

# Analysis of Japanese Folktales for the Purpose of Story Generation

Takaaki Kato<sup>1</sup>, Takenori Wama<sup>1</sup>, Koji Miyazaki<sup>1</sup>, Ryohei Nakatsu<sup>2</sup>

<sup>1</sup> Kwansei Gakuin University, School of Science and Technology  
2-1 Gakuen, Sanda, Hyogo, 669-1337 Japan  
{bqz61917, scbc0057, miyazaki}@kwansei.ac.jp

<sup>2</sup>National University of Singapore, Interactive & Digital Media Institute  
Blk E3A, #02-04, 7 Engineering Drive 1, Singapore 117574  
idmdir@nus.edu.sg

**Abstract.** We are developing an interactive folktale system in which users can enjoy both the generation of Japanese folktales as well as interaction with the system. In order to generate any kind of Japanese folktale, the system must give each character appearing in the folktales the ability to achieve a sufficient number of motions. To determine the number of motions that need to be prepared for the system, we carried out an analysis of the verbs that appear in Japanese folktales. Based on the analysis of 50 representative Japanese folktales, we found that the total number of verbs used is more than 2,000. By deleting the overlap of verbs in different folktales, while giving consideration to their meanings in actual usage, the number decreased to about 900. In addition, by restricting the verbs to those that can be expressed by computer graphics, the total number of verbs further decreased to about 100. Then we carried out experiments on generating various scenes of Japanese folktales by using several motion sets, each of which is a subset of the 100 motions corresponding to the reduced verbs. Finally, we evaluated the least number of animations needed to generate animations of sufficient quality.

**Keywords:** folktale, story generation, verb, animation

## 1 Introduction

The authors are developing an interactive story generation system [1][2]. We adopted Japanese folktales as a sample genre for story generation. Each culture has its own folktales, and children grow up listening to their folktales or watching animations of folktales on TV or in movies. Consequently, folktales have been playing an important role in the education of children in every culture. Naturally, this holds true for the Japanese folktales recounted to the children of Japan. Therefore, we can expect huge educational potential from an interactive system that, going beyond watching or listening to folktales, gives joy to children through interaction with the folktales' characters. When we consider children as the users of such an interactive system, the story should be enchanting enough to keep the interest of children. At the same time,

the quality of computer graphics is another key issue for such a system to be accepted by people. In particular, the types of motions characters could achieve in such an interactive system have a strong impact on the impressions held by children. Therefore, an evaluation of such an interactive system, from the viewpoint of how the system can entertain children, should take these factors into consideration.

When the system generates interactive stories with computer graphics of good quality, it is necessary for the system to prepare an adequate amount of motion data in advance to generate any kind of action needed by the characters in a story. Since verbs are used in stories to reflect action, it is worthwhile to analyze the verbs used in Japanese folktales to determine what kinds of verbs and how many verbs appear in them.

Based on the above considerations, we analyzed 50 representative Japanese folktales. In this paper, we describe the details of our analysis. The results would be helpful in determining the maximum number of character motions that need to be prepared for the motion database of an interactive folktale system. Another interesting issue would be clarifying the least number of character motions that need to be prepared for the generation of Japanese folktale animations having adequate quality. To determine this, we carried out an evaluation experiment. Based on the analysis of verbs in Japanese folktales, we have prepared several motion sets that each have a different number of motions for the characters in the folktales. Then we asked subjects to watch multiple animations, generated using each of these motion sets, and to evaluate each animation. We also describe the details of the experiments.

## 2 Related Works

An interactive system, and especially one with a story, has a strong power to attract people [3]. This has been proved by the fact that video games have been so successful. Furthermore, it is expected that this power of interactive stories could be applied to education or e-learning. Based on this expectation, we are developing the Interactive Folktale System [1][2]. Several scenes generated by the system are shown in Fig. 1.



Fig. 1. Examples of scenes in Interactive Folktale System

One of the most important issues for interactive stories is controlling their development, which has been actively studied as “Interactive Storytelling”

[3][4][5][6]. The key issues in interactive storytelling include the generation of autonomous interactive characters and the emergence of storylines based on interactions between virtual characters and users, as well as those among the virtual characters. This is undoubtedly the final target of interactive storytelling. The problem with present interactive storytelling is that, since the generation of sophisticated autonomous characters is so complicated, it is difficult to maintain the consistency of the generated story for more than several minutes [7]. On the other hand, in most of these works all of the necessary motions of the characters in the system have been prepared beforehand and stored in the system, and little attention has been paid to the types and numbers of motions characters could achieve. As the number of stories and the length of the stories increase, this becomes an important issue. Therefore, it is important to study what types of motions and how many motions need to be prepared beforehand. However, so far few studies have investigated this topic.

For story generation, it is important to study the basic structure of stories. One of the most famous studies in this area is that by Vladimir Propp [8]. He analyzed more than 100 Russian magical folktales and revealed several interesting facts. One is that all the story lines consist of concatenations of sort stories. Our Interactive Folktale System is constructed based on this basic concept [1][2]. Once we could prepare motion database that contains enough number of motion animation sets, the system can generate any kinds of Japanese folktales.

### **3 Methodology of Analysis**

Various books have been published in Japan on Japanese Folktales. Among them, we selected a book as a reference source that contains 100 representative Japanese folktales [9]. A variety of folktale animations have been developed and broadcasted on TV in Japan based on this book, so we considered it a valid basis for our analysis. Fifty representative folktales were chosen from these one hundred folktales. In our analysis method, we extracted all of the verbs appearing in these 50 folktales and then counted the instances of each while considering various factors of their use. In the following, we describe the procedure followed in this analysis.

#### **3.1 Step 1 Verb extraction**

As a first step, all of the verbs that appear in these 50 folktales are extracted. If the same verbs appear in one story, we merge these verbs and count them as only one instance.

#### **3.2 Step 2 Deletion of overlaps**

In the next step, if the same verbs appear in different stories, we also merge them and count them as only one verb. For example, the same verb “say” appears in different Japanese folktales such as “Urashima-Taro (Fisherman Urashima)” and “Momo-Taro (Peach Boy),” so we merge these usages.

### **3.3 Step 3 Deletion of overlaps taking *meaning* into consideration**

We then carry out the merging of verbs with the same meaning. For example, such verbs as 'speak' and 'talk' or 'go' and 'come' would be expressed by the same kind of CG animations. Therefore, these verbs can be merged.

### **3.4 Step 4 Classification of verbs into several categories**

As the next step, we consider more deeply the meaning of each verb. Here, we try to classify all of the extracted verbs into several categories from the viewpoint of meaning. For this, we prepared six categories: (1) verbs corresponding to action, (2) verbs corresponding to situation, (3) verbs corresponding to action depending on objects, (4) verbs corresponding to emotion, (5) verbs describing natural phenomena, and (6) other verbs. For example, the verb 'go' is classified into the first category. Such verbs as 'think' and 'decide,' corresponding to mental behaviors, are classified into the second category. Such verbs as 'hold' or 'hit' are classified into the third category. Since actions based on computer graphics would vary depending on the object of such verbs, we prepared this category. For example, in the case of 'hold,' computer-graphics animations would vary depending on whether this means holding a light object or a heavy one, a large object or a small one, etc. Verbs describing emotional states such as 'get angry' or 'joyful' are classified into the fourth category. Such verbs as 'rain' or 'snow' that are used to describe natural events or phenomena are classified into the fifth category. The last category was prepared to include those verbs that cannot be classified into the above categories. For example, such verbs as 'stop doing something' or 'do not do something' are classified into this category.

### **3.5 Step 5 Detailed analysis and reduction of verbs considering computer graphics expression**

Based on the results of step 4, we carry out more detailed analysis of the verbs. Since the first category is most important when preparing CG-based animation, we focus on the verbs in this category. This type of verb can be classified into two subcategories: verbs expressing the movement of a character and verbs expressing actions it takes toward objects or other characters.

The former contains such verbs as 'return,' 'accompany,' 'come back,' and so on. These verbs can be expressed by computer graphics in such actions as walk, run, swim, and fly. Since all of these verbs imply actions of movement from one place to another, it is not necessary to prepare special animations such as 'return' or 'come.' Instead, we need only prepare the animations corresponding to each mode of movement, such as walk, run, and so on. Now, let us consider the latter cases. Here, it is necessary to prepare a special animation for each of the actions such as cook and sew. However, if a character does action such as moving around in a kitchen, we can easily imagine it is cooking. Furthermore, in the case of sewing, if a character already has a needle and a cloth, a simple action such as moving the hands in some way would be close enough.

At the same time, what the CG animated characters need to do depends on the particular model of a character. If the model were simple, the animations the character would be able to do would be significantly restricted. However, if we reduced the number of animations too much, then the actions of the characters would become unnatural. This means that the kinds of animations we have to prepare for CG characters depend on the characteristics of the characters as well as the kind of stories we have to regenerate. Consequently, we have to consider what would be the least acceptable number and fewest kinds of character actions in the case of Japanese folktales.

Taking these issues into consideration, we carefully checked the verbs in the first category (verbs corresponding to actions) and reduced the number of verbs by merging them. We then prepared computer graphics animations corresponding to each of the merged verbs.

### 3 Analysis Results

Based on the methodology described in the previous chapter, we analyzed the verbs that appear in the 50 representative Japanese folktales [9]. Here we describe the analysis results. Figure 2 illustrates the result of step 1 (results of original verb data), step 2 (result after simple overlap deletion), and step 3 (result of deletion of overlap taking meaning into consideration). The horizontal axis corresponds to the number of folktales and the vertical axis corresponds to the accumulated number of verbs. The accumulated number of verbs in each step is 2324, 1613, and 884. Figures 3, 4, 5, and 6 illustrate the accumulated number of verbs in each of four verb categories, i.e., action verb, object-dependent verb, situation-based verb, and emotion-related verb. The total number of verbs corresponding to each of these categories is 306, 137, 202, and 64, respectively.

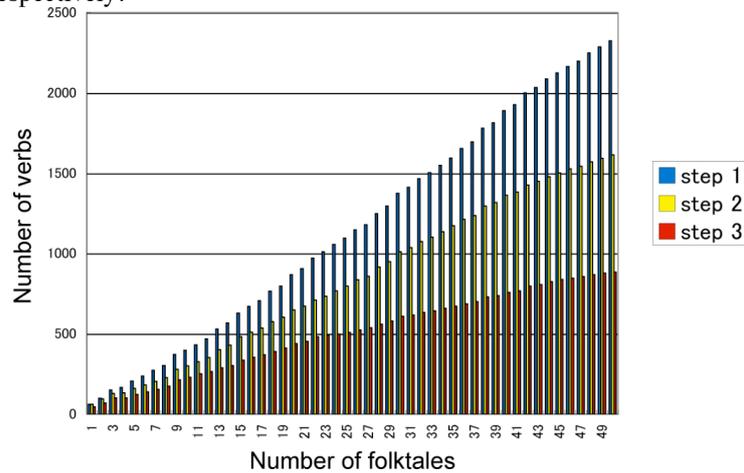
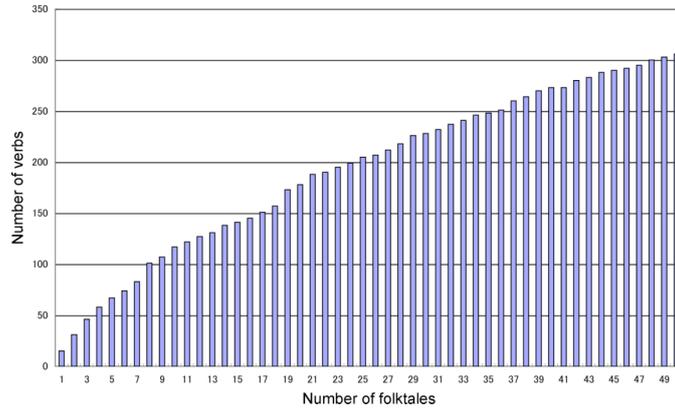
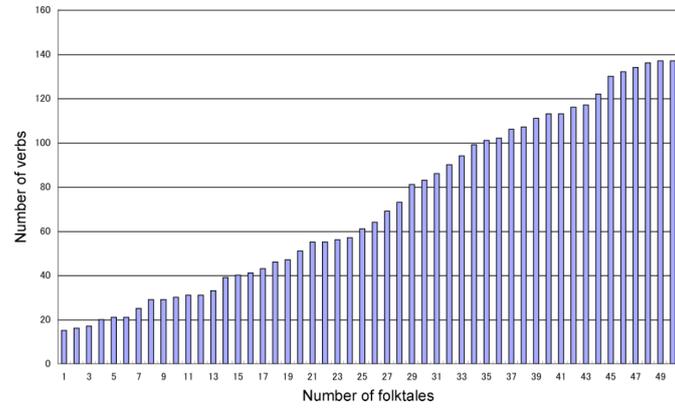


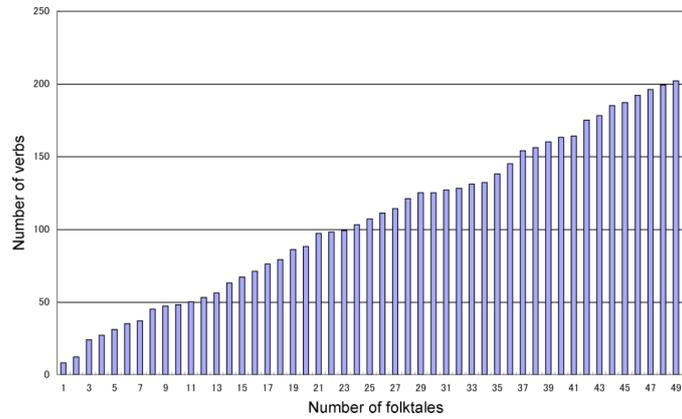
Fig. 2. Number of folktales vs. number of verbs.



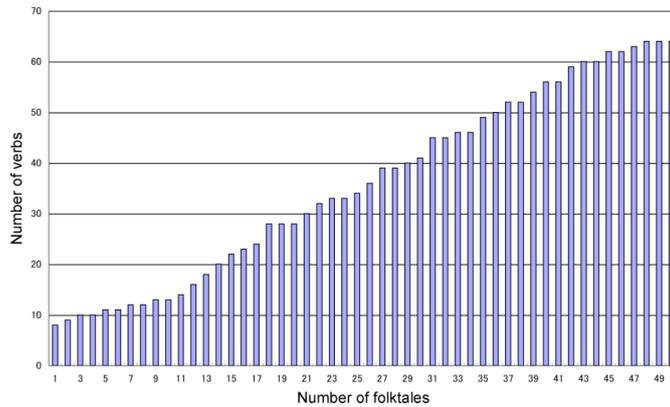
**Fig. 3.** Number of folktales vs. number of verbs (action verbs).



**Fig. 4.** Number of folktales vs. number of verbs (object-dependent verbs).

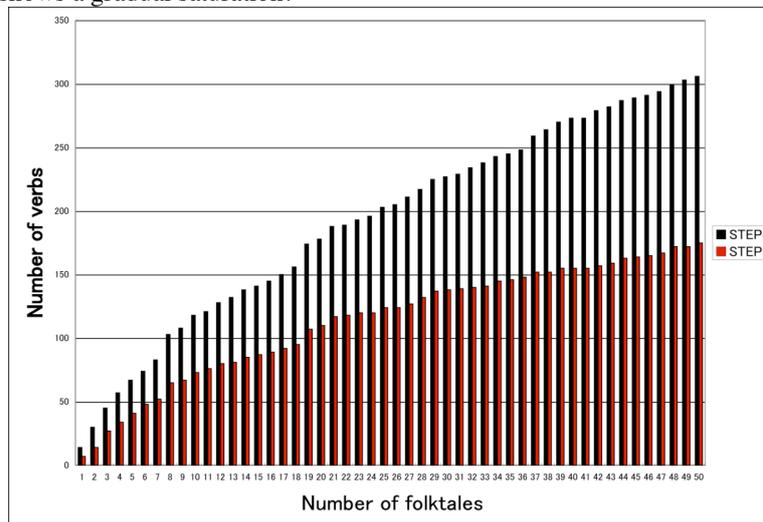


**Fig. 5.** Number of folktales vs. number of verbs (situation-based verbs).



**Fig. 6.** Number of folktales vs. number of verbs (emotion-related verbs).

In Fig. 2, the trend before deleting overlap shows a simple increase. On the other hand, after step 2, the increase ratio became smaller, and after step 3, a saturation tendency became visible. Figures 3–6 show how the number of verbs in each category increases depending on the number of folktales. Although we have to further investigate the trends in these verbs by increasing the number of analyzed folktales, each graph already shows a gradual saturation.



**Fig. 7.** Number of folktales vs. number of verbs (Step 5).

Figure 7 illustrates the results of step 5. The number of verbs after step 5 is reduced to 175. Moreover, the effect of saturation became clearer. This graph shows that about 170–200 actions should be prepared as computer graphics animations in order to generate a sufficient number of Japanese folktales.

## 5 Evaluation Experiment

Based on the results of the above analysis, we carried out evaluation experiments to determine how many character actions would be sufficient to generate animations of Japanese folktales with high enough quality. We selected two representative stories from the 50 stories (story 1 and story 2) to generate as computer graphics animations. Among 170 total verbs, only about 50 verbs appear in these two stories. Among these 50 verbs, we carefully extracted ten, twenty, and forty representative verbs and called them group 1, group 2, and group 3, respectively. Story 1 consists of eleven scenes, where seven of them are generated using the three types of motion sets described above. Story 2 consists of thirteen scenes, ten of which are generated using the three motion groups. Figure 8 shows a comparison of different motions in the same scene. In Fig. 8 (a), the scene shows the character of an old man running. In the case of Fig. 8 (b), however, the motion “run” is not included in the motion group, so the motion “walk” is used instead of run.



Fig. 8. Comparison of different motions in the same scene.

### 5.1 Evaluation experiment 1

In this experiment, we evaluated each scene generated using the three motion sets from the viewpoints of naturalness and fun. Ten subjects, aged around twenty years, were selected and asked to watch three types of generated scenes shown in random order. Then they were asked to evaluate these animations from the viewpoints of “naturalness” and “fun.” The evaluation scores consist of five rankings; 1: worst, 2: worse, 3: neutral, 4: better, and 5: best. The results of the evaluation test are shown in Table 1.

Table 1. Results of evaluation experiment 1

| Evaluation item | Group 1 (10 motions) |      | Group 2 (20 motions) |      | Group 3 (40 motions) |      |
|-----------------|----------------------|------|----------------------|------|----------------------|------|
|                 | Naturalness          | Fun  | Naturalness          | Fun  | Naturalness          | Fun  |
| Mean value      | 2.40                 | 2.61 | 2.71                 | 2.85 | 2.81                 | 2.93 |
| Variance        | 0.16                 | 0.10 | 0.28                 | 0.32 | 0.12                 | 0.18 |

These results show that as the number of actions prepared as character animations increases, the evaluation scores for both naturalness and fun also increase slightly. However, the differences are not large in light of the variances. Therefore, these results show that for naturalness and fun the number of actions is not so important. On the other hand, it can be easily conjectured that for the quality of the generated animations, the story itself would have a big effect. Since our evaluation involved only short scenes, the effect of story may not have been fully expressed.

## 5.2 Evaluation experiment 2

To clarify the effect of story, we carried out a second experiment. In this case, the subjects are asked to watch the scenes generated sequentially according to the story. All of the other experimental conditions are the same. The results of the experiments are shown in Table 2.

**Table 2.** Results of evaluation experiment 2

|                 | Group 1 (10 motions) |      | Group 2 (20 motions) |      | Group 3 (40 motions) |      |
|-----------------|----------------------|------|----------------------|------|----------------------|------|
| Evaluation item | Naturalness          | Fun  | Naturalness          | Fun  | Naturalness          | Fun  |
| Mean value      | 2.13                 | 2.53 | 2.93                 | 3.07 | 3.6                  | 3.73 |
| Variance        | 0.52                 | 0.25 | 0.6                  | 0.2  | 0.91                 | 0.2  |

These results show the following. In the case of animations using group 1 motion sets, evaluation scores for both naturalness and fun are around two and thus not adequate for generating animations. For the animations using group 2, both naturalness and fun achieved scores around three. For the generation of folktales, this implies that at least twenty basic motions are necessary for maintaining the minimum acceptable level of quality. On the other hand, the animations using the group 3 motion set earn a relatively high score between three and four. This suggests that it would be possible to generate relatively good quality by using forty motions. At the same time, however, it is still necessary to carry out further experiments using larger numbers of motions to determine where quality becomes saturated.

## 6 Considerations and Future Work

Here, based on the results of the analysis and evaluation, we discuss additional considerations as well as several points that should be further studied.

As described in the previous section, the animations achieved by CG characters would depend on the characteristics of the characters themselves. If characters are based on a precise model, such as having five fingers on each hand and the capability to mimic human actions, then we need to prepare all of the actions corresponding to all of the verbs. On the other hand, however, if the models of the characters are simple enough

that they don't have all ten fingers, then the actions using fingers should be expressed in some other way.

Other than the verbs corresponding to actions, let's consider verbs corresponding to emotions. Such emotional verbs as 'cry' or 'get angry' can be expressed either by facial expressions or body motions or combinations of these. In any case, the facial expression plays the most important role. Therefore, verbs corresponding to emotions are somewhat special.

In the case of situation-based verbs corresponding to mental behaviors such as think, decide and so on, it is fairly difficult to express these verbs only by body motions. Therefore, it is necessary to introduce narrations in addition to body motions. Moreover, the motions corresponding to these actions largely depend on culture. Therefore, the preparation of animations for these verbs would require extraordinary care.

As future work, we first have to increase the number of folktales for analysis to further clarify the trend of each type of verb. In addition, we intend to carry out further evaluation experiments by using more folktales.

## 6 Conclusion

We have analyzed the verbs appearing in 50 representative Japanese folktales for the purpose of developing an automatic story-generation system. Although we need to carry out further research, so far we have found that there is some limit on the number of animations that need to be prepared for our system. This limitation will closely correspond to the particular actions of the characters in the folktales.

## References

1. Miyazaki, K., Nagai, Y., Bosser, A. G., and Nakatsu, R.: Architecture of an Authoring System to Support the Creation of Interactive Contents. Entertainment Computing – ICEC2006, Springer LNCS 4161, 165-174 (2006).
2. Miyazaki, K., Nagai, Y., Wama, T., and Nakatsu, R.: Concept and Construction of An Interactive Folktale System. Proceedings of ICEC2007, Springer LNCS 4740, 162-170 (2007).
3. Murray, J.: Hamlet on the Holodeck. The Free Press (1997).
4. Swartout, W. et al.: Toward the Holodeck: Integrating Graphics, Sound, Character and Story. Proceedings of the Autonomous Agents 2001 Conference (2001).
5. Mateas, M. and Stern, A.: Socially Intelligent Agents: The Human in the Loop. AAAI Fall Symposium (2000).
6. Young, R. M.: Creating Interactive Narrative Structures: The Potential for AI Approaches. AAAI Spring Symposium in Artificial Intelligence and Interactive Entertainment, AAAI Press (2000).
7. Charles, F. and Cavazza, M.: Exploring Scalability of Character-based Storytelling. Proceedings of ACM AAMAS'04 (2004).
8. Prop, V.: Morphology of the Folktale. University of Texas Press (1968).
9. 100 Japanese Folktales, Kodansha (in Japanese) (1997).