Children's Story Authoring with Propp's Morphology: An Exploratory Study

Sean Hammond Tim J. Smith Helen Pain

10th March 2008

This paper is part of an investigation of the use of Propp's morphology as a story writing tool for children, with the aim of creating a computational tool for teaching storytelling. Early quantitative results are presented from an exploratory study in which participants aged 7-11 wrote stories using a paper-and-card interface based on Propp's morphology. The general research questions are: can children apply this tool to story writing? how successfully do they use it? and what do they do with it? An experimental methodology suitable for answering these questions was devised, and can now be developed based on our experience. The quantitative results show that children were able to apply the tool to story writing with a high degree of success. Further qualitative analysis of the results of the study will identify the adequacy of Propp's morphology and inform the design of a computational tool.

1 INTRODUCTION

The long-term objective of this research is to construct a tool to help children write stories by allowing them to view and interact with their plots at different structural levels, and keep track of the constant and variable elements. The tool requires a formal model of narrative structure, which we aim to develop based on Propp's morphology. The current study aims to empirically verify whether children are able to recast and generate stories using Propp's morphology. Previous applications of Propp's morphology in this field, e.g. Rodari (1996) and Charles (2006), have been somewhat informal and are underspecified. This study provides the first step in developing a more formal model.

In *Morphology of the Folktale* (Propp, 1968) Propp "devised a very simple method of analysing wondertales in accordance with the characters' actions—regardless of their concrete form" (Propp, 1984, p.69). Propp noticed that "if the hero leaves home in quest of something, and the object of his desires is far away, he can reach it by magic horse, eagle, flying carpet, flying ship, astride the devil etc." (p.73), but in each case the function is "the transfer of the hero to the place where the object of his search is located." The tale consists of both constants and variables. The concrete form of the characters, their world and actions is variable. But the *function*, "an act of a character, defined from the point of view of its significance for the course of the action" (Propp, 1968, p.21) is constant.[1] "This explains the two-fold quality of a tale: its amazing multiformity, picturesqueness, and color, and on the other hand, its no less striking uniformity, its repetition." (p.20)

Propp's thesis rests on four basic observations. First, that functions are the "stable, constant elements in a tale" (Propp, 1968, p.21). Second, the number of functions in his corpus was "startlingly small," 31 functions in total. Third, Propp discovered that although all tales do not contain all functions, "the sequence of functions is always identical" (p.22). Fourth, Propp concludes that all tales studied are of a single structural type.

1 This definition of function has two aspects. As well as the action itself, the position of a function relative to other functions in the tale (what Levi-Strauss called the context of the function (Levi-Strauss, 1963)) also forms part of its definition. "An action cannot be defined apart from its place in the course of narration"(Propp, 1968, p.21). We have seen that different actions can instantiate the same function. In some cases different functions may be instantiated by the same action if the contexts of the action differ.

Propp's morphology has seen many applications beyond its intended use as a system for classifying folktales. This study builds on two examples of Propp's morphology applied to creative story writing with children. Rodari, 1996 describes several games for use in schools that use popular folk tales as raw material to invent new stories. A game which Rodari calls "recasting" a fairy tale (Rodari, 1996, p.39) creates new tales out of old ones by revealing the underlying structure of the tales. Recasting a story is a three step process: first the tale is reduced to a bare synopsis of the plot events, then the names of characters, places etc. are replaced by signs, reducing the plot to an abstract formula. Finally the abstract expression is reinterpreted as a new story. For another game Rodari created a set of playing cards ("The Cards of Propp") each marked with the title and an illustration of one of Propp's functions, which were used in story writing games in which the story must be structured according to the cards. Rodari's experiment was a success, the children easily succeeded in creating fairy tales that followed the cards (Rodari, 1996, p.47).

A similar use of Propp's morphology by Charles (2006) aimed to empower children to create complete, original fairy tales, to "explore whether children could take ownership of folktales by recreating them," (Charles, 2006, p.46), imbuing them with personal meaning and using them to rehearse real-life confrontations. Charles developed a prototype teaching tool that combines Rodari's ideas of recasting and the cards of Propp. The Propp functions of three fairy tales were each represented by a series of Propp's cards called a "story map," and the story maps were used to reveal the structure of the sample tales to recast them. Each session began with a telling of one of the sample stories before the story was collaboratively recast by a group of children facilitated by Charles, who helped the children follow the story map as well as encouraging their ideas. Charles found that "in all sessions, original and different stories were created that reflected the childrens' personal lives, the places they lived, and their knowledge of the world. The results of the study proved that contemporary children took ownership of folktales when provided with permission, opportunity, and guidance" (Charles, 2006, p.iv).

As Rodari's work is not an empirical study it presents no empirical evidence of the effectiveness of the Propp's cards, such as stories written by the children. Charles uses data from tape recordings of the class stories, children's drawings of the stories, and children's recollections of the stories three weeks later. But the stories in Charles' study were created by the class, with an adult facilitator. Charles studied the recasting process with Propp's cards, and did not explore other interesting processes, such as the generation of new story maps by the participants.

First, our study aims to empirically verify and clarify in greater detail whether children are able to recast and generate stories using Propp's morphology. Because we aim to develop a formal model for implementation, we find that the models presented by Rodari and Charles are underspecified. Quantitative and qualitative data was collected on each individual participant and on each individual function, giving us a wealth of empirical data from which to develop a formal model. The use of each function will be analysed to identify the more difficult functions and provide suggestions on how their use may be facilitated.

Second, our study aims to investigate the potential of the model beyond the process of recasting a story. The second part of our study concerns the process of generating new plot structures, as opposed to recasting existing plots.

The study illustrates a prototype methodology for testing the story model. Further studies will develop the methodology based on our findings. This paper presents early, quantitative results from the study. It considers the question of whether the children were able to apply Propp's morphology to story writing. Setting aside the quality or originality of the stories for now, our quantitative question is, how *accurate* are the stories? How closely do the stories produced match the Propp functions used?

2 METHODS

A two-part study was carried out across two weeks at Collingwood Primary School, Newcastle. The first week, recasting, investigated the process of recasting stories using Propp's cards as story maps. The second week, generating, investigated the process of constructing new story maps without a sample story, and using them to write stories.

2.1 *Participants*

42 students of ages 7-11 participated in the first week of the study. 35 of these participants, aged 7-10, returned to participate in the second week. Signed informed consent forms were collected from the students and their guardians, confirming that they understood and accepted the conditions of the study, including the collection of video and audio data.

2.2 *Materials*

The three sample stories from Charles were used as stories to recast in the first week of the study. Cards for each of Propp's functions were developed with the name of the function, a brief description, and a picture.[2] Large copies of the Propp cards were used to make story maps representing the structures of the sample stories. Smaller copies with space for writing were used as writing materials for the participants, who were asked to write each function of their story on the card representing that function.

2.3 *Procedures*

The materials and recording equipment were setup in the school's library, and participants visited the library in groups of five at a time to participate in one to two hour story writing sessions. At the beginning of each session the participants were told that we were testing the idea of story maps for writing stories, and needed their opinions and ideas. They were reminded that it was the story maps, and not themselves, that were being tested.

WEEK ONE, RECASTING STORIES First, the teacher read out one of the sample stories while the students listened. The researcher then showed the participants the story map for the story they had heard, and explained which part of the story each card on the map represented.[3] The participants were then told that it was time to write their own stories that would follow the map of the story cards. They were told that they could write about whatever they liked, using their own characters, places and actions. A short group discussion was used to generate ideas for stories.[4] Finally the children were asked to begin writing when they were ready.

Three additional groups participated in sessions that used the facilitated, collaborative approach described by Charles (2006), in which the participants and the researcher recast the story together. No quantitative data was collected during these sessions, but the experience allowed for reflection on the process.

WEEK TWO, GENERATING STORIES The participants were reminded of the story cards and story maps concepts from the previous week. They were shown the complete selection of Propp's cards and given rules for choosing the cards for their story maps.[5] After a discussion of themes and keys to generate ideas, the participants began choosing their cards and writing their stories.

One group in week two participated in the alternative, collaborative approach instead of individual story writing.

DATA COLLECTION Data was collected from several sources in all sessions. All participants filled out cover sheets with their name, age and sex. Audio and video of each session was recorded by a

2 Simpler and sometimes modernised names and descriptions had to be developed for the functions, as Propp's work (translated from Russian) uses many difficult words: interdiction, reconnaissance, complicity, mediation, transfiguration, etc. Our replacements were based on those used by Charles. The cards were reviewed by the teacher, who suggested improvements before the study.

3 We found it effective to explain the story map interactively, so that the participants did not get bored. For each card participants were asked "Who knows what this card means?" or "Who knows what part of the story this card stands for?"

4 The discussion revolved around potential theme elements for the stories, similar to Rodari's "keys" of time and place (Rodari, 1996, p.51): a time or place, a character, a genre, a magical object or power.

5 The rules used were: The Choosing Rule: you choose which cards to use for your story map, The Numbers Rule: you must use the cards in the right order, The Pairs Rule: if you choose the first card of a pair you must choose the second card of that pair. Sessions were tried with and without each of these rules. The Pairs Rule refers to the pairs of functions identified by Propp, for which the second function always follows if the first function is found.

fixed-position camera focused on the table on which the participants were placing and writing on their Propp cards. A dictaphone was used to record participants' responses to emergent questions from the researcher. Participants were asked to explain their stories, asked about observations made by the researcher, or asked about their opinions of the story cards: what they liked or did not like, what they found easy or difficult. The researcher made paper and pencil notes of observations during each session which were written up at the end of each day. At the end of each session each participant's Propp cards containing the story written by the participant were collected.

Information about the school and its teaching methods for story writing was collected from the school's OFSTED report and interviews with the children and their teachers.

2.4 *Analysis*

The quantitative analysis focuses on one source of data: the written stories from the participants, and asks: how closely do the stories match the Propp functions of the cards used? To measure the accuracy of the stories, we developed a means of quantifying the written text. The problem was to find a consistent and reliable way of deriving the sequence of Propp functions from a story. The sequence could then be analysed quantitatively.

We developed an interpretation of Propp's functions that was more formal than the descriptions given by Propp. Because Propp's functions are highly abstract, identifying instances based on definitions of the functions is a difficult and interpretive analytical task. Instead we identify instances using the mid-level of abstraction given by Propp's groups, which lie between functions and concrete examples, related to the functions "as *species* to *genus*" (Propp, 1968, p.25). We developed an initial set of formal definitions of the species from Propp's work,[6] then refined this interpretation based on examples in our data.[7] We define an instance of a function as an instance of any species belonging to the function.

With this set of definitions as the materials for coding, we developed a process of analysis that produces results similar to the examples of completed analyses given by Propp (Propp, 1968, Chapter IX B and Appendix II) and also reflects the two-step process described by Rodari for reducing a tale to its abstract formula before recasting. It was found to be helpful to read a story from start to finish once to get an overview before beginning the analysis:

1. Reduce the writing on each Propp card to its bare plot events. Note which characters act in each event. This produces a plot synopsis of the story in which the events are connected to their Propp cards.

2. Considering this synopsis, decide which Propp role(s) are filled by each character. Knowledge of the roles of each character helps to identify the species, as roles form part of the definition of some species.

3. Decide which species each plot event instances. A plot event may instance zero-or-more species, and may instance multiple species of the same function or species of different functions.

From these steps the sequence of species, and therefore functions, contained in the story is derived, and each species is connected to the Propp card on which it is written. Given this data, quantitative analysis can finally be performed.

To measure the fit between the sequences of functions in the stories and the sequences of functions on the Propp cards two specific measures were used, a lower-bound (under-estimate) and an upper-bound (over-estimate). The lower-bound measure asks: how often is the function represented by a Propp card instanced on that card? This measure suffers from false negatives, as the story may contain the right

6 For some functions we found that the species given by Propp do not lie at the correct, mid-level of abstraction between functions and concrete instances, but are much closer to the concrete level of instances or varieties of instances. For Levi-Strauss this inconsistency is the core weakness of Propp's morphology. Levi-Strauss criticised Propp for unwittingly reintroducing aspects of concrete content into his species categories, particularly in the twenty-two species of the crucial function *Villainy*: "A drawer filled with unclassified forms does not constitute a species" (Levi-Strauss, 1963, p.179). In such cases we group the species given by Propp under new, more abstract types, deriving a new set of species for the function at the most useful level of abstraction.

7 Where a clear example of a function appears in the data that does not conform to any of our species for the function, either a species was revised or a new species added to the function.

sequence of functions without every function being on the right card. (For example, if the author has used one card to write several functions.)

The upper-bound measure asks: how often is the function represented by a Propp card instanced on any of the Propp cards of the story? This measure will count the instances missed by the lower-bound measure, but suffers from false positives because it does not account for the order of the sequences. (An anomalous instance of a function may be counted if it happens to match a card elsewhere in the story.) Both measures ignore additional functions instanced in a story that are not represented by any of the cards of the story.

These two simple measures provide lower- and upper-bounds and offer a good indication for the degree of fit between the sequences of functions represented by the Propp cards used and the sequences of functions implemented in the stories written.

3 RESULTS & DISCUSSION

3.1 *How accurately did the participants recast the sample stories in week one?*

The lower- and upper-bound accuracy measures were applied to 35 stories[8] coded from week one (table 1 and figure 1). The majority of students were able to successfully recast the sample stories using the Propp cards (upper-bound mean 90.27%). The lower-bound measure is significantly lower (74.89%), and with greater variance, than the upper-bound.

3.2 *How accurately did the participants construct stories in week two?*

The lower- and upper-bound measures were applied to 35 stories coded from week two (table 1 and figure 1). Again, the stories fit the Propp cards accurately with only a few exceptions (upper-bound mean: 79.28%). The difference between the lower- and upper-bound is much smaller in week two than in week one.

In both weeks, no obvious pattern (age, sample story used, day or session, etc.) identifies the few participants with low accuracy measures. We note that with the exception of participant `111D` the same participants have relatively low accuracy in both weeks. It seems probable that the difference is due to individual differences between the participants.

3.3 *Why is the difference between the lower- and upper-bounds larger in week one than in week two?*

Many of the week one data show a large difference between their lower- and upper-bound measures. The mean difference between the lower- and upper-bound measure of each story in week one is 15.38% (standard deviation 10.6%). But in week two this difference is much smaller, the mean being 4.79% (standard deviation 7.63%). During the study participants were often observed writing two or three correct functions on a single Propp card, and then leaving the next one or two cards blank when they realised their mistake. The story 'runs ahead' of the story cards. In these instances the story contains the correct functions (and in the correct sequence) and they will be counted by the upper-bound measure, but because the functions are not on the right cards they will not be counted by the lower-bound measure. Our hypothesis is that these run-aheads occurred frequently in week one but not in week two, causing the lower-bound accuracy measures in week one to suffer.

To test this hypothesis, a run-ahead is counted wherever a participant has failed to implement a function on the correct card, but has implemented the function on either one of the previous two cards. By this definition 62 run-aheads occur in week one, 14 in week two.

If we add the run-aheads to the lower-bound count of each story in both weeks the differences between the lower- and upper-bounds shrink, to a mean of 3.3% in week one and 1.96% in week two (variances 5.7% and 3.55%). When run-aheads are added to the lower-bound measures, the mean for week one increases from 74.89% to 86.96%. The mean from week two changes little, from 74.49% to 77.32%.

8 Of 42 participants in week one, 35 stories were coded. 7 more stories were marked as compromised because they could not be accurately coded. No stories were compromised in week two.

	Lower-bound		Upper-bound	
	Mean	Standard deviation	Mean	Standard deviation
Week One (recasting)	74.89%	18.74%	90.27%	13.35%
Week two (generating)	74.49%	17.99%	79.28%	16.76%

Table 1: Mean per-participant accuracy measures, both weeks.

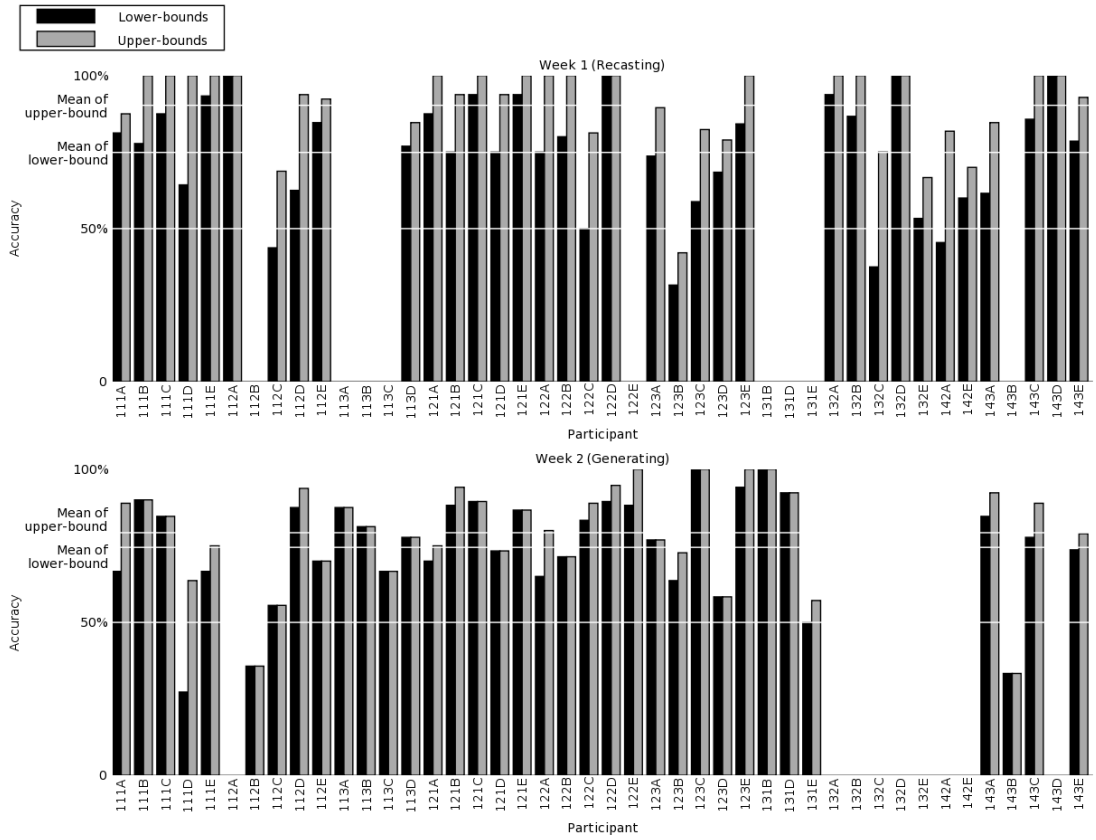

Figure 1: Lower- and upper-bound accuracy measures per participant, week one against week two

A quantitative analysis of the types of run-ahead shows that run-aheads occur most often when certain pairs of Propp cards appear in sequence in a story map, and that these sequences appear more often in the story maps from week one than in those from week two. Our hypothesis is that run-aheads occur because the author is thinking about the sample story that they are recasting, particularly at points where events are closely related one event in the sample story brings the next to mind and both are recast on the same card. Because the participants were not recasting in week two run-aheads were less frequent.

When run-aheads are added to the lower-bound counts, the means of both the lower- and the upper-bounds are about 10% higher when recasting stories than when generating stories. Although the generated stories have lower accuracy than the recast ones, the results from week two appear to confirm that children can accurately use the Propp cards to generate stories from their own story maps as well as to recast stories. But this conclusion rests on the assumption that the participants were constructing original stories in week two. The participants may instead have attempted to recast the sample stories from week one again, from memory, with slightly less accurate results than when the sample story was more fresh in their minds. A quantitative comparison of the cards used by each participant in week one and week two can only give an indication. Without taking order into account, 64% of cards chosen

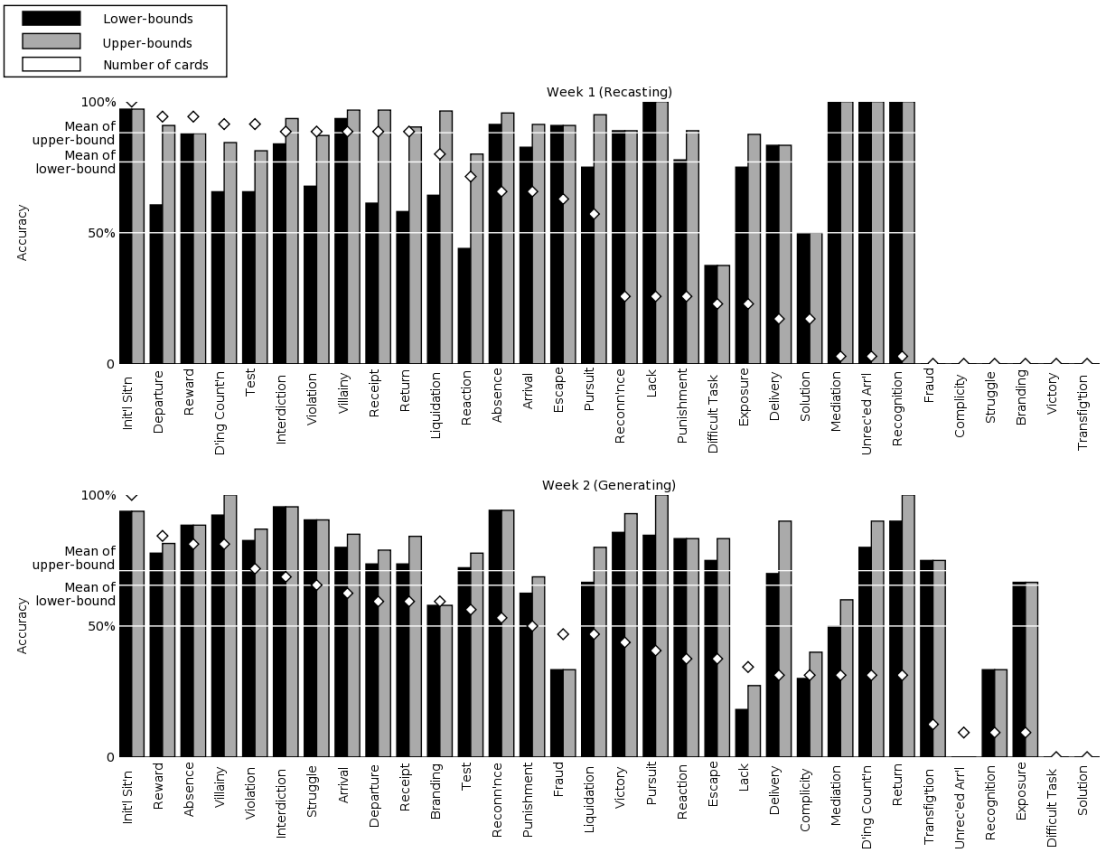

Figure 2: Lower- and upper-bound accuracy measures per function, week one against week two. The diamonds show the number of times each function was used on a scale of 0-35 (week 1) and 0-32 (week 2).

by participants in week two were used by the same participant in week one, but the overlap may be due to deliberate behaviour by the participant or it may be a feature of the Propp functions that some are more useful than others. Qualitative analysis will be needed to assess how much the first week's sessions affected the results from the second week.

3.4 *Were some functions recast more accurately than others?*

To answer this question the lower- and upper-bound accuracy measures were applied to each function (figure 2 top). When assessing these results the sample stories that were recast need to be taken into account. The instances of the functions in these stories acted as examples from which the participants recast the functions. The frequency of use of each function reflects the functions of the sample stories that were used. The difficulty of recasting of course varies depending on the story being recast. Not only did the sample stories used vary in complexity, but during the study it became apparent that two of the sample stories did not make very clean-cut use of some of the functions. As a result these two sample stories were used less frequently, and the remaining, clear story was used more often.

The lower-bound accuracy measures for many functions are much lower than the upper-bounds, because these functions were victims of run-aheads. Considering the upper-bound measures, a group of functions has both high accuracy and a high frequency of use (from left to right, Initial Situation to Pursuit). These turn out to be exactly the functions of the clearest sample story. We can conclude that, given a clear example of their use, children can recast these functions reliably and accurately. No frequently used function has a low accuracy.

The remaining functions are those that appear in one or both of the less clear sample stories but not the clearest one. Because the frequency of use of these functions is low we cannot place too much confidence in these results, but we note that all have high accuracy except Difficult Task and Solution which measure 37.5% and 50% respectively. Even without the low frequency of use, a clear conclusion could not be drawn about a function that measured low accuracy when recast from just one or two sample stories. Blame may lie with the function itself, or with an unclear example in the sample story.

3.5 *Were some functions used more accurately than others when generating stories?*

In week two the lower-bound measures for each function follow the upper-bound measures closely (figure 2 bottom), but the upper-bound measures are more variable than in week one. The frequency of use of each function in week two reflects how popular the function was with the participants.[9] Again there is a group of functions with high accuracy and high frequency. These are the same functions as in week one, with the addition of Struggle and Reconnaissance. Fraud and to a lesser extent Branding stand out as functions with a high frequency of use and low accuracy. Otherwise, functions with low accuracy have low frequency, as in week one.

4 CONCLUSION

The quantitative results verify that children are able to successfully recast stories using Propp's morphology: with few exceptions participants' stories implemented the correct functions. The same result appears to hold also for the process of generating stories using Propp's morphology, but qualitative analysis is needed to determine to what extent participants were generating new stories or repeating their recasting of the sample stories from earlier in the study. Quantitative measures do not capture the quality of the stories. It will be interesting to see how the accuracy of the stories coincides with the degree of originality and personalisation.

Further analysis will combine the quantitative results with qualitative observations and examples from the data to identify how participants use the tool. When discussed together these aspects will give context to each other and present a more thorough picture of the study.

The work so far verifies the model as a story writing tool for children, and illustrates a methodology for testing the model. Based on the experience of the study and further analysis of the data, the next step will be to develop a more formal model to be implemented in a computational tool, and to develop our methodology to test the implementation.

REFERENCES

Charles, V. (2006). Making traditional tales relevant for contemporary children. Master's thesis, York University, Toronto, Ontario.

Levi-Strauss, C. (1963). Structure and Form: Reflections on a Work by Vladimir Propp. In *Structural Anthropology*, volume 2, chapter 8. New York: Basic Books.

Propp, V. (1968). *Morphology of the Folktale*. University of Texas Press; 2nd edition.

Propp, V. (1984). The Structural and Historical Study of the Wondertale. In Liberman, A., editor, *Theory and History of Folklore*, pages 67–81. Manchester University Press. This is Propp's response to Levi-Strauss' Structure and Form: Reflections on a Work by Vladimir Propp.

Rodari, G. (1996). *The Grammar of Fantasy*. Teachers and Writers Collaborative, New York.

9 The functions unique to Propp's *second move* (Unrecognised Arrival, Recognition, Difficult Task, Solution, Exposure) are at a disadvantage because they were not used in some of the later sessions in week two.