelligence. In: *Conversational Informatics*, pp. 1–16. Springer, Tokyo (2014). https://doi.org/10.1007/978-4-431-55040-2_1
2. Nishida, T., Nakazawa, A., Ohmoto, Y., Mohammad, Y.: Conversation: above and beneath the surface. In: *Conversational Informatics*, pp. 17–41. Springer, Tokyo (2014). https://doi.org/10.1007/978-4-431-55040-2_2
3. Teeter, P., Sandberg, J.: Cracking the Enigma of asset bubbles with narratives. *Strat. Organ.* **15**(1), 91–99 (2017)
4. Jefferson, G.: Sequential aspects of storytelling in conversation. *Studies in the Organization of Conversational Interaction*, pp. 219–248 (1978)
5. Sacks, H., Jefferson, G.: *Lectures on conversation* (1995)
6. Battaglino, C., Bickmore, T.: Increasing the engagement of conversational agents through co-constructed storytelling. In: *Eighth Workshop on Intelligent Narrative Technologies* (2015)
7. Mirzaei, M.S., Zhang, Q., Nishida, T.: Conversation envisioning to train intercultural interactions. In: Meiselwitz, G. (ed.) *SCSM 2018, Part II. LNCS*, vol. 10914, pp. 68–80. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91485-5_5
8. Mirzaei, M.S., Zhang, Q., van der Struijk, S., Nishida, T.: Conversation envisioning framework for situated conversation. In: Mouhoub, M., Sadaoui, S., Ait Mohamed, O., Ali, M. (eds.) *IEA/AIE 2018. LNCS (LNAI)*, vol. 10868, pp. 517–529. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-92058-0_50
9. Coyne, B., Sproat, R.: WordsEye: an automatic text-to-scene conversion system. In: *SIGGRAPH 2001* (2001)
10. Åkerberg, O., Svensson, H., Schulz, B., Nugues, P.: CarSim: an automatic 3D text-to-scene conversion system applied to road accident reports. In: Copestake, A., Hajic, J. (eds.) *Proceedings of the Tenth Conference on European Chapter of the Association for Computational Linguistics*, vol. 2, pp. 191–194. ACM Press, New York (2003)

11. Johansson, R., Williams, D., Berglund, A., Nugues, P.: CarSim: a system to visualize written road accident reports as animated 3D scenes. In: Proceedings of the 2nd Workshop on Text Meaning and Interpretation, pp. 57–64. Association for Computational Linguistics, Stroudsburg (2004)
12. Ma, M.: Automatic Conversion of Natural Language to 3D Animation. Ph.D. thesis, University of Ulster (2006)
13. Chang, A.X., Savva, M., Manning, C.D.: Learning spatial knowledge for text to 3D scene generation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 2028–2038 (2014)
14. Chang, A., Monroe, W., Savva, M., Potts, C., Manning, C.D.: Text to 3D scene generation with rich lexical grounding (2015). arXiv preprint: [arXiv:1505.06289](https://arxiv.org/abs/1505.06289)
15. Johnson, W.L., Valente, A.: Tactical language and culture training systems: using artificial intelligence to teach Foreign languages and cultures. In: Proceedings of the 20th Innovative Applications of Artificial Intelligence (IAAI) Conference, Los Angeles, Alelo (2008)
16. Cassell, J., Bickmore, T.: Negotiated collusion: modeling social language and its relationship effects in intelligent agents. *User Model. Adapt. Interfaces* **13**(1–2), 89–132 (2003)
17. Swanson, R., Gordon, A.S.: Say anything: using textual case-based reasoning to enable open-domain interactive storytelling. *ACM Trans. Interact. Intell. Syst.* **2**(3), 1–35 (2012)