Generating Instructions in a 3D Game Environment: Efficiency or Entertainment?

Roan Boer Rookhuiszen and Mariët Theune

Human Media Interaction
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands
a.r.boerrookhuiszen@student.utwente.nl, m.theune@ewi.utwente.nl

Abstract. The GIVE Challenge was designed for the evaluation of natural language generation (NLG) systems. It involved the automatic generation of instructions for users in a 3D environment. In this paper we introduce two NLG systems that we developed for this challenge. One system focused on generating optimally helpful instructions while the other focused on entertainment. We used the data gathered in the Challenge to compare the efficiency and entertainment value of both systems. We found a clear difference in efficiency, but were unable to prove that one system was more entertaining than the other. This could be explained by the fact that the set-up and evaluation methods of the GIVE Challenge were not aimed at entertainment.

Keywords: instructions, 3D environment, Natural Language Generation, game, evaluation, efficiency vs. entertainment.

1 Introduction

Natural Language Generation (NLG) is the automatic conversion of some non-linguistic representation of information to written text in natural language (e.g., English). Most NLG systems focus on efficiency and effectiveness, generating texts that are aimed at getting the information across in an optimal way. Common applications of NLG are the generation of weather forecasts and various other kinds of reports. NLG is also used for the generation of system utterances in dialogue systems such as interactive travel guides or virtual tutors. So far, NLG has rarely been used in entertainment-oriented applications such as games.

In this paper, we present the NLG systems we developed for the Challenge on Giving Instructions in Virtual Environments (GIVE), an NLG evaluation challenge to generate instructions for users in a game-like 3D environment. We participated in the GIVE Challenge with two NLG systems: one system that was focused on generating maximally helpful instructions (the Twente system) and one that was intended to be more game-like and thus entertaining (the Warm/Cold system). Although the GIVE Challenge was presented as a game

A. Nijholt, D. Reidsma, and H. Hondorp (Eds.): INTETAIN 2009, LNICST 9, pp. 32–43, 2009.

to its users, who were invited to ‘play a game’, the evaluation criteria used in the Challenge still focused on effectiveness and efficiency of the generated instructions. In other words, the NLG systems were evaluated as if used in a serious application rather than a game. Nevertheless, in this paper we will try to use the data collected in the evaluation period of the GIVE Challenge to compare our own systems in terms of not only efficiency, but also entertainment.

Overview of Paper. We introduce the GIVE Challenge in more detail in Section 2. In Section 3 we describe the NLG systems we have developed. Our hypotheses on the differences between the systems, and the methods to measure those differences are discussed in Section 4. The evaluation results of our systems are provided in Section 5 and we conclude with a discussion in Section 6.

2 The GIVE Challenge

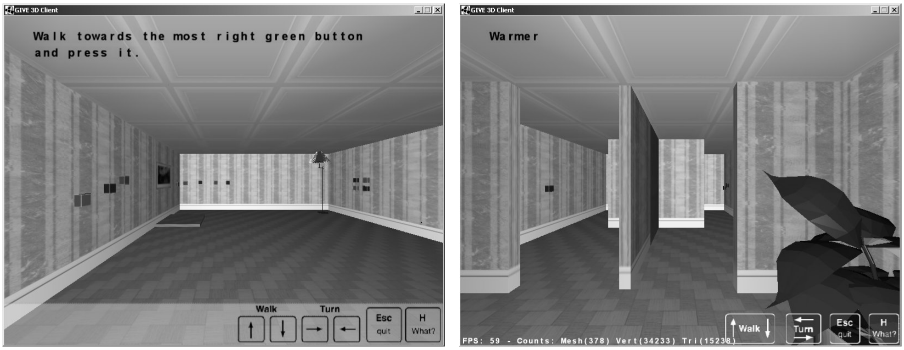
The GIVE Challenge was designed for the evaluation of automatically generated instructions that help users carry out a task in a game-like 3D environment. We participated in the first installment of the Challenge: GIVE-1 [1].

The GIVE Challenge tries to tackle a difficult problem in the field of natural language generation: evaluation. Since multiple outputs of an NLG system may be equally good (the same information can be expressed in natural language in a variety of ways), it is difficult to automatically evaluate generated texts against a ‘gold standard’ of texts written by humans. Therefore, the evaluation in GIVE is done based on performance data and subjective ratings gathered with a user questionnaire. Because the GIVE game could be played in an ordinary web browser, it was expected that a large amount of data and human judgements could be collected relatively easily [2]. A website was set up with some short instructions and the game. Players were recruited via (mainly NLG related) email distribution lists and postings on other Internet websites. In total 1143 games were played by people from all over the world, with the largest number of users coming from the USA, Germany and China.

2.1 The Task

Users of the GIVE system were asked to perform a task in a game-like 3D virtual environment. To ‘win’ the game, they had to follow the instructions that the NLG system produced.

The 3D environment presented to the player of the GIVE-game consists of one or more rooms, connected with doors. There are some objects (e.g. a chair, a lamp) in the world that can be used as ‘landmarks’ for navigation. On several walls there are buttons of various colors. The objective of the GIVE-game for the player is to find a trophy without triggering an alarm. The trophy is hidden in a safe behind a picture on one of the walls. The safe can only be opened by pressing multiple buttons in the right order. The user has a first person view of the world and can walk through it and turn to left or right (but he cannot walk through walls and closed doors). The user can also press buttons. The function


Walk towards the most right green button and press it.

Warmer

FPS: 59 - Counts: Mesh(378) Vert(34233) Tri(15238)

(a)

(b)

Fig. 1. Screenshots of the GIVE game, showing instructions from the Twente system (a) and the Warm/Cold system (b)

of each button however is unknown to the user: a button can open a door, move a picture, but also trigger an alarm. If the user is in the wrong location and passes a detector the alarm is also triggered. It is sometimes necessary to press multiple buttons in a specific order to perform one of the actions described above.

The interface for the user is shown in Figure 1. At the top of the screen instruction sentences are presented to the user. These instructions tell the user which actions he should perform and help him to achieve the goal. The NLG system that generates those instructions has complete knowledge of the world and the actions to be performed in order to win the game; see section 2.2. There are three different game worlds available; for each game one is randomly selected. The worlds have a different layout and provide different levels of difficulty for the instruction-giving system.

2.2 Architecture

The goal of the GIVE Challenge was to develop an NLG system and not to implement a whole client-server architecture. Each participant of the challenge only had to implement the language generation part of the game. All other software needed to run the game was provided by the GIVE organizers. Below we list the main components of the GIVE game environment.

The Client. The client is the actual program the users used to play the game. It could be started from the GIVE website. The client displayed the 3D environment in which a user could walk around and perform several actions. It also displayed the text generated by the NLG system. Before and after the game, the client presented the users with a questionnaire (see section 2.3).

The Matchmaker. During the evaluation period of the GIVE Challenge (7 November 2008 - 5 February 2009), the GIVE organizers ran a matchmaker server. This server held a list of all NLG systems made by the participants of the challenge. As soon as a user started a client, the matchmaker randomly assigned a NLG system to this client. After the game was finished (with or without success), a complete log of all actions performed by both the NLG system and the user was saved in a database for later evaluation.

The NLG System. The language generation part of the game was implemented by each team participating in the Challenge. The input for language generation consisted of a plan containing the sequence of actions the user should perform to successfully achieve the task (i.e., win the game). This plan was updated after each user action. Furthermore the system had complete knowledge of the virtual environment; it knew the position and properties of all objects in the environment, and which objects were currently visible to the user. Based on this information it generated sentences informing the user about what he had to do. The only feedback on the system's instructions were the actions a user performed after having received the instruction, and a notification whenever the user pressed a 'Help' button.

7-point scale items
overall: What is your overall evaluation of the quality of the direction-giving system? (1 = very bad, 7 = very good)

5-point scale items
task difficulty: How easy or difficult was the task for you to solve? (1 = very difficult, 5 = very easy)

goal clarity: How easy was it to understand what you were supposed to do? (1 = very difficult, 5 = very easy)

play again: Would you want to play this game again? (1 = no way!, 5 = yes please!)

instruction clarity: How clear were the directions? (1 = totally unclear, 5 = very clear)

instruction helpfulness: How effective were the directions at helping you complete the task? (1 = not effective, 5 = very effective)

choice of words: How easy to understand was the system's choice of wording in its directions to you? (1 = very unclear, 5 = very clear)

referring expressions: How easy was it to pick out which object in the world the system was referring to? (1 = very hard, 5 = very easy)

navigation instructions: How easy was it to navigate to a particular spot, based on the system's directions? (1 = very hard, 5 = very easy)

friendliness: How would you rate the friendliness of the system? (1 = very unfriendly, 5 = very friendly)

Nominal items
informativity: Did you feel the amount of information you were given was: too little / just right / too much

timing: Did the directions come: too early / just at the right time / too late

Fig. 2. The post-game evaluation questions

In total 4 teams participated in the Challenge, with 5 different NLG systems [3]. As our contribution we created two NLG systems, discussed in this paper.

2.3 Questionnaire

Before and after the game, the user was confronted with an optional questionnaire. This questionnaire was designed by the organizers of the GIVE Challenge; it was the same for each NLG system. Before the game, the user was asked for the following personal information: age, profession, level of computer expertise, level of proficiency in English, and experience playing video games. Then the user played a game, with a randomly assigned combination of game world and NLG system. After the game was finished, the user was asked to rate various aspects of the game experience such as the clarity and helpfulness of the instructions, and the friendliness of the system. The user was also asked to rate the quality of the direction-giving system with an overall score. Most questions had to be answered with a rating on a 5-point scale. The full list of questions asked in the post-questionnaire can be found in Figure 2.

3 Our NLG Systems

For the Challenge, we designed two NLG systems, each with a different goal:

1. The Twente system, focusing on efficiency
2. The Warm/Cold system, focusing on entertainment

The first system, the Twente system, is purely task-oriented and tries to guide the user through the game as efficiently as possible. The Warm/Cold system on the other hand tries to make the game more entertaining for the user even if a consequence is a decrease of the efficiency. Below we describe both systems. A more detailed description of the systems' designs is given in [4].

3.1 The Twente System

The organization of the GIVE Challenge provided all participating teams with an example implementation of an NLG system. This system was very basic and gave only one instruction at a time. This was easy to understand, especially for new users; however it was very annoying for more experienced users. In our first attempt at implementing our own NLG system, all instructions to get to a button were combined into one sentence. More experienced users did perform better with this system than with the example system (they used less time, and found the instructions clearer), but inexperienced users could not handle the increased complexity of the instructions. Because of this difference between new and more experienced users we decided to design a framework with three different levels. The first level generates very basic instructions, explicitly mentioning every step of the plan. The higher levels generate more abstract, global instructions that are expressed using more complex sentences. Some example sentences generated by the different levels:

– Level 1: Only one instruction at a time: “Walk forward 3 steps”, “Press the blue button”, “Turn right.”
– Level 2: A combination of a walk instruction and another action instruction: “Walk forward 3 steps then press the button.”
– Level 3: Also a combination, but only referring to objects when the user can see them: “Turn right and walk forward”, followed by “Press the blue button.” See Figure 1(a) for an example.

In the third level we thus do not give the exact route to the next button to be pushed, but try to encourage users to walk to it on their own. Only whenever the user goes into the wrong direction the system will give an extra instruction.

Our framework is adaptive; the NLG system will try to fit the level to the user's needs. It is expected that novice users learn while they are playing the game. The system is *able to detect* the level of experience of the user and automatically change the level during the game. When the game starts the level used is 2. Every second, the system checks the number of actions the user performed in the last 5 seconds. Whenever this number exceeds a certain threshold the user will probably perform better on a higher level, so the level is switched upward. On the other hand the level is switched down as soon as the number of actions is low or the user presses the ‘Help’ button.

All levels are generated using the same general framework. The sentences generated by the different levels are very similar. Certain actions are the same for all levels: interpreting events, the generation of referring expressions (“The blue button”) and the check whether users are performing the right actions. Only the timing and specific realization of a sentence is different between levels. In our framework a new level can simply be added by making a new class containing the functions that realize the text of all types of sentences. It is very easy to switch between two levels: only the realization of the sentence is different.

We asked a few users to play each level of the game separately (no automatic switching of levels). These test users suggested several small adaptations. For example, in our first version of level 2 a sentence consisted of an action followed by a move instruction. For example:“Turn right then walk 3 steps”. People found it more natural and easier to understand when the order was changed to having a move followed by an action: “Walk 3 steps then turn right.”

3.2 The Warm/Cold System

To make the task more interesting for the users, we created a more game-like situation with a system that tries to simulate a warm/cold game. Instead of telling the user exactly what to do, the only instructions given are “warmer” and “colder” to tell the user if he comes closer to the next button to be pushed, “turn” to indicate that the user only has to turn to see that button and of course the instruction to push it. Before the user gets his first ‘hint’, he has to walk around in any direction. To find the next button it is not always enough to follow the instruction “warmer”. Sometimes the user has to make a small detour to get around a wall or another obstacle. To encourage the user, some exaggerated statements are used when the user is very close to the target ("Feel the heat!"). The instructions do not prevent the user from triggering an alarm while walking around. As soon as he triggers an alarm he has lost the game.

In short, the system does not exactly tell the user where to go but leaves the navigation choices open. This is illustrated in Figure 1(b): the user is warned that he is getting closer ("warmer") to the button to be pushed, but he still has to decide for himself whether to go left or right. It is expected that this makes it more interesting to play the game, although it will probably decrease the efficiency and increase the playing time. As game studies have shown, player enjoyment increases if a game is more challenging, and if the players have more control and freedom to play the game in the way they want [5].

It was not expected that the Warm/Cold system would perform well in the GIVE Challenge, because the GIVE evaluation focused on efficiency and clarity of instructions. The overview of the results of all participating systems confirms this expectation [3].

4 Evaluation

To test if our systems have achieved their respective goals (being efficient versus being entertaining), we evaluate them using the data collected for our systems in the GIVE Challenge, including the action logs of the system and the answers to the questionnaires. We will compare the results of the Twente system and the Warm/Cold system in light of the two goals. Our main hypotheses are:

Hypothesis 1 - The Twente system is more efficient than the Warm/Cold system.

Hypothesis 2 - The Warm/Cold system is more entertaining than the Twente system.

To test these hypotheses, the only information we have available are the data collected in the GIVE Challenge. A disadvantage is that the evaluation questions used in the Challenge were about clarity and task performance rather than the users' experiences. This means they are most suitable to test Hypothesis 1, whereas for Hypothesis 2 we will have to rely on more indirect clues. Below, we describe how we intend to measure the efficiency and entertainment value of our NLG systems in terms of the data available from GIVE.

4.1 Measuring Efficiency

The efficiency of a system can be measured objectively by using the logged performance results. One system is more efficient than another if using this system, the users successfully performed the task in *less time* and with less detours (i.e., using *fewer steps*) than when using the other system. We also take *task success rate* as an objective indicator of efficiency.

Most questions in the post-questionnaire (Figure 2) deal with the subjective perception of efficiency. We assume that one system is perceived as more efficient than another if it scores better on *task difficulty*, *goal clarity*, *instruction clarity*, *instruction helpfulness*, *choice of words*, *referring expressions*, *navigation instructions*, *informativity* and *timing*. Also the overall rating of the quality of the direction-giving system *(overall)* is expected to be better, based on the assumption that the users mostly based this rating on the clarity and helpfulness of the instructions, rather than on the entertainment value of the game.

4.2 Measuring Entertainment

It is expected that users find a game more interesting if they have to try harder to finally achieve the goal of the game, as is the case in the Warm/Cold system when compared to the Twente system. The GIVE action logs provide some information that may indicate how entertaining the users found each game. First, *cancellation frequency*: if the user is more interested in the game he is less likely to cancel it. Second, *playing time until cancellation*: if the user does cancel, this is expected to be after a longer period.

As said, the GIVE questionnaire was primarily aimed at measuring clarity and effectiveness of the system's instructions. However, one of the questions can be directly related to the system's entertainment value: if the game is entertaining, the user is more likely to want to play it again. So, in the user questionnaire we expect to find that the score given for *play again* is higher for Warm/Cold than for Twente, even after the user has lost the game.

Finally, we think that if users find a game entertaining, they are at least as interested in the process of playing as in the outcome of the game. Therefore we expect that the more entertaining the users find a system, the less they care about losing. Overall, our prediction is that when the 'game-play' merely consists of carrying out instructions (as with the Twente system), failing to achieve the task ('losing' the game) will negatively influence the users' subjective judgement of the system, whereas in a more entertaining situation (as with the Warm/Cold system) the users' judgement will be much less influenced by the game's outcome.

5 Results

The results presented here are based on the data gathered in the GIVE Challenge. The subjective user ratings for the Twente and Warm/Cold systems, given in Table 1, and a few other results reported here, were taken from the overview paper [3] discussing the outcomes of the Challenge. We computed the other results presented in this section from the raw evaluation data for our two systems, which were made available to us by the GIVE organizers.

Roughly one third of the games was played from an IP address in the USA, another third from Germany and the rest from other countries. Around 80% of the games were played by male users, 10% by female users and for 10% the gender was not specified. Unfortunately we were unable to determine whether all

Table 1. Results of the GIVE user questionnaire taken from [3]. Results that are **significantly different** (with $p < 0.05$) are given in bold face. For *informativity* and *timing* we give the percentages of "just right" answers; these were computed for successful games only.

Question	Twente	Warm/Cold
overall	**4.3**	**3.6**
task difficulty	**4.0**	**3.5**
goal clarity	**3.9**	**3.3**
play again	2.4	2.5
instruction clarity	**3.8**	**3.0**
instruction helpfulness	**3.6**	**2.9**
choice of words	**4.1**	**3.5**
referring expressions	3.7	3.5
navigation instructions	**4.0**	**3.2**
friendliness	3.1	3.1
informativity	51%	51%
timing	60%	49%

games also represent different users, as the GIVE Challenge only distinguished between game plays, not between users. It is possible that users played a game with another NLG system before they used one of our systems.

In Table 1, the results from the questionnaire are reported as the mean ratings given by the users of each system. Each mean value was calculated from roughly 50 answers. Significance was tested by using Tukey tests. The means that are significantly different (with $p < 0.05$) are shown in bold face.

Hypothesis 1. The Twente system is more efficient than the Warm/Cold system if we look at the objective measurements: the task is performed in less time (207.0 vs. 312.2 seconds), using fewer steps (160.9 vs. 307.4). Also the task success rate is significantly higher (35% vs. 18%) [3].

As we have seen in Section 4.1, most of the questions in the questionnaire consider the subjective perception of efficiency. The results shown in Table 1 clearly show that for all questions related to efficiency, except *referring expressions*, there is a significant difference between the means of the two systems.

Hypothesis 1 is confirmed: the Twente system is more efficient than the Warm/Cold system.

Hypothesis 2. In relation to this hypothesis, we predicted that when a game is more entertaining, the player is less likely to cancel it. However, the game logs show almost no difference: 25.8% of the games with the Twente system were cancelled, against 24.6% of the games with the Warm/Cold system. We also expected that entertaining games would be cancelled after a longer period. However, the mean playing time before cancellation was 234 seconds for the Twente system and 233 seconds for the Warm/Cold system. These results contradict our expectation; there is no significant difference between the two systems. The scores for *play again* are not significantly different either (see Table 1).

We also suggested that when a game is entertaining, the outcome is less important than when it is not. To investigate the extent to which the outcome influenced the subjective ratings of each system, we compared our systems' ratings for the games in which the user won and the games in which the user lost. For each system, we tested the significance of the differences between the means of the successful and lost games by using Tukey tests. In Table 2 the means with a significant or near-significant difference are indicated by ** (with $p < 0.05$) or * (with $p < 0.10$). For the Twente system, *task difficulty*, *play again*, *instruction clarity* and *referring expressions* show a significant difference between the user ratings, when distinguishing between won and lost games. This shows that losing a game did cause users to judge the Twente system more negatively on these aspects, whereas for the Warm/Cold system no such negative influence of losing was found. This is in line with our hypothesis. However for one question, *goal clarity*, a significant difference between won or lost games was found for the Warm/Cold system, but not for the Twente system. We will try to give an explanation for this in the discussion.

Based on these results, we can neither confirm nor reject Hypothesis 2.

Table 2. Results of the GIVE user questionnaire. Significant differences are indicated by ** (with $p < 0.05$) and * (with $p < 0.10$)

Question	Twente			Warm/Cold		
	Won	Lost		Won	Lost	
overall	4.34	4.26		3.93	3.60	
task difficulty	**2.15**	**3.83**	**	3.55	3.57	
goal clarity	4.10	3.64	*	**3.62**	**2.94**	**
play again	**2.14**	**3.06**	**	2.56	2.54	
instruction clarity	**4.06**	**3.46**	**	3.22	2.93	
instruction helpfulness	3.64	3.64		3.02	2.91	
choice of words	4.22	3.74	*	3.89	3.62	
referring expressions	**3.96**	**3.33**	**	3.76	3.36	
navigation instructions	3.96	3.76		3.38	3.29	
friendliness	3.27	2.94		3.29	3.07	
informativity	2.26	2.08		1.67	1.69	

6 Discussion

Some of the results presented in the previous section differ from what we expected. For example, Table 2 shows a significant difference in *goal clarity* between lost and successful games for the Warm/Cold system, but not for the Twente system. Our hypothesis however was that this should be the other way around. We can explain this because in the GIVE Challenge, the users were led to expect a system aimed at efficiency. The Warm/Cold system has another goal, but this was not (clearly) communicated to the user. It seems that the users were confused about the goal of the Warm/Cold game, and 'blamed' the explanation after losing a game.

In general, the evaluation results for both systems were probably strongly influenced by the users' expectations. In the introduction of the GIVE game, the NLG system was presented to the user as a 'partner' or 'assistant' who would "tell you what to do to find the trophy. Follow its instructions, and you will solve the puzzle much faster." In short, all information provided to the users suggested that the instructions would be as helpful as possible. The players thus expected a co-operative assistant that would obey the Cooperative Principle proposed by the philosopher Grice: "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose of the talk exchange in which you are engaged." ([6], p. 45). In accordance with the Cooperative Principle, Grice proposed four conversational maxims:

- Maxim of Quantity: Make your contribution as informative as needed, but not more informative than required
- Maxim of Quality: Do not say what you believe to be false or for which you lack adequate evidence
- Maxim of Relation: Be relevant
- Maxim of Manner: Be perspicuous, and avoid ambiguity.

These maxims can be seen as rules a co-operative speaker uses in a conversation. They underlie most work in NLG, and we have obeyed them for the instructions generated by the Twente system. In contrast, we intentionally failed to fulfill some of the maxims to make the Warm/Cold system more challenging. We flouted the Maxim of Manner: our instructions were obscure, and we introduced ambiguity in our direction giving. This is also in violation of the Maxim of Quantity: we gave less information than we could. This made it much harder for the users to understand the direction giving instructions of the system. Instead of just blindly following the instructions, in the Warm/Cold version the user should think of a strategy to be able to win the game, which we expected would make the game more entertaining.

Note that the conversational behaviour of the Warm/Cold system could still be seen as cooperative, in the sense that its instructions were "such as is required (...) by the accepted purpose of the talk exchange" if this purpose was defined as achieving entertainment. However, as mentioned above, this purpose was not clearly presented to the users of the GIVE game. Rather, the accepted purpose was to guide the users as efficiently as possible to the trophy. This probably explains the lower ratings on all questions for the Warm/Cold system compared to the Twente system.

In short, the GIVE Challenge was set up to measure efficiency-related quality aspects of generated instructions. In this paper, we have tried to extract the level of entertainment provided by our systems' instructions from data that were not fully suitable to measure this. In future editions of the GIVE Challenge, it would be good if the participating teams could adapt the user questionnaire to their own research questions. In our case, this would allow us to use better methods for measuring entertainment value, such as the FUN questionnaire developed by Newman [7]. This questionnaire was designed to evaluate player enjoyment in roleplaying games, measuring the degree in which (1) the user lost track of time while playing, (2) felt immersed in the game, (3) enjoyed the game, (4) felt engaged with the narrative aspects of the game, and (5) would like to play the game again. The FUN questionnaire looks like a good starting point for our evaluation purposes, and could easily be adapted to our own game context, as was also done by Tychsen et al. [8].

Acknowledgements. We thank Michel Obbink for his contribution to the development and testing of the Twente system. We are also grateful to the organizers of the GIVE Challenge for the development of the framework and for collecting the evaluation data. In particular we thank Alexander Koller for his quick responses to all our questions. This research has been supported by the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie).

References

1. Koller, A.: First NLG Challenge on Generating Instructions in Virtual Environments (GIVE-1), `http://www.give-challenge.org/research/page.php?id=give-1-index` (last visited March 20, 2009)
2. Koller, A., Moore, J., di Eugenio, B., Lester, J., Stoia, L., Byron, D., Oberlander, J., Striegnitz, K.: Instruction Giving in Virtual Worlds. In: Dale, R., White, M. (eds.) Report of the Workshop on Shared Tasks and Comparative Evaluation in NLG, pp. 48–55 (2007)
3. Byron, D., Koller, A., Striegnitz, K., Cassell, J., Dale, R., Moore, J., Oberlander, J.: Report on the First NLG Challenge on Generating Instructions in Virtual Environments (GIVE). In: Proceedings of the 12th European Workshop on NLG, special session on Generation Challenges (2009)
4. Rookhuiszen, R.B., Obbink, M., Theune, M.: Two Approaches to GIVE: Dynamic Level Adaptation versus Playfulness. GIVE-1 Report (2009), `http://www.give-challenge.org/research/files/GIVE-09-Twente-WarmCold.pdf`
5. Sweetser, P., Wyeth, P.: GameFlow: A Model for Evaluating Player Enjoyment in Games. ACM Computers in Entertainment 3(3) (2005)
6. Grice, H.P.: Logic and Conversation. In: Cole, P., Morgan, J.L. (eds.) Syntax and Semantics 3: Speech Acts, pp. 41–58. Academic Press, New York (1975)
7. Newman, K.: Albert In Africa: Online Role-Playing and Lessons From Improvisational Theatre. ACM Computers In Entertainment 3(3) (2005)
8. Tychsen, A., Newman, K., Brolund, T., Hitchens, M.: Cross-Format Analysis of the Gaming Experience in Multi-Player Role-Playing Games. In: Situated Play, Proceedings of DiGRA 2007 Conference, pp. 49–57 (2007)