POSTPRINT

Annotating with Propp's *Morphology of the Folktale*: reproducibility and trainability

Bernhard Fisseni

Fachbereich Mathematik, Universität Hamburg, Hamburg, Germany and
Fakultät für Geisteswissenschaften, Universität Duisburg-Essen, Essen, Germany

Aadil Kurji

Department of Philosophy, University of Bristol, Bristol, UK

Benedikt Löwe

Fachbereich Mathematik, Universität Hamburg, Hamburg, Germany, and
Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, The Netherlands

Correspondence:
Bernhard Fisseni, Universität Duisburg-Essen, Fakultät für Geisteswissenschaften, Germanistik/Linguistik
Universitätsstraße 12, 45117 Essen, Germany
E-mail:
bernhard.fisseni@uni-due.de

Abstract: We continue the study of the reproducibility of Propp's annotations from Bod et al. (2012). We present four experiments in which test subjects were taught Propp's annotation system; we conclude that Propp's system needs a significant amount of training, but that with sufficient time investment, it can be reliably trained for simple tales.

1 Introduction

In this article, we report on four experiments in which test subjects were taught Propp's formal system and assess the reproducibility of formalizations in Propp's system on the basis of our results. The results of the experiments allowed us to assess the reproducibility of formalizations in Propp's system. The article is a continuation of the research presented by Bod *et al.* (2012) where the first two of our four experiments were already discussed. In these two experiments, during which the test subjects had only a brief training session before annotating using Propp's system, we observed that the annotations had very little inter-annotator agreement and

differed considerably from Propp's own annotations. In the third and fourth experiment, we trained the test subjects much more intensively, obtained much more inter-annotator agreement, and were even able to reproduce Propp's annotations on the function level. We also had to introduce some simplifications into the framework, and focused on tales whose structure is relatively simple. We discuss the differences between the results in this article and suggest that while Propp's system is not easily trainable (as Bod *et al.*, 2012, argued), it is trainable with a sufficient investment of time and effort.

In this article, we first describe our motivation in Section 2 and then discuss related work in Section 3 and Propp's formal system in Section 4. The main

488

part of the article is Section 5, where we discuss the four experiments. Finally, we conclude in Section 6.

2 Motivation

Vladimir Propp's study *Morphology of the Folktale* was published in 1928 in Russian (second edition 1969); the English translations (1958; second edition 1968) made the work accessible in the West. The *Morphology* then inspired much work in folkloristics, literature, psychology, and artificial intelligence. Researchers in the fields of structuralism and later in the new field called *computational models of narrative* have extended the scope of the approach, and have developed and adapted the general Proppian methodology into formal and computational frameworks for the analysis, automated understanding, and generation of narratives (see, e.g. Rumelhart, 1975; Lehnert, 1981; Schank, 1982; Dyer, 1983; Turner, 1994).

In recent years, there has been an increased interest in the methodological and conceptual issues involved in the computational study of narratives. For computational applications, formal structures are needed; a formal framework for representing narratives consists of a formal language, a class of mathematical structure, and a description of a procedure (called *formalization* by Löwe, 2011) of assigning to each narrative a structure. This procedure is not a *function* in the mathematical sense, but (the result of) an activity by trained formalizers who follow given guidelines.

A formal framework will always model some structural aspects of narratives, which may differ depending on the context of usage.[1] For the framework to be useful, it is necessary that given a narrative, there is substantial intersubjective agreement on the modelled structural aspects. As a side condition, we need to assume that the cultural context of recipients is sufficiently similar to warrant a similar interpretation of the narrative. Such intersubjective agreement would imply *reproducibility*, i.e. agreement of different annotators with respect to which formal structure is assigned to a given narrative.[2] As in psychology and computational linguistics, we assume such *inter-rater (more precisely here: inter-annotator) agreement* to be an indicator of the reliability or objectivity of a formal system. Here we use *reliability* and *objectivity* as in empirical social science research (cf. e.g. Bortz and Döring, 2002); if we assume that the test subjects are the measurers, i.e. the experimenters, then inter-rater agreement measures *objectivity*, if we assume that they are providing measurements assessed by the experimenters, then inter-rater agreement measures *reliability*.

In computational linguistics, inter-annotator agreement has been studied at all linguistic levels, but no such analysis has been done for the formalization of narratives, not even for Propp's *Morphology of the Folktale*, the oldest and best-known formal approach (cf. Bod *et al.*, 2011). We focus on this formal framework due to its prominent place in the history of formal representations of narratives.

3 Related Work

Bremond and Verrier (1982) document experiences when teaching Propp in depth. Reading their paper is very instructive and highlights many difficulties and inconsistencies of the system and corpus as presented by Propp, in particular with respect to atypical and complex tales. Bremond and Verrier's (1982) discussion was carried out in a seminar on Propp over a whole semester, so the training of the discussants was even more intensive than the training reported here, even for the experiments **Propp III** and **IV** discussed in Sections 5.4 and 5.5.

Lendvai *et al.* (2010) discuss how linguistic information may be helpful in approaching the annotation in Propp's framework with machine learning.

Finlayson (2011) discusses learning Propp functions from an annotated corpus. This study is being continued (Finlayson, personal communication) by extending the corpus.

The original inspiration for the work reported in this article comes from the studies on inter-annotator agreement in computational linguistics, as documented by Bod *et al.* (2011) and discussed above.

Lendvai *et al.*'s approach is comparable to ours insofar as it discusses the kind of linguistic information that may be used to determine the annotation. In the long run, one could compare the usefulness of certain data in machine learning and in working with test subjects. However, the article mainly describes preparations for a study whose results are not yet available.

The direction of Finlayson's (2011) approach is opposite to ours: a given representation by Propp is fit to a story, and if annotators agree it does not fit, it is adjusted. This is an attempt to test the plausibility of Propp's system, yet it is not designed to test the annotations annotators come up with when not primed by Propp's annotation; they might be completely different. In contrast to this approach, we are focussing on intersubjective agreement when the goal annotation is not known. Any procedure of validating Propp's annotation is complicated by the fact that Propp changed his strings from the first to the second edition (see also Section 5). For our approach, the fact that Propp strings differ between editions is not problematic since we are mainly interested in the agreement between test subjects rather than agreement with Propp.

4 Propp's Formal System

Working with a corpus of 100 Russian folktales from the collection of *Narodnye Russkie Skazki* by Alexander Afanas'ev (a selection is presented in English in Afanas'ev, 1973), Vladimir Propp (1928, 1968)[3] developed a formal system that could describe the structure of each folktale by short annotation strings consisting of symbols representing what Propp calls *functions*.

The functions are tied to specific roles that characters may play within the tales, referred to as the *dramatis personae*. Not every *dramatis persona* occurs in each story, not every character represents a *dramatis persona*, some *dramatis personae* can be represented by the same character, and a *dramatis persona* can be represented by a group of characters acting together.

Propp identified seven[4] representing roles the characters may play within the tales. They are the hero (**H**), the villain (**V**), the princess (**P**), the dispatcher (**Di**), the donor (**Do**), the (magical) helper (**MH**), and the false hero (**FH**) (Propp, 1968, Chapter 3).

The actions of the *dramatis personae* are described by a set of thirty-one functions described by Propp (1968, Chapter 3) by means of numbered examples, to which we refer as *subfunctions*. For example, the function *Wedding* (**W**) is summarized as follows: 'The Hero is married and ascends the throne.' and has, among others, the following subfunctions: '1. A bride and kingdom are awarded at once, or the hero receives half the kingdom at first, and the whole kingdom upon the death of the parents[...].' and '6. The hero sometimes receives a monetary reward or some other form of compensation in place of the princess's hand [...].' (Propp, 1968, 63f; the summary is typeset in upper case in the original).

Functions are marked by symbols in the order of their occurrence in the folktale; the first seven functions, marked with lowercase Greek letters, are called the *preliminary functions:* β Absentation; γ Interdiction; δ Violation, ϵ Reconnaissance, ζ Delivery, η Trickery, θ Complicity. The *preliminary functions* are not fully developed by Propp (1968) and are not included in Propp's own annotation strings. The annotated functions are as follows: **A** Villainy, **a** Lack, **B** Mediation, **C** Beginning counteraction, ↑ Departure, **D** First function of the Donor, **E** Hero's reaction, **F** Provision or receipt of magical agent, **G** Spatial transference between two kingdoms, **H** Struggle, **J** Branding, **I** Victory, **K** Liquidation, ↓ Return, **Pr** Pursuit, **Rs** Rescue, **o** Unrecognized Arrival, **L** Unfounded Claims, **M** Difficult Task, **N** Solution, **Q** Recognition, **Ex** Exposure, **T** Transfiguration, **U** Punishment, **W*** Wedding. Most of the Proppian functions represent events, or in Propp's nomenclature *acts*.[5] The only one that does not necessarily (but still often) represents an event in the story is the function *Lack*.

Functions occur in strict sequential order, i.e. they have to occur in the folktale in the order they are given in the list above. In his own annotations, Propp often uses double subfunctions, e.g. $\mathbf{B}_2^4$ in the string for the story *Nikita the Tanner*, 148, used below. This means: the function *Mediation*, with

the subfunctions '2. *The hero is dispatched directly*' and '4. *Misfortune is announced*' (Propp, 1968, p. 37, original italics).

The status of the subfunctions is not completely clear: Propp states that in general, they are non-exhaustive even for his corpus,[6] but occasionally mentions that he gives all examples of a function that he found (cf. Propp, 1968, p. 50f on function $\mathbf{G}$). Subfunctions mainly serve to illustrate abstraction: Propp's functions are formal symbols which have an intensional label that is often too restrictive for the task; e.g. a *Wedding* can comprise just monetary rewards ($\mathbf{w}^0$) (but it would be in clear violation of Propp's intention to consider the death of the donor a wedding), in which case the princess's role need not be filled, and the donor's refusing to transfer a magical helper ($\mathbf{F}_{neg}$) can fulfill the same function (First Function of the Donor) as transferring it. In our experiments, we only gave test subjects the subfunctions which Propp gives in the arabic-numbered lists of chapter III.[7] However, subfunctions are also supposedly relevant for dependencies between functions, so that e.g. the subtype of $\mathbf{H}$ 'Struggle' and $\mathbf{I}$ 'Victory' generally match (Propp, 1968, p. 109).

Propp explains mutual dependencies of functions (Propp, 1968, p. 108ff) with respect to subfunctions, but also explains that any function can be omitted,[8] and licenses certain inversions (most notably, the reception of the helper as $\mathbf{DEF}$ before the villainy $\mathbf{A}$ or lack $\mathbf{a}$ (Propp, 1968, p. 107).

A further source of complexity of Propp's system is the fact that formal frameworks of narrative normally do not apply to the text level of a narrative, but to a more semantic level referred to as *story* (Todorov). Generally, the discourse (Todorov) of the narrative will not represent all the events of the story, so that annotation cannot be assumed to apply directly to the text. Hence, Propp sometimes notes that a function can be inferred by the reader[9], and it is not marked in the table of annotations that there is an inference. Sometimes he assumes that a function is not present and hence not shown in the function string, yet there are also intermediate cases of assimilation (Propp 1968, Chapter 4) and 'rudiments' (e.g. Propp, 1968, pp. 97, 133). Regarding inference, Finlayson (2011)

notes regarding what he calls 'implicit functions': '[Propp] gives no way to identify when they happen, and marks them inconsistently.' (p. 56; later, p. 73, he proceeds by merging pairs of functions in which one often appears as implicit $\mathbf{E} - \mathbf{F}$ and $\mathbf{H} - \mathbf{I}$ into one function, which is justified by their tendency to co-occur).

In the full Proppian system, there are a few specific ways to break strict sequentiality (Propp, 1968, Section IX.D): The most important one is that some folktales contain a series of individual tale units, called *moves*, which contain their own sequential list of functions. Examples are *trebling*, the triple repetition of moves within the tale, and moves in which a magical agent is obtained in the first move but only used in the second move of the tale.

Propp summarizes his system in the following formula, which abstracts from interdependencies, repetitions of (groups of) functions, and does not show that Propp allows omitting any function and relaxes order in some cases, as mentioned above.[10]

$$\mathbf{A\,B\,C}\uparrow\mathbf{D\,E\,F\,G}\begin{Bmatrix} \mathbf{H} & \mathbf{I} \\ \mathbf{J} \\ \mathbf{M} & \mathbf{N} \end{Bmatrix}\mathbf{K}\downarrow\mathbf{Pr\text{-}Rs\,o\,L\,Q\,Ex\,T\,U\,W}^*$$

(Propp, 1969, p. 95; 1972a, p. 104; 1972b, p. 130)

This expression shows that Propp distinguishes two kinds of tales: those involving struggle and victory ($\mathbf{H}, \mathbf{I}$) and those involving a difficult task with its resolution ($\mathbf{M}, \mathbf{N}$). According to Bremond and Verrier (1982), the whole framework is designed to fit the first kind of stories.[11] Whether or not we share the critical note of their text, for experimental purposes, special attention has to be paid to comparison of stories of the two different types; in this study we choose $\mathbf{H}, \mathbf{I}$ rather than $\mathbf{M}, \mathbf{N}$ for an experiment.

Designing an annotation measure for Propp's system is not trivial, as (Finlayson, 2011, 57f) observes: 'Propp's monograph was not intended as an annotation at all, but rather a theory of narrative'.

There are two difficulties when choosing and justifying a concrete measure of inter-annotator agreement. First, as mentioned above in this section, Propp's system operates on a 'hidden layer' (*story*) rather than on the textual layer, as explained above,

where furthermore text spans corresponding to functions can be very long or very short. Also, large spans of the text (and hence: a good part of the events) may go unlabelled in Propp's system, and will then be omitted from the structure, as noted and criticized by Bremond and Verrier (1982). This fact was also observed as a quantitative outcome of the experiments reported here as **Propp II** and **III** (Sections 5.3 and 5.4: test subjects annotated between 4 and 74% of the sentences of a story, but except for one annotation per story, all stayed below 40%, and the means per story were between 10% for the longest story (*Seven Semyons*; labels are explained in the next section) and 33% for the shortest (*Nikita*).

Secondly, Propp's system contains dependencies between the functions and hierarchical elements (moves)[12], and therefore any statistical analysis of inter-annotator agreement would have to reflect these dependencies. Since standard statistical accounts of inter-annotator agreement do not respect this property, it is not trivial to choose a formula (cf. Artstein and Poesio, 2008). We therefore decided not to calculate a precise number for inter-annotator agreement, but use it as an intuitive concept, and thus avoid the need to justify the choice of a specific measure.

5 Empirical Work

The following empirical studies were conducted as group experiments: A group of test subjects was trained in Propp's system, and then asked to apply the training to tales from Afanas'ev's corpus.

We first report on the two experiments that were conducted first (**Propp I** and **II**, covered in Bod *et al.*, 2012), and then turn to the later experiments that incorporated lessons learned from the first experiments (**Propp III** and **IV**).

5.1 Method and Material
5.1.1 *Procedure and Data*
In all experiments, test subjects were trained in the Proppian framework and then asked to annotate two to four of the folktales formalized in Propp (1968).

Major changes between the development of the experiments were the following: (1) the time devoted to the training of test subjects was increased from 45 minutes in **Propp I** and **II** to 5 hours in **Propp III** and **IV**, the presentation of the Proppian system and the way of discussing example stories was extended and deepened, and (2) the kind of data that we gathered from the test subjects were slightly different. The first type of changes also reflects how the present authors' understanding of Propp's system changed during the process. Regarding the second change: the primary focus was always on gathering the string resulting from Proppian annotation, but in **Propp I** (*group 1*), **III**, and **IV**, we also asked for the subfunctions; in **Propp II** and **III**, emphasis in the training was put on annotating spans of text; **Propp II** and **IV** asked for the assignment of the *dramatis personae*. Table 1 gives an overview of some changes in the experiments. Changes will also be explained in detail in the respective sections.

5.1.2 *Material*
We were naturally confined to material from Propp's corpus, which is a subset of Afanas'ev's collections of Russian fairytales (*Narodnye Russkie Skazki*). Criteria for the selection of stories were the following.

First, the story had to be available in an English translation; later, when we did the experiments in Germany, we also used German translations.[13] Second, Propp's annotation of the story should have no repetition and not more than one move,[14] and retain strict sequential ordering; it should also use relatively few functions (we started with stories of six to eight functions, see Table 2). Third (this was decided after the first experiment discussed below, Section 5.3), the story had to be short, in order to keep the amount of annotation time within reasonable bounds.

We used the folktales *The Seven Semyons, 147, Shabarsha, 151*, and *Ivan the Bear's Son, 152, Nikita the Tanner, 148*, and *Ivan Popyalov, 135*; in the following, we refer to these folktales as *Semyons*[15], *Shabarsha*, and *Ivanko*, *Nikita*, and *Popyalov*. In **Propp I**, we also used the folktale *The Enchanted Princess*, but because of its length, it took considerably more time to annotate than the other stories

Table 1 Parameters of Propp training side by side

	I	II	III	IV
Setting: Parameters				
Time for training	$^3/_4$ h	$^3/_4$ h	prep. + [5] h	prep. + [5] h
Language of training/questionnaires	en	en	en	de
Language of stories	en	en	en/de	de
Payment	yes	yes	no	no
Test subjects	5 + 4	6	9	18
Task: Aspects annotated				
Dramatis personae	y	(g)	n	y
Functions	y	y	y	y
Subfunctions	y/n	n	y	y
Indications of text spans	n	y	y	(y)

y, part of the task; (y), officially part of the task, but largely ignored by test subjects; n, not part of the task; (g) given before the task; en, English; de, German; prep., preparatory course on (formal) narratology.

Table 2 Overview of the stories used in experiments and training

Story	I	II	III	IV	Functions
(Invented fairy tale)	t	t	–	–	(N/A)
The Enchanted Princess	E	–	–	–	(N/A)
The Seven Semyons	E	E	t	t	(N/A)
Shabarsha (151)	E	E	E	E	8
Ivanko (152)	E	E	–	–	6
Ivan Popyalov (135)	–	–	t	t	10
Nikita the Tanner (148)	–	–	E	E	8

E, used in experiment; t, used in training; –, not used; (N/A), not applicable.

and was hence not reused; we ignore it in the following discussion. An overview of the use of the stories is in Table 2.

The English translations we worked with were Gutermann's (Afanas'ev 1973) for *Semyons*, *Ivanko*, and *The Enchanted Princess*, and Cook's (Afanas'ev 1985b) for *Shabarsha*; *Popyalov* was translated by William Ralston Shedden Ralston (no year).[16] From **Propp III** on, we also used German translations. These were printed in parallel (paragraphs aligned) in **Propp III**, and only German versions were used in **Propp IV**. For *Ivanko*, *Semyons*, and *Shabarsha*, we prepared translations of the stories from the English versions, originally for use in

another experiment reported by Fisseni and Löwe (2014). For *Popyalov* and *Nikita the Tanner*, we used translations from Afanas'ev (2008).

5.1.3 *Propp's Own Annotation*

We give Propp's annotation here, for reference and because some comments on the strings are in order.

One important issue when working with Propp's annotation is that strings differ between the first and second edition of Propp's work; we used the German edition (Propp 1972a) as a reference for the second edition. In cases of doubt, we also consulted the Russian first edition and the second Russian edition. Strings are quoted using the symbols of the English edition.

As mentioned, Propp's annotations only contain the function strings and do not include the *preliminary functions α–θ*. Therefore, we do not take the preliminary functions into account for comparison between Propp's original strings and the strings produced by the test subjects.

Shabarsha: $\mathbf{a^5 B^4 C \uparrow H^2 I^2 K^1 \downarrow}$

Ivanko: $\mathbf{A^9 \uparrow H^2 I^2 K^1 \downarrow}$

Semyons:—[17]

Nikita: $\mathbf{A^1 B_2^4 C \uparrow H^2 I^1 K_4 \downarrow}$

The first edition has a $\mathbf{W_-^*}$ (*Wedding* without accession to the throne, but negated, which one could interpret to reflect Nikita's refusing rewards); this was dropped in the second edition.

Ivan Popyalov: $\mathbf{A^4 F^2 C \uparrow H^1 I^1 K^4 \downarrow Pr^5 Rs^5}$

The inversion ($\mathbf{F^2}$ before $\mathbf{C}$) was not present in the first edition of Propp's work; subfunction for receiving the magical helper was different: $\mathbf{F^3}$ instead of $\mathbf{F^2}$. Surprisingly $\uparrow$ was missing in the first edition: $\mathbf{A^4 C F^3 H^1 I^1 K^4 \downarrow} \begin{Bmatrix} \mathbf{Pr^4 Rs^7} \\ \mathbf{Pr^1 Rs^5} \end{Bmatrix}$

Invented Fairy Tale:

The string for the example story we wrote for pedagogical use was: $\boldsymbol{\beta\gamma\delta\epsilon}\mathbf{ABC\uparrow DEFGHIK\downarrow W}$

5.1.4 Remarks on Story Similarity

Ivanko and *Shabarsha* are superficially similar as they revolve around tricking devils inhabiting a lake (even with similar tricks), but have quite different Proppian annotations.

Nikita and *Shabarsha* are almost identical with respect to Propp's annotation (not considering subfunctions), but they are superficially quite different: Nikita is a classical hero who slays a dragon to save a princess and the whole empire, while farm hand Shabarsha tricks devils into giving him a lot of money to economically support his master.

Popyalov is superficially closer to Nikita (since it also involves dragon slaying), but has a much more complicated course of events.

5.1.5 The Notion of Stability

As it is not possible to do a serious statistical analysis on the basis of few annotations, we instead do a mostly qualitative analysis based on very simple quantitative data. We say that a function occurs *stably* if it is in more than half of the annotations. We further distinguish *strong stability* when the marked text of the annotators overlaps for more than half of the test subjects, and *weak* otherwise.

Doing such a qualitative analysis is most important for **Propp I** and **II**, where no two strings produced by the test subjects are the same, and comparison would have to be per function; calculations of statistics per function, however, are not useful there because of the variation in the assignment of *dramatis personae* and the small amount of data. Using the notion of stability, we can construct a *majority string* consisting of those functions that were listed by the majority of test subjects.

For the later experiments, showing stability is still helpful, and the strings are so similar that calculating agreement is not really interesting. It is also interesting to see to what extent different spans triggered the same functions.

5.2 Propp I

Test subjects were trained in the Proppian framework and then asked to annotate four of the folktales formalized in Propp (1968). An annotation in **Propp I** consisted of (1) the assignment of story characters to the *dramatis personae*, and (2) a list of either only the *functions* (group 1) or also the *subfunctions* (group 2) occurring in the folktale.

5.2.1 Procedure

We had nine test subjects, all students of the *Universiteit van Amsterdam*, and all with native or near-native competence of English. We split them into two groups: Test subjects 1–5 were in group 1 (no subfunction marking), and test subjects 6–9 were in group 2 (subfunction marking). Test subjects were instructed that the experiment would last 3 hours and received a moderate financial compensation for participation.

The experiment started with a 45 minutes introduction to Propp's system given by a native speaker of English supported by a projector presentation explaining the relevant fragment of Propp's system. Only a selection of the subfunctions was included (labelled 'examples' for group 1 and 'subfunctions' for group 2). We analyzed a simple example story, of our own design, as an illustration of how the functions and *dramatis personae* interact; discussion and questions were encouraged. Brief definitions of the *dramatis personae* and a condensed version of the function definitions were distributed as a leaflet for use during the annotation. The leaflet also included a brief historical background on Propp and his system, including: a direct reference to the sequential nature of the functions, that the same character can serve as more than one dramatis persona and that the entire story must be read to completion before the task of assigning *dramatis personae* should commence.

5.2.2 Results

We give the results of the assignments of *dramatis personae* in Table 3. The results indicate that the test subjects did not fully understand the Proppian scheme. Note in particular the variation in the three main *dramatis personae*, Hero, Villain, and Princess. Especially having Shabarsha as both Hero and Villain (and potentially False Hero) at the same time is in violation with one of the principles of Propp's framework, as is the coupling of

Table 3 The assignment of the *dramatis personae* for the three folktales in **Propp I**

Test subject	H	V	P	PF	Di	Do	MH	FH
Ivanko								
1	Ivanko	Devils			Peasant			
2	Bearlet	Bear/devil	Bearlet/Wife	Peasant	Peasant			
3	Ivanko	Thieves/dogs/ devil	Wife	Peasant	Peasant	Grandfather	Horse	
4	Ivanko/Mother	Devil/peasant			Peasant			
5	Ivanko	Father	Father's satisfaction	Grandfather	Father	Little devil	Horse	
6	Ivanko	Devil			Peasant			
7	Ivanko	Devil	Bear	Ivanko/wife		Peasant	Devil	
8	Ivanko	Devil			Peasant		Horse	
9	Bearlet	Father	Father, money		Father	Devil	Hare	

Test subject	H	V	P	PF	Di	Do	MH	FH
Semyons								
1	Semyons		Elena		Tsar			
2	Semyons	Tsar	Elena		7			
3	Semyons	7th Semyon	Elena	Tsar	Tsar		Kitten/stone	7th Semyon
4	7th Semyon	Elena's father	Elena	Elena's father	Tsar		Semyon bros	
5	Semyons	Tsar	Elena	Elena's father	Tsar	Tsar	Cat	
6	7th Semyon	Tsar	Elena	Elena's father	Tsar		Semyons	
7	Semyons		Elena	Tsar	Tsar		Ship	
8	7th Semyon		Elena		Tsar		6 Semyons	
9	7th Semyon	Tsar	Elena		Tsar		Semyons	Tsar

Test subject	H	V	P	PF	Di	Do	MH	FH
Shabarsha								
1	Shabarsha		Gold	Little Devil/ Grandad	Master			
2	Shabarsha/little devil	Shabarsha			Grandad			
3	Shabarsha	Little devil	Gold	Grandad	Grandad	Master	Cap	
4	Shabarsha	Little devil	Gold	Grandad	Master	Master		
5	Shabarsha	Shabarsha	Gold	Grandad	Master		Bear/hare	Shabarsha
6	Shabarsha	Little devil	Gold	Grandad	Master		Bear/hare	
7	Shabarsha	Little devil/ grandad	Gold		Master	Master	Twine	
8	Shabarsha	Little boy	Gold	Grandad	Master		Bear/hare	
9	Little Devil	Shabarsha	Peace	Grandad	Grandad			

H, hero; V, villain; P, princess; PF, princess's father; Di, dispatcher; Do, donor; FH, false hero.

Shabarsha with his adversary, the Little Devil. Also the lack of correspondence between the Donor and the Magical Helper shows that the intention of Propp's system was not clear at all.

Correspondingly, the annotation strings vary widely and are given in Table 4 (subfunctions are marked by superscripts, with a missing subfunction marked by $\varnothing$). The strings are also longer than Propp's strings (compare an average of 14.2, 13.2, and 12.8 functions with the Propp's of 6, 8, and 8 for *Ivanko*, *Semyons*, and *Shabarsha*, respectively), because test subjects could not filter out the functions relevant for the structure of the tale (in Propp's sense).

Table 4 The annotation strings for the three folktales in **Propp I** (cf. note 15)

Subject	Proppian functions for **Ivanko**	
Propp	A^9 ↑	H^2 I^2K^1↓
Stable β γ δ $\zeta\eta$ θ	a B C↑D	G H I K ↓
1 $\quad\beta$	↑	G H I ↓
2 $\quad\beta$ γ δ $\zeta\eta$ θ	A a B C↑	G H I K ↓ $\quad$ U
3 $\quad\beta$ γ δ $\zeta\eta$ θ	A a B C↑D E	H I K ↓ N Ex
4 $\quad\beta$ γ $\delta\varepsilon$ $\zeta\eta$ θ	a B C↑D	G H I K ↓ZNQ U
5 $\quad\beta$ γ δ $\zeta\eta$ θ	a B C↑D E F	JI K
6 $\quad\beta^{\varnothing}\gamma^2\delta\varepsilon^2$ $\zeta\eta^1\theta^1$	B^5 C↑	G^3 H^2 I^2K^1↓
7 $\quad\beta^1$ $\quad\delta\varepsilon^{\varnothing}$	a^5B^2 C↑D^1E^1	$H^{\varnothing}$ I^2K^2↓ $\quad$ U
8 $\quad\beta^1$ γ^2	↑	G^3 H^2 I^2 ↓ $\quad W^6$
9 $\quad\beta^1$ $\quad\zeta$ θ^1	$a^5B^{\varnothing}$ ↑$D^1E^9F^9G^{\varnothing}$	K^1↓

Subject	Proppian functions for **Shabarsha**	
Propp	a^5B^4C↑	$H^2I^2K^1$↓
Stable γ η	a B C↑	G H I K
1	a B C↑	G H K
2 $\quad\beta$ $\varepsilon\zeta\eta$ θ	Aa B C↑	G H I K ↓ $\quad$ W
3 $\quad\beta\gamma$ $\varepsilon\zeta\eta$ θ	A B C D EF	G H I K
4 $\quad\beta\gamma$ $\varepsilon\zeta$	a B C↑ E	I K Pr QU
5 $\quad\beta\gamma$ η	a B C D F	K ↓ QU
6 $\quad\gamma^2\delta$ $\eta^1\theta^1$	a^2	$H^2I^2K^1$↓ N UW^6
7 $\quad\gamma^2$	a^2B^2C↑D^1	$F^1G^3H^2I^2K^1$ o W^6
8	a^2B^1C↑	$G^3H^2I^2K^1$
9 $\quad\eta^{\varnothing}\theta^1$	a^2B^2	H^2 $\quad W^{\varnothing}$

Subject	Proppian functions for **Semyons**	
Stable γ δ η	a B C↑	G K ↓Pr W
1 $\quad\beta$	a B C↑	G K ↓
2 $\quad\beta$ γ $\delta\varepsilon\zeta\eta\theta$	Aa B C↑	G H K ↓PrRs W
3 $\quad\beta$ γ $\delta\varepsilon\zeta\eta\theta$	Aa	DEF G K ↓Pr NQTU
4 $\quad\beta$ γ δ	a B C↑D	G HIK ↓PrRs Q
5 $\quad\beta$ $\quad\zeta\eta\theta$	a B C↑	F G K ↓
6 $\quad\beta^{\varnothing}\gamma^1$ δ ζ	a^1B^2 ↑	$G^{\varnothing}$ K^1 Pr N W^6
7	a^5B^1C↑	F^3G^3 K^1↓ W^6
8 $\quad\beta^2$ γ^1	a^1B^1C↑	G^3 K^2↓ W^6
9 $\quad\beta^2$ $\gamma^{\varnothing}\delta$ ζ	a^5B^2 ↑	$G^{\varnothing}$ K^2↓Pr N W^6

5.2.3 *Methodological Conclusion*

Four out of nine test subjects reported that the generated example story was considerably simpler than the folktales. The generated story was quite short, but every sentence of the text (very straightforwardly) corresponded to (at least) one of Propp's functions.

The variation in the assignment of characters to *dramatis personae* suggests that the description of the *dramatis personae* was not precise enough. For

instance, our description of the hero used the relative clause 'who is good'. Arguably, Shabarsha's behaviour in *Shabarsha* cannot be described as 'good', which caused some of the variation in the assignment of the hero.[18]

A number of functions are consistently annotated which do not show up in Propp's own annotations. On the other hand, we see that some of Propp's functions show up in all or almost all annotations strings: e.g. ↑, **H**, **I**, **K**, and ↓ are reliably reproduced in the *Ivanko* annotation strings. However, since we do not know which events in the tale the annotators marked with these functions, we cannot be sure whether these are actual reproductions of Propp's assignments.

5.3 Propp II

The experiment **Propp II** was a modified version of **Propp I**, taking the problems discussed in Section 5.2 into account.

The main changes to **Propp I** were: the test subjects were given the assignment of *dramatis personae*; subfunctions were not discussed at all; the example story now was from Propp's own corpus. As Propp only recorded the annotation strings, we extrapolated his choice of *dramatis personae* (see Table 5).

We used *Shabarsha*, *Ivanko*, and *Semyons* for the experiment and *Popyalov* for training. An annotation of a folktale in **Propp II** consisted of (1) a list of the functions occurring in the folktale, and (2) marked text passages for each of the functions that occurred, in order to tie function identification to events in the story.

5.3.1 Procedure

We had six test subjects, all students of the *Universiteit van Amsterdam*, and all with native or near-native competence of English. Test subjects were instructed that the experiment would last 3 hours and received a moderate financial compensation for participation.

The experiment started with a 45-minutes introduction to Propp's system as before, but was improved by giving short descriptions of the *dramatis personae* roughly based on Propp's original text and the descriptions of the functions (with subfunctions labelled as 'examples') from Propp's text. During the training, we analyzed *Popyalov*. Again the condensed version of the *dramatis personae* and functions was distributed as a leaflet to for use during the annotation; the leaflet given was the same as in **Propp I**, e.g. it contained the same subset of the subfunctions listed as 'examples', the historical background, etc. Test subjects were given the generated assignment of characters to the dramatis personae together with each folktale.

5.3.2 Results

We give the results of the function annotation in Table 6. The annotation strings are noticeably shorter than in **Propp I** (on average 6.8 functions per annotator, compared with 13.4 functions in **Propp I** and 7.3 functions in the original Propp strings), and more similar to Propp's own strings, but we still do not have matching strings among the test subjects.

In *Ivanko*, $\boldsymbol{\beta}$, **I**, and ↓ were strongly stable, and ↑ and **H** were weakly stable (of which all are annotated by Propp); in *Shabarsha*, **a** and ↑ were strongly stable and **H** and **I** were weakly stable (of which ↑, **H**, and **I** were annotated by Propp); in *Semyons*, **a**, **B**, **G**, and **W** were strongly stable and ↑ and **K** were weakly stable. Note that in both *Ivanko* and *Shabarsha*, there is a strongly stable function not annotated by Propp (**B** and **a**, respectively).

Table 5 *Dramatis personae* as presented to test subjects of **Propp II**

Story	Hero	Villain	Princess	Dispatcher
Ivanko	Ivanko	Little devil + grandfather		
Semyons	7th Semyon		Elena the Fair	Tsar
Shabarsha	Shabarsha	Little devil + grandfather		

Table 6 The annotation strings for the three folktales in **Propp II** (cf. note 15); 'Str.St.' indicates the strongly stable functions

	Ivanko		Semyons		Shabarsha	
Test subject	Proppian Functions	Test subject	Proppian Functions		Test subject	Proppian Functions
Propp	A ↑ HIK↓				Propp	aBC↑HIK↓
Stable	β ⎟ ↑ HI ↓	**Stable**	aB ↑GK W		**Stable**	a ↑HI
Str. St.	β ⎟ I ↓	**Str. St.**	aB G W		**Str. St.**	a ↑
1	β ⎟ a ↑GHIK↓	1	aB ↑GK W		1	a HIK N
2	βγ ⎟ ↑ MN W	2	β ⎟ aB ↑ K↓ W		2	aB ↑ MN W
3	β ⎟ B↑ HI U	3	β ⎟ aB G oNW		3	a C↑HI M UW
4	β ⎟ ↑ HI ↓ U	4	η ⎟ a ↑G Pr W		4	a ↑HI MN
5	β ⎟ aB↑ HI ↓	5	aB ↑ K↓PrRs W		5	a ↑H K
6	β ⎟ aB↑ HIK↓ W	6	β ⎟ aBC↑GK↓PrRs W		6	aBC HIK W

5.3.3 Discussion

Most likely, a reason for the much longer strings in **Propp I** was the assignment of superfluous *dramatis personae* by the test subjects in that experiment. By fixing the *dramatis personae* ourselves, we controlled this source of variation. Other than that, the strings are still very different and contain a lot of noise.

5.3.4 Methodological Conclusion

The differences between **Propp I** and **Propp II** show that the assignment of the characters to the *dramatis personae* has an important effect on the assignment of the functions. Even with pre-assigned *dramatis personae*, there are marked differences between Propp's and the test subjects' annotations, and among the test subjects. Some of this effect can be explained by the vagueness of the description of Propp's functions: remember (cf. Section 4) that subfunction 6 of **W** (*Wedding*) is listed as 'Other form of compensation like a monetary reward'. This vague description fits in much more general situations than Propp apparently intended. Making these vague descriptions understandable for the test subjects may require considerably more time and training than we gave the test subjects in our experiments.

The detailed study of human annotations of Propp's framework highlights weaknesses such as vague descriptions of *dramatis personae* and functions, and in general, points to some important obstacles for an automatization of the process of formalization in a computational setting.

5.4 Propp III

5.4.1 Experiment description

The experiment **Propp III** was a modification of **Propp II**, adjusting for the problems highlighted in the preceding section: it emphasized the overall system, thus suggesting a more top-down approach, and trying to reduce the vagueness of the definitions provided for the atoms of the system (functions, *dramatis personae*) and their interconnectedness.

We changed to only two folktales, reusing *Shabarsha* (151) and adding *Nikita* (148). An annotation in **Propp III** consisted of (1) a sequential list of the functions, including subfunctions, and (2) marked text passages for each of the functions that occurred.[19]

The main changes from **Propp II** were: the test subjects were not given the *dramatis personae* assignment; subfunctions were discussed and required, the complete set of subfunctions were given; there were two example stories used, both from Propp's corpus. As noted in Section 5.3, the assignment of the *dramatis personae* was not part of Propp's annotation.

5.4.2 Procedure

The experiment was carried out in a weekend seminar on formal approaches to narrative of the programme *Stipendiaten machen Programm* by the *Studienstiftung des deutschen Volkes*.

We had nine test subjects, all of them part of the *Studienstiftung des deutschen Volkes* programme, most of them undergraduate students. Test subjects were instructed that the experiment would take place in English and was comprised of two parts, 3 hours of teaching Propp's morphology followed by 2 hours of an annotation experiment (including the annotation of the two stories in two extremely simple formalizations).

During **Propp III**, we discussed Propp's system in more detail than in **Propp I** and **II** and highlighted the interdependency of function pairs on the one hand, and of functions and *dramatis personae* on the other hand. Subfunctions were discussed.

The experiment started with a 3-hour introduction of Propp's system given by a native speaker of English (the same individual who provided the instruction in **Propp I** and **II**) supported by a projector presentation. We gave short descriptions of Propp's *dramatis personae* based on Propp's original text and the character functions. Here, we moved to seven *dramatis personae*, merging the Princess and Princess's Father, as is more commonly done. Each function was discussed including every subfunction. Popular culture examples of the functions were given (e.g. Disney animated features or Joanne K. Rowling's *Harry Potter* book series); sometimes an improvised tale about a magic goat was used for examples. The test subjects were instructed to ask questions throughout, making sure that the group had understood the definition of a function before moving on. The test subjects were given the folktale (*Ivan Popyalov, 135*) from the Propp corpus, and tasked with reading it at least two times, while 'thinking about how to annotate the story', the group then worked together to annotate the story, with Propp's original string given at the end of discussion. Subsequently the test subjects were given the folktale (*The Seven Seymons*, see note 15) and participated in a mock experiment, followed by a group discussion of the participant's and Propp's annotation strings.

The test subjects were then given 90 minutes, 45 min per folktale, to annotate the two experimental folktales, within an environment similar to a written exam, i.e. silence and no notes. The guidelines leaflet was not used, as all subfunctions were listed on the annotation forms.

5.4.3 Group discussion

The initial annotation of *Ivan Popyalov* showed more disagreement between test subjects on what constitutes the presence of a function. The group engaged in a discussion amongst each other and with the instructor to achieve consensus on a string produced by the group; afterwards, they discussed where their string differed from the one provided by Propp and why.

Both example stories happened to include, in Propp's annotation, the function F_3: *The provision or receipt of a magical agent – the agent is prepared*. This particular subfunction was the source of much of the discussion, as it allows for the sequential nature of the annotations to be broken, and it is not clear from Propp's descriptions what constitutes 'prepared' versus incidental nascent ability of the hero. As seen in the results, three test subjects annotated one story as containing F_3, while most others (and Propp) did not. This lends credibility that the function/subfunction pairing is troublesome, at the least.

The test subjects tended to immediate agreement at the functional level though there was less agreement at the subfunctional level. For example, the group was in agreement that the tale contained an instance of function **K** *Liquidation:* 'The initial misfortune of lack is liquidated'. However the group was split between it being K^1 *The object of a search is seized by the use of force or cleverness*, or K^2 *The object of search is obtained by several personages at once, through a rapid interchange of their actions*, with one individual arguing for K^8. The object of search is obtained with the help of enticements. In the tale the seven Semyons trick the princess to get on board their ship with the promise of beautiful jewels and a cat. Each in turn uses their unique skills to escape with the princess on board unscathed, and without being followed by her kingdom's navy.

5.4.4 Results

We give the results of the function annotation in Tables 7 and 8. We observe that the strings are very uniform, with most only differing by one or

Table 7 Propp Strings for *Shabarsha* in **Propp III**; 'Str.St.' indicates the strongly stable functions

Propp		a^5	B^4	C	$\uparrow$		H^2	I^2	K^1			$\downarrow$
Stable		a^5	B^3	C	$\uparrow$		H^2	J^2	K^1			
Str.St.		a^5	B^3				H^2	J^2				
1		a^5	B^3	C	$\uparrow$		H^2	I^2	K^1			$\downarrow$
2		a^5	B^3		$\uparrow$	F^3			K^1			T^4
3		a^3	B^3	C		F^3			K^3			
4	β^3	a^5		C			H^3	I^2	K^1			
5		a^5	B^3	$C_\varnothing$	$\uparrow_\varnothing$		H^2	I^2	$K^1_\varnothing$	$M_\varnothing$	$N_\varnothing$	
6		a^5	B^3	C	$\uparrow$		H^2	I^2	K^1			
7		a^5	B^3	C	$\uparrow$		H^2	I^2	K^1			
8		a^5		C	$\uparrow$		$\left\{ \begin{matrix} H^2 & I^2 \\ H^2 & I^2 \end{matrix} \right\}$		K^4			
9		a^5	B^3		$\uparrow$				K^1			

TS 8: 'if repetition is allowed'.

Table 8 Propp Strings for *Nikita* in **Propp III**; 'Str.St.' indicates the strongly stable functions

Propp		A^1		B^4_2	C	$\uparrow$		H^2	I^1	K^4	$\downarrow$			$[W^*_-]$
Stable		A^1		B^1	C	$\uparrow$		H^1	I^1		$\downarrow$			
Str.St.		A^1		B^1	C	$\uparrow$		H^1	I^1		$\downarrow$			
1		A^1		B^2	C	$\uparrow$		H^1	I^1	K^1	$\downarrow$			
2		A^8		B^1	C	$\uparrow$				K^4	$\downarrow$			
3		A^1		B^2	C			H^1	I^1		$\downarrow$			
4		A^{15}	a^6	B^1		$\uparrow$	G^2	H^1	I		$\downarrow$			
5	γ^2	A^1		B^1	C	$\uparrow$	F^3	H^1	I		$\downarrow$	$M_\natural$	$N_\varnothing$	
6		A^1		B^1	C	$\uparrow$		H^1	I^1		$\downarrow$			
7		A^{15}		B^1	C	$\uparrow$		H^1	I^1	$K^1_\varnothing$			U	
8		A^1		B^2	C	$\uparrow$		H^1	I^1		$\downarrow$			
9		A^1		B^1	C	$\uparrow$		H^1	I^1	K^{10}	$\downarrow$			

Subscript $\varnothing$ indicates that no text span was annotated. Subscript $\natural$ indicates that the string was written in order, but the annotation is not.

two functions. There are few functions not anno-tated by Propp and there is a high correlation between the experimenter and Propp's strings. In *Shabarsha*, every function in Propp's string except for $\downarrow$ was stable. These results contrast the conclu-sions we drew from **Propp I** and **II**.

The text of the story does not include an explicit return, though it does end with Shabarsha leading a

rich life. The only test subject, out of nine, to include this function in their annotation string (1) commented: 'most likely, but it doesn't say it in the text'. This is clearly the case of an implicit function: Nowhere in the tale is it said that Shabarsha returned to his master (text level); however, it can be inferred (and included in the story) that he returned home because staying at the lake seems implausible, and it is reported that he lived in wealth.

In *Nikita*, we see a similar comment (9) 'Most likely [K^{10}, i.e. *A captive is freed*], but it does not say directly that the tsar's daughter was freed'.

The choice of what text span supports the functions varies substantially between annotators. This was especially apparent with the implicit functions. In addition, some test subjects chose to annotate a discrete act/event that could indicate the presence of the function, while some others annotated the entire scene in which the function took place, and others still somewhere in between; examples are **H** in *Shabarsha* and **I**, and to a lesser extent **H** and **I**, in *Nikita*. For some functions, delineation of the events seems unclear (**B**-↑ and **I**-↓ in *Shabarsha* and **K**/↓ in *Nikita*), so that spans are annotated as overlapping.

We note that the annotated spans of text still wildly differ on length, and overlap. Though we have functional agreement by the annotators, what text corresponds to the presence of a function is still non-uniform.

While the test subjects agree on the presence of the functions, almost uniformly, the annotated spans were not as uniform. For example, the function **H** 'Struggle' is mostly annotated as one of the following two sentences, which are very close, and semantically overlapping.

> (1) The dragon, seeing that he could not avoid trouble, went out to fight in the open field. Nikita fought him for a long time or a short time; in any event, he defeated him.

This again reflects that the Proppian functions exist on the *story* level, while the annotations (must) occur at the *text* level.

The training in **Propp III** did not cover the interdependencies, and mutual exclusions between functions. Test subject (4) submitted a string that included both **A** *Villainy* and **a** *Lack*. Propp's framework has both of these functions as discrete 'initiating incidents', i.e. the corpus can be divided between villainy tales and lack tales. Similarly test subject (5) included an inversion function as well as an impossible string $M − N$, where the annotation spans were in inverted order, while the string was written in sequential order.

5.4.5 Methodological conclusion

The results of **Propp III** are in contrast to the results from **Propp I** and **II**. By increasing the training time substantially, and focusing on the known problematic areas of Propp's framework, we can get much closer to the original Proppian string, and more importantly the test subjects produce very similar, if not the same, strings. However, this focused intensive training indicates that the Proppian framework is not natural, but is (at least) understandable.

We also note that **Propp III** was part of a weekend seminar on formal approaches to narrative for which the participants had to sign up. Therefore, participants were not a representative sample of the student population, but a selection of students who make a considerable time commitment to learn about this particular topic. In addition, the test subjects had a general introduction to narratological questions with an added potential effect on their understanding of the framework.

5.5 Propp IV
5.5.1 Experiment description

We conducted another experiment in a setting similar to **Propp III**, at a *Sommerakademie* of the *Studienstiftung des deutschen Volkes*, where the first and third authors were teaching a course on formal and computational approaches to narrative. Before the Propp training, test subjects had spent 1 week (5 days, total of at least 15 work hours) getting to know basic narratology and some formal approaches (e.g. Plot Units by Lehnert, 1981, Story Grammars from Rumelhart, 1975, Move Grammars of Pavel, 1982, and the system by Labov, 1997).

An annotation in **Propp IV** consisted of (1) a Propp string and (2) a table of *dramatis personae*. We did not ask for the *Princess's Father*.

5.5.2 Procedure

The training was similar to **Propp III**, but was administered to a group of 18 native German speakers in German by two native German speakers; the function list used the short descriptions from Propp (1972a), accordingly. The first part lasted one morning session (about 3 hours) with in-depth discussion of the Proppian system and an example story analysis; the second part consisted of an experiment-like 1-hour training session. Both parts made use of the experiences of the previous experiments. After this, a discussion about another part of the course and a coffee break took up about 1.5 h. Then, 2 hours were used for the experiment.

In the training, annotation of concrete text spans was less important than the discussion of functionhood, presence of functions, exhaustivity of the subfunction inventory. Furthermore, the training taught a clear bottom up approach: starting from key functions like **A/a** *Villainy/Lack* and **K** *Liquidation* or **W** *Wedding* move to the identification of hero (and potentially villain) and other *dramatis personae*, then continue to find the other functions. During the training, Propp's text Propp (1972a) was available to the test subjects; the introductory chapter was discussed during training.

In this experiment, stories were only given in German, but the versions used were the same as above: *Ivan Popyalov* for a first training, *The Seven Semyons* in an experiment-like situation and finally both *Nikita the Tanner* and *Shabarsha* during the experiment on the next day.

5.5.3 Results

As in **Propp III**, there is considerable agreement among test subjects, and the agreement with Propp is remarkable: ignoring the subfunctions, the stable strings (i.e. the string of those functions annotated by the majority of test subject) matches almost exactly with Propp's. Results are given in Tables 9 and 10, where the stable strings are juxtaposed to Propp's strings. Most test subjects did not annotate text spans in a useful way, so that we do not report corresponding results.

Regarding the theme of *story – discourse* from the discussion of **Propp III**: the function ↓ is missing again for the majority of test subjects, which to us

confirms that its 'implicit' presence (as described in the previous experiment) makes it difficult to detect for formalizers. Furthermore, two test subjects noted that functions were implicit, namely **K** (Liquidation) is implicit: this is not implausible, as Shabarsha initially wants to get money for his master, but during the story Shabarsha does not return to his master, as just pointed out.

The lists of *dramatis personae* (see Tables 11 and 12) is now quite uniform, much more than for **Propp I** (see Table 3), similar to **Propp III**. The main question seems to be which of the devils is the villain in *Shabarsha*, which is quite irrelevant for the function assignment. During the training, the fact that multiple characters could fill the same dramatis persona slot was not explicitly addressed, though it was the case in the example story *Semyons*.

Overall, the minor roles of dispatcher, donor, and magical helper seem to be the most difficult; but even there, there is little noise and much agreement. Some decisions are still surprising in several respects and may highlight difficulties of the system: First, there are some assignments that are not in line with Propp's intentions, at least for *Shabarsha*: the devil cannot be the donor, even though he gives gold to Shabarsha, because the donor generally occurs in a phase that prepares the main story, and provides means to fulfil his quest/task; despite Propp's statement that almost any function can be omitted, Propp's own annotations never end with a **D**. Secondly, test subject 2 has an alternative set of *dramatis personae* (given as 2* in Table 12), where the devil and Shabarsha exchange roles; interestingly, no alternative list of functions is provided, even though this would be necessary with the different assignment of *dramatis personae*. One test subject notes that Shabarsha could be a false hero because of his trick; in a way this may be an artefact of the vague formulation that the hero presents 'unfounded claims' (Propp, 1968, p. 60). Test subject 4 notes that Shabarsha could be a false hero (and the devil a hero that loses) because the tale ends with Shabarsha's beer and mead being bad. However, both analyses follow our colloquial understanding of *hero*, but seem to contradict the whole system, which only permits a false hero if a hero is

Table 9 Propp strings for *Shabarsha* in **Propp IV**

			F	G/H	I	K	
Propp		a^5 B^4 $C\uparrow$		H^2	I^2	$K^1 \downarrow$	
Stable		a^5 B^3 $C\uparrow$		H^2	I^2	K^1	
1		a^5 B^3 $C\uparrow D^{2,7}$ E^1		H^2	I^2		T^4
2		a^6 B^3 $\uparrow$		H^2	I^2		W
3		a^5 B^4 $\uparrow$	F^3	$G^2 H^2$	I^2	$K^1 \downarrow M$	W^6
4		a^5 B^4 $C\uparrow$		$G^2 H^2$	I^2	$K^4 \downarrow M$	W^6
5		a^5 B^3 $C\uparrow$		H^2	I^2	$K^1 \downarrow$	
6		a^5 B^3		H^2	I^2	$K^1 \downarrow$	T^4
7		a^5 B^3 $C\uparrow$		H^2	I^2	$K^{1,4}$	
8		a^5 B^3 $C\uparrow$		$G^2 H^2$⋮	I^2	K^1	
9		A^{19} $\quad D^9$ $\quad E^9$				K^1	
10		a^5 B^3 $\uparrow$		H^2⋮	I^2	K^1	
11		$a^5 B^{3,4} C\uparrow$	F^1	$\begin{Bmatrix} H^1 & I^2 \\ H^2 & I^2 \\ H^2 & I^2 \\ H^2 & I^1 \end{Bmatrix}$		$K^1 \downarrow$	W^6
12		a^5 B^3 $C\uparrow$		H^2	I^2	$K^{1,4}\downarrow$	
13		a^5 $C\uparrow$		$G^2 H^2$	I^2	K^4	
14	$\beta^3\gamma^2\delta^1\varepsilon^1\zeta^1$ a^5 B^2 C $\begin{Bmatrix} D^9 & E^9 & F^9 \\ D^9 & E^9 & F^9 \end{Bmatrix}$			H^1	I^1	$K^1 \downarrow$	
15		a^5 B^1 $C\uparrow$		H^2	I^2		
16		a^5 B^3 C	F	$\left(H^2 I^2\right)$⋮		K^1	
17		a^5 B^3 $C\uparrow$ $\begin{Bmatrix} D^1 & E^1 \\ D^1 & E^1 \\ D^1 & E^1 \\ D^1 & E^1 \end{Bmatrix}$			I^2	K^1	
18		a^5 B^3 $\uparrow$		G H^2	I^2	$K \downarrow$	

also present, and where false heroes are always unmasked.

5.5.4 *Methodological problem: dramatis personae*

In the experiments which required the test subjects to provide *dramatis personae*, i.e. **Propp I** and **IV**, the nature of what can, and cannot, be a dramatis persona was not well understood, e.g. the donor is not just the person who transfers a good, but said transfer has to be in the actual functions that correspond to the sphere of action of the donor. This confusion was most apparent with the role of the *Princess*. The princess is often the 'object of search' for the hero's quest, and the functions in her sphere of action reflect this. However Propp's framework only allows characters, that is, things that fill the *actor role*, to be *dramatis personae*. Thus quests

Table 10 Propp strings for *Nikita* in **Propp IV**

Propp		A	a	B	C	H/G/F	I	K	↓	
Propp		A^1		B^4_2	$C\uparrow$	H^2	I^1	K_4	$\downarrow$	$[W^*_-]$
Stable		A^1		B^1	$C\uparrow$	H^1	I^1	$K_{1\vee10}\downarrow$		
1		A^1		B^1	$C\uparrow$	H^1	I^1	K^1	$\downarrow$	
2		A^{15}		B^1	$\uparrow$	H^1	I	K^{10}	$\downarrow$	
3		$A^{1,7}$		$B^{1,4}$	$C\uparrow$	H^1	$I^{1,2}$	K^1	$\downarrow$	MU
4		$A^{1,8,15}$		B^1	$C\uparrow$	G^2H^1	I^1	K^{10}		MU
5		A^1		B^1	$C\uparrow$	H^1	I^1	K^{10}	$\downarrow$	
6	$\varepsilon^3\zeta^2$	A^1	a^1	B^1	$\uparrow$	H^1	I^1	K^1	$\downarrow$	U
7		$A^{1,(15,16,17)}$		$B^{(2)}$	$C\uparrow$	H^1	I^1	$(K^{10})\downarrow$		
8		A^1		B^1	C	$F^3G^2H^1$	I^1	K^{10}	$\downarrow$	
9		A^1	a^1	B^7	$\uparrow$	H^1	I^3		$\downarrow$	
10		A^1		B^1	C	H^1	I^1	$(K^1)?$		
11		$A^{1,17}$		B^1	$C\uparrow$	G^2H^1	I^1	K^{10}	$\downarrow$	
12		$A^{1,15}$		B^1	$C\uparrow$	H^1	I^1	$K^{1,10}\downarrow$		
13		A^1		B^1	$\uparrow$	G^2H^1	I^1	K^{10}	$\downarrow$	
14		A^1		B^1	$C\uparrow$	$\left\{\begin{array}{cc} H^1 & I^1 \\ H^1 & I^1 \end{array}\right\}$		K^1	$\downarrow$	
15		A^1		$\left\{\begin{array}{c} B^1 \\ B^7 \end{array}\right\}C\uparrow$		H^1	I^1		$\downarrow$	
16	η^3	A^1		$B^{1,7}$	$C\uparrow$	H^1	I^1		$\downarrow$	
17	β^3	$A^{1,15}$		$B^{1,7}$	$C\uparrow D^5E^5$		I^2		$\downarrow$	
18		A^1		B^1	$C\uparrow$	H^1	I^1	$K^{1,10}\downarrow$		

Regarding K^1 (ts 10) and K^{10} (ts 7), we find a remark that the test subject considered these as (potentially, 10) implicit.

which involve an 'object of search' which is not an acting character do not contain a *Princess*. Similarly, not all tales which have a character which is a princess contain, or have said princess fill the role of, a *Princess*.

There is also a tension between the *dramatis personae* as a category restricted to agents, and the fact that the variable slots in some functions normally take a dramatis persona. When *Wedding* corresponds to a monetary reward, one is tempted to conceptualize the money as analogue to the princess. However, the money cannot perform actions, and hence cannot be dramatis persona, so that the *dramatis personae* assignment appears as an artefact of the formalization process, so that it may be useful to drop this part of the experiment for further experiments.

All these problems may show that despite training, the abstract meaning of the *dramatis personae* is difficult to convey to test subjects and the colloquial intension of the terms (cf. note 4) is difficult to get rid of.

5.5.5 *Methodological conclusion*

We interpret our data to show that giving theoretical background and teaching Propp in depth

Table 11 *Dramatis Personae* for *Nikita* in **Propp IV**

Test subject	H	V	P	Di	MH
1	Nikita	Der drache	Die zarentochter	Die zarentochter	
	Nikita	*The dragon*	*Tsar's daughter*	*Tsar's daughter*	
2	Nikita	Drache	Zarewna	Zar (/)	
	Nikita	*Dragon*	*Tsarevna*	*Tsar (/)*	
3	Nikita der Gerber	Drache	Zarewna	Vater + mutter	
	Nikita the Tanner	*Dragon*	*Tsarevna*	*Father + mother*	
4	Nikita	Drache	Zarentochter	Zar/vater	
	Nikita	*Dragon*	*Tsar's daughter*	*Tsar/father*	
5	Nikita der Gerber	Drache		Der zar	
	Nikita the Tanner	*Dragon*		*The tsar*	
6	Nikita	Drache	Zarewna	Zar	5000 kinder
	Nikita	*Dragon*	*Tsarevna*	*Tsar*	5000 *children*
7	Nikita	Drache	Zarentochter	Zar	
	Nikita	*Dragon*	*Tsar's daughter*	*Tsar*	
8	Nikita	Drache	Zarentochter	Zar	
	Nikita	*Dragon*	*Tsar's daughter*	*Tsar*	
9	Nikita	Drache	Prinzessin		
	Nikita	*Dragon*	*Princess*		
10	Nikita	Drache	Zarentochter	Zar	
	Nikita	*Dragon*	*Tsar's daughter*	*Tsar*	
11	Nikita der Gerber	Drache	Zarentochter	Zar	
	Nikita the Tanner	*Dragon*	*Tsar's daughter*	*Tsar*	
12	Nikita	Drache	Zarentochter	Zar	
	Nikita	*Dragon*	Tsar's daughter	*Tsar*	
13	Nikita	Drache	Die zarentochter	Zar	
	Nikita	*Dragon*	*Tsar's daughter*	*Tsar*	
14	Nikita	Drache	Zarentochter	Zar	
	Nikita	*Dragon*	*Tsar's daughter*	*Tsar*	
15	Nikita	Drache	Zarentochter	Zar	
	Nikita	*Dragon*	*Tsar's daughter*	*Tsar*	
16	Gerber	Drache	Zarentochter „Zarewna"	Zar	
	Nikita	*Dragon*	*Tsar's daughter 'Tsarevna'*	*Tsar*	
17	Nikita	Drache	Zarewa	Zar	
	Nikita	*Dragon*	*Tsarev[n]a*	*Tsar*	
18	Nikita	Drache	Zarentochter	Zar	
	Nikita	*Dragon*	*Tsar's daughter*	*Tsar*	

allowed test subjects to annotate simple stories in his system consistently.

5.6 General discussion

During the experiments, the level of expertise of the experimenters with respect to Propp's system and how to teach it increased. The group discussions revealed methodological issues for training test subjects in Propp's framework, e.g. our experience informed us which issues required more intensive discussion during the training, and this helped us to improve our training. Methodologically, this means that the experiments are not directly comparable, but the observations presented here allow us to assess the reproducibility of Propp's system. Two aspects are particularly important.

First, the abstract nature of both the functions and the *dramatis personae* is not easy to apprehend, as it requires us to ignore or suppress the naïve natural language meaning of the labels (e.g. Wedding) to a large extent. The interdependent definition of the functions and *dramatis personae* must be stressed. Secondly, it is important to emphasize that functions are not only defined by their content, but in addition by their place in the story (cf. the quote in note 5). This is one of the

Table 12 *Dramatis Personae* for *Shabarsha* in **Propp IV**

Test subject	H	V	P	Di	Do	MH
1	Schabarscha	Der Teufel	–	Der Herr	Der Teufel	
	Shabarsha	*The devil*	*–*	*The master*	*The devil*	
2	Schabarscha	Junger Teichteufel	Gold	Herr		
	Shabarsha	*Young pond devil*	*Gold*	*Master*		
2*	Junger Teufel	Schabarscha	Gold	Großvater		
	Young devil	*Shabarsha*	*Gold*	*Grandfather*		
3	Schabarscha	Teufelchen/Großvater	das Geld/Gold	Herr v. Schabarscha		
	Shabarsha	*Little devil / grandfather*	*The Money/Gold*	*Shabarsha's master*		
4	Schabarscha	Großvater + Knete, Wolf	Gold + Silber	Schabarschas Herr		
	Shabarsha	*Grandfather + money, wolf*	*Gold + silver*	*Shabarsha's master*		
5	Schabarscha	Kleiner Teufel		Herr		
	Shabarsha	*Little devil*		*Master*		
6	Schabarscha	Teufelchen	Gold & Silber			
	Shabarsha	*Little devil*	*Gold & silver*			
7	Schabarscha	Die Teufel	Geld	der Herr		
	Shabarscha	*The devils*	*Money*	*the master*		
8	Schabarscha	Teufel	Gold	Herr		
	Schabarscha	*Devil(s)*	*Gold*	*Master*		
9	Schabarscha	Knabe / dessen Stamm				
	Shabarscha	*Boy / his tribe*				
10	Schabarscha (und tierische Freunde)	Wasserteufel + Sohn	Geld	(Herr)		
	Shabarsha (and animal friends)	*Water devil + son*	*Money*	*(Master)*		
11	Schabarscha	Kleine Teufel	Gold	Schabarschas Herr	Müller	
	Shabarsha	*Little devils*	*Gold*	*Shabarsha's master*	*miller*	
12	Schabarscha	Wasserteufel	Existenzmittel	Herr		
	Shabarsha	*Water devil(s)*	*Means of existence*	*Master*		
13	Schabarscha	Der Großvater		Der Herr		
	Shabarsha	*The grandfather*		*The master*		
14	Schabarscha	Teufel & Großvater	Gold & Silber	Herr		Bär, Hase
	Shabarsha	*Devil & grandfather*	*Gold and Silver*	*Master*		*bear, hare*
15	Schabarscha	Teufel (klein)	Geld	Bauer		
	Shabarsha	*Devil(s), small*	*Money*	*Farmer*		
16	Schabarscha	kleiner Wasserteufel	Geld / Gold	der Herr		
	Shabarsha	*Little water devil*	*Money / gold*	*the master*		
17	Schabarscha	Wasserteufel	Geld	Herr	Teufelchen	
	Shabarsha	*Water devil(s)*	*Money*	*Master*	*Little devil*	
18	Schabarscha	2 Teufel	Geld	Bauer ('Herr')		
	Shabarsha	*2 devils*	*Money*	*Farmer ('master')*		

Only roles that were assigned by at least one test subject are in the table.

German *Teufel* (*devil(s)*) is ambiguous between singular and plural. We indicate ambiguity by adding *(s)* in the translation where neither context nor determiners/adjectives disambiguate the number.

Row 2*: alternative assignment of *dramatis personae* by subject 2

Row 12, regarding *means of existence*: cf. the formulation by Propp (1968, p. 35f)

central ideas of Propp's system, and it must be repeatedly stressed during the training. The (seeming) occurrence of a function out of sequence should be scrutinized, as this is only allowable under a small set of circumstances. Similarly, inferred or implicit functions must be discussed.

6 Conclusions

We conclude from our results that Propp's system is not easily grasped, and does indeed need relatively sophisticated training. Given the complexity of narrative both as a phenomenon and as a concept,

this is to be expected; however, we also could isolate some of its difficulties, as has just been discussed.

When reviewing our data, we understood better that the interpretation of text-spans assigned to functions is also not trivial because Propp's system works on the story level rather than the discourse level. This is a theoretically important point that must be respected when tales are annotated and used in machine learning, but also for human understanding: one could discuss whether genre expectations lead us to infer certain events such as Shabarsha's returned, discussed above. Such interaction of levels, and of top-down and bottom-up constraints in narrative interpretations is an important and well-known challenge for computational modelling.[20]

7 Future Work

Future work could go in two directions: deeper investigation of Propp's system, and turning to other systems.

We have deliberately used simple tales in our Propp experiments (where we defined 'simple' in terms of the complexity of Propp's original annotation string). Now that we established learnability, it would be interesting to see whether agreement can also be achieved on more difficult tales. However, we should also keep in mind that this would turn our research into the evaluation of one particular system. Even if it is a venerable and admirably pioneering system, it would be at least as important to turn to other candidates for narrative frameworks.

Especially frameworks that have been designed for a special purpose such as Plot Units or Story Grammars (both designed to help summarization) should be put to a similar test of inter-annotator agreement.

References

Afanas'ev, A. N. (1973). *Russian Fairy Tales*. 2nd edn. New York: Pantheon. (Translation by Norbert Guterman from the collections of Aleksandr Afanas'ev. Folkloristic commentary by Roman Jakobson).

Afanas'ev, A. N. (1985a). *Russische Volksmärchen. In neuer Übertragung von Swetlana Geier*. München: Winkler.

Afanas'ev, A. N. (1985b). Shabarsha, translated by Kathleen Cook. *The three kingdoms. Russian fairy tales from Alexander Afanasiev's collection, illustrated by Alexander Kurkin*. Moscow: Raduga Publisher.

Afanas'ev, A. N. (2008). *Russische Volksmärchen, übersetzt von Anna Meyer*. Köln: Anaconda.

Artstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, **34**(4): 555–96.

Bod, R., Fisseni, B., Kurji, A., and Löwe, B. (2012). Objectivity and reproducibility of Proppian narrative annotations. In Finlayson, M. A. (ed.), *The Third Workshop on Computational Models of Narrative*. Istanbul, pp. 17–21.

Bod, R., Löwe, B., and Saraf, S. (2011). How much do formal narrative annotations differ? A Proppian case study. In Ess, C. and Hagengruber, R. (eds), *The Computational Turn: Past, Presents, Futures?, Aarhus University, July 4–6, 2011*. Münster: MV-Wissenschaft, pp. 242–245.

Bortz, J. and Döring, N. (2002). *Forschungsmethoden und Evaluation. für Human- und Sozialwissenschaftler*. 3rd edn. Berlin, New York: Springer.

Bremond, C. and Verrier, J. (1982). Afanassiev et Propp. *Littérature*, **451**: 61–78.

Dyer, M. G. (1983). *In-depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension*. Cambridge, MA: MIT Press.

Finlayson, M. A. (2011). *Learning Narrative Structure from Annotated Folktales*. Doctoral dissertation, Massachusetts Institute of Technology. http://www.mit.edu/markaf/doc/finlayson.phdthesis.mit.2011.pdf.

Fisseni, B., Kurji, A., Sarikaya, D., and Viehstädt, M. (2013). Story comparisons: Evidence from film reviews. In Finlayson, M. A., Fisseni, B., Löwe, B., and Meister, J. C. (eds), *2013 Workshop on Computational Models of Narrative*, Vol. 32. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, pp. 94–9.

Fisseni, B. and Löwe, B. (2014). Event-mappings for comparing frameworks for narratives. *Logique et Analyse*, **226**: 181–222.

Gervás, P. (2013). Propp's Morphology of the Folk Tale as a Grammar for Generation. In Finlayson, M. A., Fisseni, B., Löwe, B., and Meister, J. C. (eds), *2013 Workshop on Computational Models of Narrative*, Vol. 32. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, pp. 106–22.

Labov, W. (1997). Some further steps in narrative analysis. *Journal of Narrative and Life History*, **7**: 395–415.

Lehnert, W. G. (1981). Plot units and narrative summarization. *Cognitive Science*, **5**(4): 293–331.

Lendvai, P., Declerck, T., Darányi, S., Gervás, P., Hervás, R., Malec, S., and Peinado, F. (2010). Integration of linguistic markup into semantic models of folk narratives: The fairy tale use case. In Calzolari, N. *et al.* (eds), *Proceedings of the Seventh Conference on International Language Resources and Evaluation*. Paris: European Language Resources Association (ELRA), pp. 1996–2001.

Löwe, B. (2011). Methodological remarks about comparing formal frameworks for narratives. In Allo, P. and Primiero, G. (eds), *Third Workshop in the Philosophy of Information, Contactforum van de Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten*. Brussel: KVAB, pp. 10–28.

Pavel, T. G. (1982). Plot-structure and style: Remarks on an unstable relationship. *Canadian Review of Comparative Literature*, **9**(2): 27–45.

Propp, V. Ya. (1928). *Morfologiya skazki*. 1st edn. Leningrad: Akademiya.

Propp, V. Ya. (1958). *Morphology of the folktale*. 1st edn. Bloomington: Indiana University Press. English translation of Propp, 1928.

Propp, V. Ya. (1968). *Morphology of the folktale*. 2nd edn. Austin, Texas: University of Texas Press. English translation of Propp, 1928.

Propp, V. Ya. (1969). *Morfologiya skazki*. 2nd edn. Moskva: Nauka.

Propp, V. Ya. (1972a). *Morphologie des Märchens*. München: Carl Hanser. Deutsche Übersetzung von Propp, 1969.

Propp, V. Ya. (1972b). *Morphologie du conte*. Paris: Points. traduction française de Propp, 1969.

Ralston, W. R. S. (no year). *Russian Fairy Tales*. New York: Hurst & Co. http://www.gutenberg.org/ebooks/22373.

Rumelhart, D. E. (1975). Notes on a schema for stories. In Bobrow, D. G. and Collins, A. M. (eds), *Representation and Understanding: Studies in Cognitive Science*. New York: Academic Press, pp. 211–236.

Schank, R. C. (1982). *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*. Cambridge, UK: Cambridge University Press.

Turner, S. (1994). *The Creative Process. A Computer Model of Storytelling*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Notes

1 For instance, for the case of film reviews, Fisseni *et al.* (2013) report the need to integrate pragmatic and reception-oriented factors into the model, both which go beyond current practice in formal or computational models of narrative.

2 Reproducibility is not a sufficient condition for intersubjective agreement with respect to the structural core, but must be supported by some kind of *content validity* (Bortz and Döring, 2002, p. 200), i.e. some indication that the structural core is indeed represented adequately by the formal structure.

3 We generally quote from the English translation Propp (1968) for ease of use. However, this edition was made from the first edition of Propp's work. As Propp changed his annotations in the second edition, we sometimes report wherever we noticed differences in the French and German translations Propp (1972a,b) of the second edition Propp (1969).

4 The persona of *Princess/Princess's Father* (**P/PF**) can be split into two with a slightly difficult delineation. We used the resulting list of eight *dramatis personae* in the first experiment, but changed this after a reviewer of Bod *et al.* (2012) pointed out that it was preferable to only use seven *dramatis personae*.

5 '*Function is understood as an act of a character, defined from the point of view of its significance for the course of the action*'. (Propp, 1968, p. 21, original emphasis)

6 'For the most part, the samples far from exhaust our material,' (Propp, 1968, p. 25)

7 This means that mainly roman-numbered and negated subfunctions were not discussed at length.

8 'In doing this [i.e. composing a tale], any elements may be dropped (except possibly for A or *a*), or repeated three times, or repeated in various forms.' (Propp, 1968, p. 111); cf. also (Propp, 1968, p. 102).

9 For example, 'Return is implied, but not mentioned. ($\downarrow$).' (Propp, 1968, p. 98)

10 In the first edition, this expression is not yet given, but only the preceding step of the derivation:

$$\mathbf{A\,B\,C}{\uparrow}\mathbf{D\,E\,F\,G}\frac{\mathbf{H\,J\ \ I\,K}{\downarrow}\mathbf{Pr-Rs\,oL}}{\mathbf{L\,M\,J\,N\,K}{\downarrow}\mathbf{Pr-Rs}}\mathbf{Q\,Ex\,T\,U\,W}^*.$$

This expression is corrected in accordance with the second edition with respect to **Rs o**, which—after symbol translation—is **Rs o**.

11 Most pointedly: 'Propp tranche la difficulté en couchant toute espèce de contes merveilleux, quelles que soient leur provenance, leur histoire et leur forme, sur le lit de Procuste du conte type AT 300, *Le Tueur de dragon*.' (Bremond and Verrier, 1982, p. 77). Bremond and Verrier also doubt the typological consistency of Propp's corpus (p. 61f); this might imply serious problems for investigations such as ours or Finlayson's (2011).

12 Gervás (2013) discusses how Propp's system could be transformed into a modern formal system or grammar.

13 The translation into German by Geier (Afanas'ev 1985a) did not come to our attention until after the second experiment. It comprises, among others, all stories Propp lists in his corpus and the Afanas'ev number, which is missing from many editions of the Russian fairytales. For reasons of readability, we decided to use this translation mainly to check our understanding of the Russian tales, and check whether we had the right version of the story. The problem with the *Semyons* (cf. note 15) was discovered thus. We then chose to not include this story in the experiments of **Propp III** and **IV**, but did use it during the example and discussion sections.

14 For *Ivan Popyalov* (135), repetition is removed from the annotation in the second edition of Propp's *Morphology* (Propp, 1972a, p. 138f). However, the second edition introduces an inversion.

15 Propp annotated version 145 of *Semyons*; however, due to an oversight we worked with one of the other non-annotated versions of the story. Thus our results cannot be compared to Propp's original annotation (cf. note 17), but it does not invalidate the discussion of inter-annotator agreement of our test subjects.

16 A brief description of the tales: In *Semyons*, seven orphans meet the Tsar and pledge to work hard in their professions. The seventh becomes a thief and, with the help of his brothers and their respective talents, journeys to capture Elena the fair as a bride for the Tsar. In *Ivanko*, Ivanko is born of a peasant woman and her kidnapper, a bear. After returning to human society, he causes some damage and is sent to a lake in which devils dwell. Through a series of tricks, Ivanko gains all of the devils' gold and the services of a little devil for a year. In *Shabarsha*, the protagonist Shabarsha takes a day off to earn some money for himself and his boss. He goes to a lake to catch fish, meets a little devil, and threatens to evict all of the devils from the lake if they do not pay rent. Through a series of tricks he acquires all of their wealth. In *Nikita*, a dragon robs and devours innocent maidens, and finally kidnaps the tsar's daughter. Nikita the Tanner fights the dragon, tricks it, and slays it, but does not want to be rewarded. In *Popyalov*, a dragon robs the daylight. Ivan, a stupid lazy boy, goes to fight the dragon and save the day, supported by his brothers. The dragon's

wife pursues Nikita who hides in a smithy; he and the smiths kill her.

17 The Propp string for *Semyons* (version 145, cf. note 15) is $\mathbf{a}^1\mathbf{B}^2\mathbf{C}{\uparrow}\mathbf{F}^3\mathbf{G}^1\mathbf{K}^1{\downarrow}$ in the English edition ($\mathbf{B}^3$ in the first Russian edition) and $\mathbf{a}^1\mathbf{B}^3\mathbf{C}{\uparrow}\mathbf{G}^1\mathbf{K}^2{\downarrow}\mathbf{W}^*_*$ (where $\mathbf{W}^*_*$ reflects the marriage to the 7th Semyon and his accession to the throne) in the second edition.

18 It is conceivable that the term 'devil' triggered a connotation in the original readers of the folktale, producing a very different reading of Shabarsha's behaviour that cannot be reproduced in contemporary test subjects due to a lack of cultural context and contemporary sympathy for harmless 'little devils'.

19 While we recognized that the problem of annotating *implicit functions* remained with annotating spans of text, we retained them to continue to use our stability measure and be able to compare results across experiments.

20 In that sense, the idea that one structure fits many stories is not only a reductionist Procrustean bed (cf. note 11) but also a strategy of semantic enrichment.