# Using Expressive Language Generation to Increase Authorial Leverage

**Marilyn A. Walker, Jennifer Sawyer, Carolynn Jimenez, Elena Rishes, Grace I. Lin**
**Zhichao Hu, Jane Pinckard, Noah Wardrip-Fruin**
Natural Language and Dialogue Systems Lab
Department of Computer Science
University of California, Santa Cruz, CA 95064, USA
maw|jsawyer|cjimenez|erishes|glin|zhu|jpinckard|nwf@soe.ucsc.edu

## Abstract

It is widely agreed that the interaction possibilities in interactive narrative are limited by the current approach to dialog creation, which typically relies on teams of scriptwriters. This paper reports on experiments testing whether authorial leverage can be increased by methods for natural language generation of dialog in a story world called *Heart of Shadows*. Our experiments show that (1) expert writers spend less time authoring dialogue variations by editing automatically generated content than creating it from scratch and (2) expert writers are more critical of automatically generated dialogue than amateur writers.

## Introduction

It is widely agreed that the interaction possibilities in interactive narrative are limited by the current approach to dialog creation, which typically relies on teams of scriptwriters. This approach either means that the number of story paths available to the player are limited, or that lengthy scripts are needed to develop a range of story paths and dialog interactions (e.g., scripts for *Heavy Rain* and *Beyond: Two Souls* are both over 2000 pages). Recent work has begun to explore whether expressive natural language generation (NLG) engines previously developed for dialogue systems could be used to automatically expand the possible number of story paths and their associated dialog interactions (Lin and Walker 2011; Beskow et al. 2004; Rowe, Ha, and Lester 2008; Cavazza and Charles 2005; André et al. 2000).

This paper presents an experiment exploring whether a creative synergy for authoring dialog can result from integrating a deep model of a story with an NLG engine. We describe how we extend the PERSONAGE language generation engine (Mairesse and Walker 2011) to produce story dialog for a story world called *Heart of Shadows (HoS)*. We then test the generated utterances in an experiment that requires authors to select and edit automatically generated utterances for some scenes, and to author dialog for other scenes. We also compare the behavior and user satisfaction judgments of 30 naive and 29 expert authors.

We first describe the *HoS* story world and how PERSONAGE is used to generate character dialog, and then describe an experiment intended to test both the authoring behavior and the user perceptions of naive and expert writers when asked to author a scene from scratch as compared to selecting and editing dialog produced automatically by an NLG engine. We then present the experimental results, and finally sum up and describe future work.



## Generating Dialog for *Heart of Shadows*

In order to test our hypotheses that NLG could be used to automatically extend the story paths in an interactive story, we constructed a story world targeted at this goal (*HoS*), and adapted the PERSONAGE NLG engine to the characters in this story world.

***Heart of Shadows* Story World.** *HoS* is designed as a paranormal fantasy populated by otherworld elements layered on top of real world experiences. It draws from Celtic folklore legends of the Sidhe, a race of elfish creatures with magical abilities. Otherworlders live all around the world, in the shadows of human existence, often blending in with humans and in general keeping their existence a secret. While most humans are completely unaware of Otherworlders, some humans are Sensitive. The player is one such human. The emphasis of the storyline is on mystery-solving and character interactions, and the characters are intended to be strongly delineated with distinct personalities. In addition to mystery, there are romance subplots and the possibility of interactions with multiple characters. To explore the possibilities of generated dialog, *HoS* is explicitly constructed to allow the same scene be played by different characters in different ways. Fig. 1 provides sample descriptions of characters in the *HoS* world, while Fig. 2 provides sample story scenes that are used in our experiments.

**Natural Language Generation.** Our goal is to dynamically generate dialog that reflects and drives the state of the fictional world in order to increase the number of story trajectories that players can experience; these trajectories are dependent on the events that take place, the characters that are present, the characters' personalities and histories, and the dynamic states of relationships between characters. At present it is not possible to develop rich story dialog that expresses all of these contextual variations. What we are able to experiment with at this stage is automatic generation of different versions of dialogic utterance with stylistic variation based on a set of stylistic parameters.

PERSONAGE is a publicly available NLG engine that provides 67 parameters for varying the expressive style of generated utterances. We cannot describe PERSONAGE in detail here, interested readers are referred to (Mairesse and Walker

| Char | Description | Big Five Trait (scale 1 to 7) |
|---|---|---|
| **Rhys**: | Rhys is the Captain of the Guard for Queen Maeve of the Shadow Court and a proud man who knows and has a large variety of secrets. He has motives to murder everyone for one reason or another, and is more than capable to carry out the deed. Rhys is a persistent and vengeful person. He is both strategic and calculating, and others find him tough and intimidating. He has black hair, green eyes, silver skin, and a scar over his left eyebrow. His hair is extremely long, worn in a plain braid and he is very tall. | agreeableness=1.0, conscientiousness=7.0, emotional-stability=7.0 |
| **Mara**: | Mara is the Queen's sister and accustomed to holding positions of power. She is bitter that she was not named Captain of the Guard and her aggressive nature leads her to pick fights with others. Her primary motives to murder are manipulation, jealousy, and to usurp power. She is charismatic, insincere, proactive, aggressive, too caring, and jealous. Mara has purple hair, black eyes, silver skin, and a scar on her right hand. Her hair is elaborately braided and threaded with jewels, a sign of wealth and prestige in their culture. | extraversion=7.0, agreeableness=1.0, conscientiousness=7.0, emotional-stability=1.0 |
| **Aelyn**: | Aelyn is a hunter, renowned for her tracking skills. She is a curious, independent person who doesn't always like to follow the rules, but she has a kind heart and will always try to help someone in need. She has known Fahn since childhood and is deeply in love with him. She spends a great deal of time in the wilderness and does not care about her appearance, preferring plain clothing over ornate apparel. She has platinum hair, silver eyes, pearl skin and wears her hair uncombed and loose, sometimes with leaves or twigs in it. | extraversion=7.0, agreeableness=7.0, emotional-stability=7.0, open-to-experience=7.0 |
| **Fahn**: | Fahn was once part of the Shadow Court but left it long ago and assumed a new identity. He keeps to himself but he is patient and kind, and he is in love with Mara because of a shared past. He is an herbalist and master of poisons. His primary motives for murder are to protect those he cares about. He is submissive, weak, undisciplined, introverted, agreeable, and open-minded. Fahn has sandy blonde hair that is longer than shoulder length, which he wears brushed back and loose with a lock of hair hanging over his right eye. He has hazel eyes, gold skin, and a soft voice that is close to a whisper. | extraversion=1.0, agreeableness=7.0, open-to-experience=7.0 |

Figure 1: *Heart of Shadows* Characters

2011). In this paper, we make use of PERSONAGE's parameter estimation (PE) statistical generation models. The PE models allow a character's dialog style to be specified by values between 1 and 7 for each of the Big Five Personality Traits of extraversion, conscientiousness, agreeableness, emotional stability and openness to experience. As part of our character design, we therefore created a mapping between *HoS* character descriptions and the Big Five trait specifications as shown in Fig. 1. Traits that were not specified were assumed to default to midrange values.

The other aspect of PERSONAGE that needed to be customized for our story world was the construction of a generation dictionary for *HoS* story propositional content. This involves two aspects: (1) construction of content plans which specify for different dialog goals what type of information can be used to accomplish those goals; and (2) the specification of lexico-syntactic information for each proposition that can provide that type of information. PERSONAGE uses these deep rhetorical and lexico-syntactic representations to automatically generate dialog variations. Here we construct these by hand, but in other work we present methods for generating these automatically from story representations (Rishes et al. 2013). Another limitation of the work presented here is that the story content is more limited than we would like. This means that a large number of PERSONAGE parameters do not get invoked to any extent, such as the operations that affect the verbosity of the utterance, content structuring operations that foreground and background negative or positive information, the aggregation operations that make use of the rhetorical structure between a nucleus proposition and a satellite, and some types of lexical variations. We hope that the methods described in (Rishes et al. 2013) will allow us to more easily develop richer story worlds that will let us more fully explore which of these parameters are useful in interactive stories. Here, the primary set of parameters that PERSONAGE can consistently invoke are those called PRAGMATIC TRANSFORMATIONS, which support variations in address terms and in hedges and emphasizers. The second column of Table 2 illustrates the types of variations that were automatically produced and used in the experiment.

## Experimental Setup

We conducted a web-based experiment in which naive and expert writers were asked to complete authoring tasks. 30 amateur writers were recruited via Mechanical Turk and 29 expert writers (years of experience ranging from 3 to 46 years) were recruited through personal contacts to participate in the experiment. In both cases, the subjects were introduced to the *HoS* domain and prompted to create dialog for four scenes. Fig. 4 shows the instructions given to the participants at the beginning of the experiment. Fig. 2 provides descriptions for two of the scenes. Each scene description included background information on the characters (see Fig. 1), and a brief description of the goal for each turn in the dialog for that scene. For example, Fig. 3 presents a scene from *HoS* describing dialog goals at an abstract level, along with utterances automatically generated to realize those dialog goals, and the edits that naive users made to the automatically generated utterances. In two cases (Scene 1 and Scene 3) human subjects had to create dialog completely from scratch. In two other cases (Scene 2 and Scene 4) the subjects were presented with several versions of dialog generated by PERSONAGE and were asked to pick the best one and edit it if needed. The second column of Table 2 gives a complete list of automatically produced utterance variations used in the experiment. Our hypotheses were:

- **[H1]**: Both Amateur and Expert writers will spend less time on selecting and editing generated content than they spend on creating original content;
- **[H2]**: **(Part-1)** Expert writers will be less satisfied with the generated output and **(Part-2)** there will be greater differences between expert authored utterances and the generated output.

We therefore set up the web pages so that we could measure the time spent by experiment participants on each task, and told participants that they were being timed. Furthermore, the survey included questions measuring user perceptions of task difficulty and naturalness of machine produced dialog: *How difficult was it for you to generate dialogue for this scene?* with responses ranging from *Very Easy* to

| Scene | Description |
|---|---|
| Scene 1: | Rhys and Mara used to be friends until Rhys nabbed the Captain position out from under Mara. Since then, the two have had a bitter rivalry which is ready to explode. Recently Mara stole Queen Maeve's ring, and Rhys knows of Mara's thievery. Scene 1 includes Rhys interrogating Mara concerning the ring. |
| Scene 2: | Rhys and Aelyn have had a relationship in the past, but they are incompatible because he is too rigid in his methods and beliefs. In this scene, Rhys needs Aelyn to help him track Fahn, who is missing. Aelyn tries to dissuade him from his task. Rhys tries to get her to betray Fahn first by insulting him, and then by telling her that Fahn has given Mara a special gift. Aelyn is crushed and decides to help Rhys. |

Figure 2: *Heart of Shadows* Scene Descriptions. Used to guide the dialog authoring and selection in the experiments.

| Char | Dialog Goal | Utterances |
|---|---|---|
| Rhys | Discuss Knowledge (Fahn has run away) | **Generated:** What do you think about the fact that Fahn has run away? **Edited:** Do you have anything to say about the fact that Fahn has run away? |
| Aelyn | Endear (Fahn) | **Generated:** You are always picking on Fahn. He isn't your enemy! **Edited:** You seem to always accuse Fahn. He isn't your enemy! |
| Rhys | Discourage romance and bring to light knowledge (Fahn gave Mara an onyx hair jewel) | **Generated:** Fahn isn't right for you. You should stay away from him. Did you know that Fahn gave Mara a damn onyx jewel? **Edited:** Fahn isn't right for you. You should stay away from him. Did you know that Fahn gave Mara an onyx jewel? |
| Aelyn | Expose suffering. | **Generated:** By the world of ice and death, what do you think about the fact that Fahn has run away? **Edited:** I do not believe this! Is this why Fahn ran away? |

Figure 3: Automatically Generated Utterances for *Heart of Shadows* Scene 2 and User Edits

In this survey, you will be presented with four authoring tasks for an interactive narrative called *Heart of Shadows*. In some tasks, you will write a dialogue based on the information presented to you. In others, you will select machine-generated dialogue from a sample, and then edit it. Use your best judgment in all tasks to create believable and interesting dialogue between the characters.

Figure 4: Experimental Instructions

*Very Difficult*, and *Did you feel that you had enough information about the characters?* with responses ranging from *Not enough information* to *Too much information*, and *How natural did the machine-generated dialogue feel before edits (if any)?* with responses ranging from *Very Natural* to *Very Unnatural* measured on a five-point Likert scale.

Participants from the expert writers group were also asked to fill out an additional survey inquiring about what they view as being the state-of-the-art tools for professional writers: what features authors want and expect, what product they are currently using and whether machine generated dialogue can be helpful even if it is not very natural. The survey also encouraged expert writers to share their insights on how NLG can be improved in order to be useful for them.

Finally, we wrote scripts for automatically processing both the edited and the authored outputs for the experiment in order to analyze typical edit patterns, utterance lengths for authored and generated dialog, type/token ratio for utterances and instances of unique vocabulary.

## Experimental Results

Table 1 shows dialog written by a human author for Scene 1. Fig. 3 provides examples of automatically generated dialog utterances for Scene 2 and illustrates the type of edits that the subjects made to the generated utterances. There were no differences between naive and expert authors on the *difficulty* and the *enough information* survey questions. Only 19% of the authors thought it was *difficult* or *very difficult* to generate the dialogue. Only 25% of the authors thought they had *not enough* or *barely enough* information about the characters.

Our experimental measures include time spent on authoring tasks (measured in seconds), as well as qualitative measures of user perceptions of the generated dialog. The boxplot in Fig. 5 shows the range of time spent on the authoring tasks. It should be noted that the subjects were timed on completion of each authorial task. The time instances in Fig. 5 reflect the time spent on writing dialog for a scene (in case of scenes 1 and 3) or the time spent on selecting and editing automatically generated content (in case of scenes 2 and 4). Our experiment did not measure the time per utterance instance. Both expert and non-expert writers spent approximately the same amount of time (between 150 and 700 seconds) on selecting and editing the output from NLG engine. On the other hand, the distribution of time spent on authoring dialog scenes from scratch is very different for the two groups.

The boxplot also shows differences in time spent on authoring versus editing for both groups of subjects. In order to test whether these differences are statistically significant a paired t test was conducted. We used Wilcoxon test as the time distributions are not normally distributed (see Fig. 6a and 6b). The results show that **H1** holds for both expert ($p < 0.0001$) and non-expert ($p < 0.0001$) authors, i.e. both groups spent significantly more time generating text from scratch than editing NLG versions.

User perception of dialog produced by NLG is different for the two groups. As was predicted by **H2-(Part 1)**, professional writers were more critical of the automatically generated dialog than naive users (see Table 3). The main point of criticism was the fact that some hedges and emphasizers didn't fit well with the tone of the narrative. Analysis of dialog edits made by non-expert writers shows support for this claim as intensifiers such as *Oh gosh* and *Yeah* were among top 5 units removed from the machine generated ut-

Table 1: Example Expert Hand-Written Utterances for Scene 1.

| Char | Dialog Goal | Authored Utterances |
|---|---|---|
| Rhys | Bring up rumor of Mara stealing the Queen's ring | Surely your ears rest far above the petty gossip of the Shadow Court, Mara, but perhaps wisps of rumours of your theft of Maeve's prized ring have reached even you? |
| Mara | Deny rumor and embarrass Rhys about lack of jewels in hair | Yes, they've reached me and I've brushed them away. I need no baseless rumours clouding my head, only jewels and braids. Which reminds me, Rhys, I've been meaning to say... your head has been looking a little empty as of late. Have especially hard times befallen the pauper lord? |
| Rhys | Insult Mara's petty nature | My respect comes from my command, and her majesty the queen. Not my upbringing, and certainly not bauble adorned hair. |
| Mara | Bring up past history (Rhys took Captain of the Guard position when it was to be Mara's) | Ah, yes, your command of the Court's guard. Our scars are not the only thing we share, Rhys. We both yearn for power. But my dear sweet sister saw fit to appoint you Commander, when by law and by rights the title belonged to me. You must have pleased my sister dearly to earn such a prize. |
| Rhys | Accept Mara's accusation | Aye, you and I have shared many things Mara, but Commander of the Guard is a gift her majesty saw fit to bestow upon only me. |
| Mara | Subtly threaten Rhys | No doubt after bestowing yourself upon her. I wonder... if the nobles of the Shadow Court learned of a tryst between her Majesty and the pauper lord... why... I daresay the church would have my dear sister's head. And I assure you, in the wake of that unfortunate tragedy, nothing could prevent my ascention to the throne. How much longer do you think stale gossip of missing rings will tide the Court, Rhys? I dare say they'll soon be in the market for something fresh. |
| Rhys | Accuse Mara of being dramatic | You always did have a flair for the dramatic, Mara. If I were you, I'd be spending less time concocting wild fantasies, and more time considering the fate of whoever is caught in posession of the Queen's ring. Who knows? Perhaps once I've found the culprit, Maeve will consent to allow me to pry a gem from it and braid it as my first hair jewel. I think that would be very lovely. Wouldn't you? |

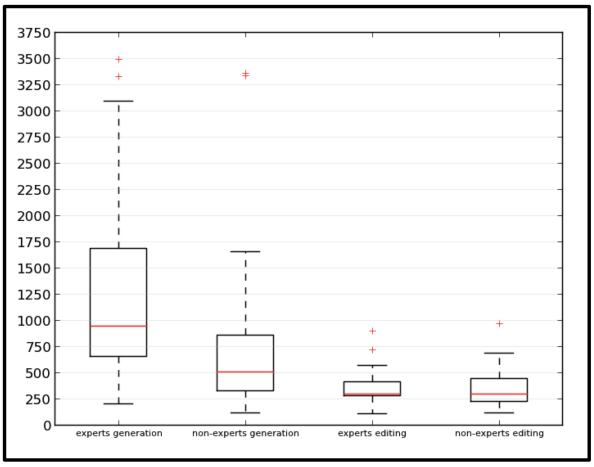
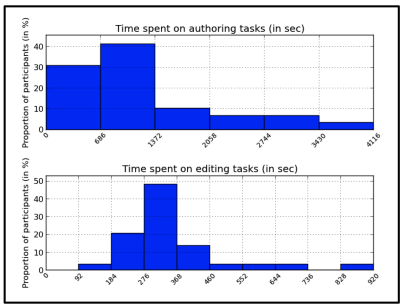
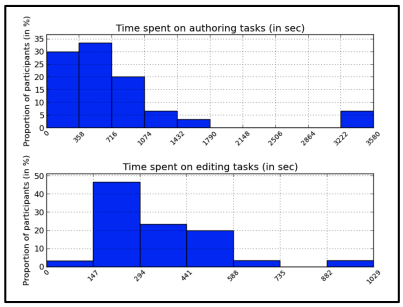



Figure 5: Comparison of time (in seconds) spent on authoring (scenes 1 and 3) vs. dialog selection/editing (scenes 2 and 4) between two groups of subjects




(a) Experts




(b) Non-experts

Figure 6: Time distribution for expert and non-expert writers

terances. In future work we plan to adapt the vocabulary used by PERSONAGE to the romantic fantasy genre of the *HoS* story world.

As for **H2-(Part 2)**, there were no significant differences between the two groups of subjects measured in terms of average length of authored utterances and average number of edits (see Table 4). This result indicates that the chosen metrics are too straightforward to capture such distinction. In future work we plan to develop better measures that go beyond the mere magnitude of changes and aim to characterize the content.

Although more than half of expert writers found the dialog generated by NLG unnatural (see Table 3), overall expert writers gave an affirmative answer to the question of whether NLG could be useful in their work (see Table 5). The majority of experts agreed that dialog produced by NLG gives a good starting point that can be used for brainstorming as it introduces many unexpected combinations. It also helps to tune the personality model by showing several ways a character could say something. As one subject put it: *"It was*

Table 2: Selection Choices and Distribution.

| Scene-Task | Choice of Utterances | Counts[1] NE | Counts[1] E | Distribution NE | Distribution E |
|---|---|---|---|---|---|
| 2-1 | 1.By the world of ice and death, what do you think about the fact that Fahn has run away? | 10 | 7 | | |
| | 2.Um... what do you think about the fact that Fahn has run away? | 5 | 2 | | |
| | **3.What do you think about the fact that Fahn has run away?** | **11**[2] | **19** | | |
| | 4.Yeah, what do you think about the fact that Fahn has run away, alright? | 1 | 0 | | |
| | 5.Yeah, what do you think about the fact that Fahn has run away? | 3 | 1 | | |
| 2-2 | 1.Damn it to Jotunheim, you are always picking on Fahn. He isn't your enemy. | 1 | 3 | | |
| | 2. Mother of a pixie, you are always picking on Fahn. He isn't your enemy. | 6 | 4 | | |
| | **3. You are always picking on Fahn dearest friend. He isn't your enemy.** | **10** | 9 | | |
| | 4. You are always picking on Fahn. He isn't your enemy, he isn't your enemy! | 4 | 0 | | |
| | **5. You are always picking on Fahn. He isn't your enemy!** | 9 | **13** | | |
| 2-3 | 1. Damn it to Jotunheim, Fahn isn't right for you. You should stay away from him. Did you know that Fahn gave Mara an onyx jewel? | 3 | 6 | | |
| | **2. Fahn isn't right for you. You should stay away from him. Did you know that Fahn gave Mara a damn onyx jewel?** | 9 | **11** | | |
| | **3. Oh Gods, Fahn isn't right for you. You should stay away from him. Did you know that Fahn gave Mara an onyx jewel?** | **12** | 8 | | |
| | 4. Okay, Fahn isn't right for you, is he? You should stay away from him. Did you know that Fahn gave Mara an onyx jewel? | 1 | 0 | | |
| | 5. Yeah, Fahn isn't right for you. You should stay away from him. Did you know that Fahn gave Mara an onyx jewel? | 5 | 4 | | |
| 2-4 | **1. By the world of ice and death, what do you think about the fact that Fahn has run away?** | **13** | 4 | | |
| | 2. Um... what do you think about the fact that Fahn has run away? | 7 | 8 | | |
| | 3. Yeah, what do you think about the fact that Fahn has run away, alright? | 2 | 1 | | |
| | **4. Yeah, what do you think about the fact that Fahn has run away?** | 8 | **16** | | |
| 4-1 | 1. Do you remember when we first met buddy? | 1 | 2 | | |
| | 2. Do you remember when we first met mate? We were so young. We didn't know how to act at court, we didn't know how to act at court! | 7 | 6 | | |
| | 3. Oh gosh do you remember when we first met? | 2 | 1 | | |
| | 4. Oh gosh do you remember when we first met? We were so young. We didn't know how to act at court friend. | 3 | 7 | | |
| | **5. Oh gosh do you remember when we first met? We were so young. We didn't know how to act at court.** | **16** | **14** | | |
| 4-2 | 1. Yeah, err... I remember, wouldn't you? | 0 | 2 | | |
| | 2. Yeah, I remember, I remember. | 2 | 4 | | |
| | **3. Yeah, I remember. I am somewhat so glad we have learned the ways of the royals. The Queen has taken notice of my progress.** | **23** | **21** | | |
| | 4. Yeah, is that so? I am also impressed you have grown much, I am also impressed you have grown much. | 3 | 1 | | |
| | 5. Yeah, ok, I remember. | 2 | 1 | | |
| 4-3 | **1. Is that so? I am also impressed you have grown much! I have come to like you quite a lot friend.** | 5 | **12** | | |
| | 2. Oh God yeah, is that so? I am also impressed you have grown much. I have come to like you quite a lot. | 6 | 3 | | |
| | 3. Oh gosh yeah, is that so? I am also impressed you have grown much. I have come to like you quite a lot, I have come to like you quite a lot. | 1 | 0 | | |
| | **4. Yeah, is that so? I am also impressed you have grown much! I have come to like you quite a lot.** | **12** | 6 | | |
| | 5. Yeah, is that so? I am also impressed you have grown much. I have come to like you quite a lot! | 6 | 8 | | |
| 4-4 | 1. Err... yeah, I feel the same way, wouldn't you? | 2 | 1 | | |
| | **2. I feel the same way!** | **14** | **19** | | |
| | 3. Oh God yeah, I feel the same way. | 4 | 3 | | |
| | 4. Ok, yeah, I feel the same way. | 2 | 0 | | |
| | 5. Yeah, I feel the same way. | 8 | 6 | | |
| 4-5 | 1. Oh gosh would you accept like, this onyx jewel for your hair? | 6 | 1 | | |
| | 2. Oh gosh would you accept this onyx jewel for your hair mate? | 1 | 0 | | |
| | **3. Oh gosh would you accept this onyx jewel for your hair?** | **18** | **25** | | |
| | 4. Oh gosh would you accept this quite onyx jewel for your hair? | 5 | 3 | | |
| 4-6 | 1. Oh God thank you. That is quite sweet of you, okay? | 3 | 3 | | |
| | 2. Yeah, ok, thank you, okay? | 1 | 0 | | |
| | 3. Yeah, thank you. That is quite sweet of you. | 2 | 5 | | |
| | **4. Yeah, thank you. That is somewhat sweet of you. It matches my eyes.** | **24** | **21** | | |

[1] NE = nonexperts(total=30); E = experts(total=29)

[2] Boldface indicates first choice of NE and/or E

Table 3: Naturalness of NLG generated dialog

| Naturalness | Expert | Non-expert |
|---|---|---|
| Very natural | 0.0% | 6.7% |
| Natural | 3.4% | 30.0% |
| Somewhat natural | 34.5% | 53.3% |
| Unnatural | 51.7% | 6.7% |
| Very unnatural | 10.3% | 3.3% |

Table 4: Different measures among the two groups

| Measure | E | NE |
|---|---|---|
| 1. Average authored dialog (in words) | 17.56 | 15.58 |
| 2. Levenshtein distance (in words) between NLG produced and author edited dialog | 7.83 | 7.70 |
| 3. Average number of unique words removed from the NLG produced utterance | 4.84 | 5.27 |
| 4. Average number of unique words added to NLG produced utterance | 3.90 | 3.48 |

E = experts; NE = non-experts

Table 5: Helpfulness of NLG dialog among experts

| Helpfulness | Experts |
|---|---|
| Yes | 52.4% |
| To some extent | 33.3% |
| No | 14.3% |

*much easier to modify the given machine dialogue than to come up with my own (though I hope my dialogue wound up more natural)"*.

Last but not least, Table 2 shows the distribution of the NLG utterance selection tasks. Of the 10 scene-tasks, six have the same most-selected utterance for experts and non-experts: 2-1, 4-1, 4-2, 4-4, 4-5 and 4-6. There seems to be a preference for less-decorated utterances, though they are not necessary the shortest sentence and not at the sacrifice of information. For example, in scene-task 4-2, both experts and non-experts picked the utterance *Yeah, I remember. I am somewhat so glad we have learned the ways of the royals. The Queen has taken notice of my progress.* It is the longest sentence with the most information, but without decorations such as *err* (utterance 1) or repetition (utterance 2, 4). While utterance 5 is short and non-decorated, it probably did not convey any useful information for the given scene-task. In future work we plan to examine the stylistic models for NLG with the hope of identifying patterns to the selected choices.

## Discussion

We present the results of an experiment testing whether automatically generated dialog can be used by either naive or expert authors to increase the number of possible story paths. Our results show that NLG dialog is acceptable to authors, but they also suggest a number of ways in which the NLG output could be improved. Two related things we noticed are the choice of less decorated utterances by expert writers and the need to customize interjections and adornments used in generation to the fantasy language. This issue will be addressed in future work. We plan to extend generic PERSONAGE hedges and emphasizers vocabulary by mining texts of a fantasy genre.

One of the limitations of the presented work is the shortage of alternative selections of content to satisfy same dialogic goal. PERSONAGE enforces a particular structure on its content plans. Content plans are required to include SATELLITE information. Satellites can be added to the main proposition without changing the primary dialog goal and produce utterances with widely differing degrees of verbosity. To date, our story world has not been rich enough to provide the content needed for satellite propositions. This limited the expressivity of PERSONAGE and prevented its utterance aggregation operations from being fully utilized. In future work we hope to use methods described in (Rishes et al. 2013) to develop richer story worlds.

In future work we also aim to compare the differences between generated and authored utterances in more detail, and to develop post-generation editing rules that can reproduce the edits that the human subjects have made. We also aim to test whether the quality of the dialog produced through the revision of automatically generated content can equal the quality of dialog authored from scratch. We plan a follow-up experiment that involves collecting quality comparisons judgments from experts of hand-crafted dialog and human edited version of NLG dialog.

## Acknowledgements

Thanks to Jessica Zamorra, April Grow, Stephanie Lukin, and Ricky Grant for their contributions to the Character Creator team. This work was supported by NSF IIS-#1002921 Pilot project funded by the NSF Creative IT program.

## References

André, E.; Rist, T.; van Mulken, S.; Klesen, M.; and Baldes, S. 2000. The automated design of believable dialogues for animated presentation teams. *Embodied conversational agents* 220–255.

Beskow, J.; Cerrato, L.; Granström, B.; House, D.; Nordenberg, M.; Nordstrand, M.; and Svanfeldt, G. 2004. Expressive animated agents for affective dialogue systems. In *Proc. of the Tutorial and Research Workshop on Affective Dialogue Systems (ADS'04)*, 301–304.

Cavazza, M., and Charles, F. 2005. Dialogue generation in character-based interactive storytelling. In *Proc. of the First AI and Interactive Digital Entertainment Conference*, AIIDE '05. AAAI.

Lin, G. I., and Walker, M. A. 2011. All the world's a stage: Learning character models from film. In *Proc. of the Seventh AI and Interactive Digital Entertainment Conference*, AIIDE '11. AAAI.

Mairesse, F., and Walker, M. A. 2011. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*.

Rishes, E.; Lukin, S.; Elson, D.; and Walker, M. 2013. Generating different story tellings from semantic representations of narrative. In *International Conference on Interactive Digital Storytelling, ICIDS'13*.

Rowe, J.; Ha, E.; and Lester, J. 2008. Archetype-Driven Character Dialogue Generation for Interactive Narrative. In *Intelligent Virtual Agents*, 45–58. Springer.