Objectivity and reproducibility of formal narrative representations or annotations: Propp's functions & narrative summarization

Alexander Block[1,5], **Rens Bod**[1], **Bernhard Fisseni**[1,2,3], **Adil Kurji**[1], **Carlos León**[3,4,5], **Benedikt Löwe**[1,3,5,6], **Deniz Sarikaya**[5]

[1] Institute for Logic, Language and Computation, Universiteit van Amsterdam
Postbus 94242, 1090 GE Amsterdam, The Netherlands
{alexander.block,aadil.kurji}@student.uva.nl, {l.w.m.bod,b.loewe}@uva.nl
[2] Fakultät für Geisteswissenschaften, Universität Duisburg-Essen
Universitätsstraße 12, 45117 Essen, Germany
bernhard.fisseni@uni-due.de
[3] Isaac Newton Institute for Mathematical Sciences
20 Clarkson Road, Cambridge CB3 0EH, United Kingdom
[4] Escuela Politécnica Superior, Universidad Alfonso X el Sabio
28691 Villanueva de la Cañada, Madrid, Spain
cleonazn@uax.es
[5] Fachbereich Mathematik, Universität Hamburg
Bundesstraße 55, 20146 Hamburg, Germany
deniz.sarikaya@studium.uni-hamburg.de
[6] Corpus Christi College, University of Cambridge
Cambridge CB2 1RH, United Kingdom

Abstract

Formal narrative representation is a procedure assigning a formal description to a natural language narrative. In general, it is a human procedure, and one of the goals of *computational models of narrative* is to understand this procedure better in order to automatise it. In order to be automatisable, a formal framework should allow for objective and reproducible representations. In this paper, we present empirical work focussing on objectivity and reproducibility of the Proppian framework and the hypothesis that narrative formalisation is summarisation. The first two experiments consider Propp's formalisation of Russian fairy tales; the third compares these results to summaries of the same stories. The data show that some features of Propp's system such as the assignment of the characters to the *dramatis personae* and some of the functions are difficult to reproduce; furthermore, natural summaries of folktales do not match the Proppian functions.

1. Introduction & Motivation

The formal study of narratives goes back to the Russian structuralist school, paradigmatically represented by the Vladimir Propp's 1928 study *Morphology of the Folktale* (Propp, 1928) in which he identifies seven dramatis personae and 31 functions that allow him to formally analyse a corpus of Russian folktales (cf. § 2.1. for details).

Researchers in the field of *computational models of narrative* have developed the general Proppian methodology into various formal and computational frameworks for the analysis, automated understanding and generation of narratives. Examples for this are Lehnert's Plot Units, Rumelhart's Story Grammars, Schank's Thematic Organization Points (TOPs), Dyer's Thematic Abstraction Units (TAUs), or Turner's Planning Advice Themes (PATs).[1]

While the biggest advances of this research community were on the technical side, recent years have seen an increased interest in the methodological and conceptual issues involved, linking this research closely to questions of the philosophy of information.[2]

[1] Cf. (Lehnert, 1981; Rumelhart, 1980; Schank, 1982; Dyer, 1983; Turner, 1994).

[2] Cf. (Löwe, 2011).

The enterprise of representing a narrative by a formal object formalizing its structure that can then be used in computational application rests on a number of assumptions:

Assumption E (Existence of a structural core). There is a structural core of narratives; or several, depending on which part of the structure we are interested in.

Assumption O (Objectivity of the structural core). Given a narrative, there is an interpersonal agreement what its structural core is; possibly after some agreement of what part of the structure should be represented.

A formal framework Λ for representing narratives consists of a formal language $\mathcal{L}_\Lambda$, a class of mathematical structures $\mathcal{M}_\Lambda$, and a description of a procedure (called *formalization* in (Löwe, 2011)) of assigning to each natural language narrative N a structure $\Sigma_\Lambda(N) \in \mathcal{M}_\Lambda$. Note that this procedure that assigns $\Sigma(N)$ to N is not a function in the mathematical sense, but an activity by expert formalizers who follow given guidelines to produce $\Sigma_\Lambda(N)$ on the basis of N.

In this paper, we mainly explore the validity of Assumption **O**: in particular, we are investigating the following property of formal frameworks Λ:

Property Obj(Λ). Sufficiently trained human formalizers, given the same narrative N will produce the same structure $\Sigma_\Lambda(N)$.

Property **Obj** is an important (and arguably necessary) feature of a formal framework Λ if it is supposed to be the basis of an automatised system. The existence or non-existence of formal frameworks Λ with property **Obj**(Λ) is closely related to Assumption **O**: if no formal frameworks proposed in practice has this property, this is evidence against Assumption **O**. Checking **Obj**(Λ) was described in (Bod et al., 2011) as a natural analogue of the study of annotator agreement in corpus linguistics and computational linguistics: whereas typical annotation tasks involve annotation of sentences or discourses,[3] the formalization or annotation of a narrative is at the next level of complexity, involving sequences or systems of discourses, connected to a narrative. At the sentence or discourse level, inter-annotator agreement has been studied (Carletta et al., 1997; Marcu et al., 1999), but for the annotation or formalization of narratives, no such analysis has ever been done, not even with the oldest and best-known formal approach to narrative structure, the Proppian narratemes.

We therefore decided to focus on this particular formal framework, not because we feel that the Proppian framework is a good candidate for a framework close to the *stable structural core*, but mostly due to its prominent place in the history of formal representations of narrative. In § 2., we describe the Proppian formal framework and discuss two empirical studies called **Propp I** and **Propp II** pertaining to it and performed at the *Universiteit van Amsterdam*.

There is a very close relationship between formal representation of narrative and summarization: simple formal representations (e.g., *Plot Units*) should serve as a summary of a narrative, and similarly, a formalized theory of narrative summarization could give rise to high-quality formal frameworks for the representation of narrative. In § 3., we describe a follow-up empirical study called **Summarization** done at the *Universität Hamburg* with the same narratives that were used in the two studies described in § 2. In this study, we investigated natural summarization techniques and their relation to Propp's functions as resulting from the two Amsterdam studies.

2. Propp's formal system

2.1. Overview of Propp

Working with a corpus of 100 Russian folktales, from the book *Narodnye Russkie Skazki* by Alexander Afanas'ev (Afanas'ev, 1973), Vladimir Propp developed a formal system to identify each folktale by short annotation strings consisting of symbols representing Proppian *functions* or *narratemes*. In the following, we give a description of the components of the Proppian system relevant for the experiments discussed in this paper. For more details, we refer the reader to (Propp, 1928).

Propp identified eight *dramatis personae* representing roles that the various characters inhabit within the stories. Not every *dramatis persona* occurs in each story, not every character represents a *dramatis persona*, and some *dramatis personae* can be represented by the same character. The Proppian *dramatis personae* are: the hero (**H**), the villain (**V**), the princess (**P**), the princess's father (**PF**), the dispatcher (**Di**), the donor (**Do**), the magical helper (**MH**) and the false hero (**FH**) (Propp, 1928, § 3).

The actions of the *dramatis personae* are described by a set of thirty-one functions described in (Propp, 1928, § 3) by means of examples and more specified subfunctions (cf. Figure 1 for an example). These functions are marked by symbols in the order of their occurrence in the folktale: $\boldsymbol{\beta}$ Absentation; $\boldsymbol{\gamma}$ Interdiction; $\boldsymbol{\delta}$ Violation, $\boldsymbol{\varepsilon}$ Reconnaissance, $\boldsymbol{\xi}$ Delivery, $\boldsymbol{\eta}$ Trickery, $\boldsymbol{\theta}$ Complicity, **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** Pu! rsuit, **Rs** Rescue, **o** Unrecognized Arrival, **L** Unfounded Claims, **M** Difficult Task, **N** Solution, **Q** Recognition, **Ex** Exposure, **T** Transfiguration, **U** Punishment, **W** Wedding.

These functions occur in strict sequential order, i.e., functions have to occur in the folktale in the order they are given in the list above: after function **I** (Victory) has occurred, none of the functions listed earlier can occur anymore.

The full Proppian system contains a number of additional features: some folktales contain a series of individual tale units; the Proppian system allows for *moves* within an annotation. Since none of the tales we used in the experiment had these features, we shall not further discuss this here. Another additional feature is a special annotation symbol for *trebling*: in folktales, a common motif is the triple repetition of some functions called *trebling*. Propp's system allows for annotating occurrences of trebling, and we'll comment on this in § 3.3.

False Hero. The FALSE HERO is a person who tries to take credit for the actions of the hero. He may try to marry the PRINCESS, or gain respect of PRINCESS'S FATHER to further complicate the HERO'S pursuit of the PRINCESS.

$\boldsymbol{\eta}$ Trickery. The villain attempts to deceive his victim in order to take possession of him or of his belongings.

Villain first assumes a disguise. Then follows the function:
Example 1. By persuasion: e.g. A witch tries to have a ring accepted.
Example 2. By direct application of magic: e.g. Step mother gives a sleeping pill.

Figure 1: Examples of descriptions of a *dramatis persona* and a function used in the experiments **Propp I** and **Propp II**.

2.2. Description of Propp I

In the experiment **Propp I**, test subjects were briefly trained in the Proppian framework and then asked to annotate some of the folktales formalized in (Propp, 1928). We used the

[3] Cf., e.g., (Marcus et al., 1993; Brants, 2000; Passonneau et al., 2006).

folktales *The Seven Semyons, 145*, *Shabarsha, 151*, and *Ivan the Bear's Son, 152*; in the following, we refer to these folktales as *Semyons*, *Shabarsha*, and *Ivanko*.[4] We chose tales that were available in English translation, did not exhibit the Proppian phenomenon of *moves*, and used few functions in Propp's own formalisation (*Ivanko* and *Semyons* use eight functions, *Shabarsha* six). An annotation of a folktale in **Propp I** consisted of

1. the assignment of story characters to the *dramatis personae*, and
2. a list of the functions (group 1) or functions specified by subfunctions (group 2) occurring in the folktale.

Procedure. We had nine test subjects, all students of the *Universiteit van Amsterdam*, and all with native or near-native competence of English. The test subjects were split into two groups: in the first group, only the functions had to be given; in the second group, the function had to be specified with the subfunction. Test subjects were instructed that the experiment would last three hours and received a moderate financial compensation for participation.

During the experiment, test subjects were given a 45-minute introduction to Propp's system given by a native speaker of English supported by a projector presentation. It explained the Propp's system (without *moves* and *trebling*), i.e., the *dramatis personae* and their functions. We gave short descriptions of the *dramatis personae* roughly based on Propp's original text and the the descriptions of the functions from Propp's text. Only a selection of the subfunctions was included (for subjects in group 1, these were marked as "examples", for subjects in group 2, they were numbered subfunctions). Finally, we analyzed a simple example story designed by the experimenters as an illustration.

Figure 1 shows examples of the description of a *dramatis persona* and a function as we used them in the presentation. A condensed version of the description of Proppian *dramatis personae* and functions was distributed as a leaflet to test subjects for use during the annotation.[5]

[4] In **Propp I**, we also used the folktale *The Enchanted Princess*, but it turned out to be too long and was omitted in the other experiments. Due to an oversight, we worked with version 147 of *Semyons* while Propp had annotated version 145. We used the translations of Gutermann from (Afanas'ev, 1973) for *Semyons* and *Ivanko*, and the translation of Cook from (Afanas'ev, 1985) for *Shabarsha*. 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 don't pay rent. Through a series of tricks he acquires all of their wealth.

[5] To simplify the annotations, we used the symbols A–Z and a–f for the Proppian functions.

Results. Propp's own annotations only contain the function string, and do not specify the assignment of characters to the *dramatis personae*. The original Propp strings for the narratives are: $\mathbf{a}^1\,\mathbf{B}^2\,\mathbf{C}\uparrow\mathbf{F}^2\,\mathbf{G}^1\,\mathbf{K}^2\downarrow$ (*Semyons*), $\mathbf{A}^8\,\mathbf{B}^4\,\mathbf{C}\uparrow\mathbf{H}^2\,\mathbf{I}^2\,\mathbf{K}^1\downarrow$ (*Shabarsha*), and $\mathbf{A}^9\uparrow\mathbf{H}^2\,\mathbf{I}^2\,\mathbf{K}^1\downarrow$ (*Ivanko*).[6]

We give the results of the assignments of the characters to the *dramatis personae* in Table 1. We see some amount of variation even in the assignment of the three main *dramatis personae*, **H**, **V**, and **P**: consider, e.g., the variations among the choices of villain in the *Ivanko* story or the *Semyons* or the choice of hero in the *Shabarsha* story (see below for a methodological remark).

The annotation strings vary widely and are given in Table 2. Test subjects 1–5 belonged to group 1 (no subfunction) and test subjects 6–9 to group 2 (subfunctions). The subfunctions are marked by superscripts in the table, and a missing subfunction is marked by a superscript of $^{\varnothing}$. No strings matched across all annotators and all stories, and thus no inter-annotator analysis was necessary. It is interesting to notice that the annotation strings are considerably longer than Propp's original strings (compare an average of 14.2, 13.2, and 12.8 functions for *Ivanko*, *Semyons*, and *Shabarsha*, respectively, with the Proppian string lengths of 6, 8 and 8 for the same folktales).

Methodological Conclusion. In their post-experiment comments, several test subjects reported that they considered the task as easy. Four out of nine test subjects reported that the example story from the presentation was considerably simpler than the actual folktales.

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 read "The hero can be a prince, a poor girl, a bear-man, a group of soldiers or anyone who is good and sets out on an adventure". Arguably, Shabarsha's behaviour in *Shabarsha* cannot be described as "good", which has caused some of the variation in the assignment of the hero in that folktale.[7]

A number of functions are consistently annotated by the annotators that do not show up in Propp's annotation: for instance, *Ivanko* has nine annotations with function $\boldsymbol{\beta}$, eight with **B** and six each with $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, $\boldsymbol{\zeta}$, $\boldsymbol{\theta}$, **a**, and **C**, none of which occur in Propp's annotation.

On the other hand, we see that some functions used by Propp show up in all or almost all annotations strings: e.g., $\uparrow$, **G**, **H**, **I**, and $\downarrow$ 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.

2.3. Description of Propp II

From **Propp I**, we learned that

[6] Cf. Footnote 4.

[7] It is conceivable that the designator "devil" created 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.

Ivanko Test subject	**H**	**V**	**P**	**PF**	**Di**	**Do**	**MH**	**FH**
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	

Semyons Test subject	**H**	**V**	**P**	**PF**	**Di**	**Do**	**MH**	**FH**
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

Shabarsha Test subject	**H**	**V**	**P**	**PF**	**Di**	**Do**	**MH**	**FH**
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			

Table 1: The assignment of the *dramatis personae* for the three folktales in **Propp I**.

1. the variations in the assignment of the characters of the *dramatis personae* made comparison of the annotation strings difficult,
2. we did not know because of which passage in the text the test subjects marked a function as present in a folktale,
3. the constructed example story was considered too simple in comparison with the actual folktales.

It should be noted that we Propp only recorded the annotation strings, so that his choice of *dramatis personae* and text passages would have to be extrapolated from (Propp, 1928).

The experiment **Propp II** was a modified version of **Propp I**, taking these lessons into account. We used the same folktales as in **Propp I**. 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.

The main changes to **Propp I** were: the test subjects were given the assignment of characters to the *dramatis personae*; subfunctions were not discussed at all; the example story was from Propp's own corpus.

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 three hours and received a moderate financial compensation for participation.

During the experiment, test subjects were given an introduction to Propp's system given by a native speaker of English supported by a projector presentation. It explained the Propp's system (without *moves* and *trebling*), i.e., the *dramatis personae* and their functions. There was no explicit mention of subfunctions: a selection of subfunctions was included as "examples". The presentation lasted 45 minutes and finished by presenting the analysis of an example folktale from the Propp corpus (*Ivan Popyalov, 135*). Again, the condensed version of the description of Proppian *dramatis personae* and functions was distributed as a leaflet to test subjects for use during the annotation. This time, test subjects were given an assignment of characters to the *dramatis personae* together with each folktale (this assignment was done by the experimenters on the basis of the Proppian annotation strings):[8] In *Ivanko*, we assigned Ivanko to **H** and the Little Devil and the Grandfather jointly to **V**; in *Semyons*, we assigned the seventh Semyon to **H**, Elena the Fair to **P**, and the Tsar to **Di**; finally, in *Shabarsha*, we assigned Shabarsha to **H** and the Little Devil and the Grandfather jointly to **V**.

Results. We give the results of the function annotation in Table 3. The annotation strings are noticeably shorter

[8] Propp's own assignment of characters to *dramatis personae* is not explicit in (Propp, 1928); our assignment is consistent with Propp's annotations.

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

Test subject	Proppian functions for *Semyons*
Propp	$\mathbf{a}^1$ **B**^{2}**C**↑ $\mathbf{F}^3\mathbf{G}^1$ $\mathbf{K}^2\downarrow$
1	β **a B C**↑ **G** **K** ↓
2	$\beta\ \gamma\ \delta\varepsilon\zeta\eta\theta$**A a B C**↑ **G H K** ↓**PrRs W**
3	$\beta\ \gamma\ \delta\varepsilon\zeta\eta\theta$**A a DEF G** **K** ↓**Pr NQTU**
4	$\beta\ \gamma\ \delta$ **a B C**↑**D** **G HIK** ↓**PrRs Q**
5	$\beta\ \ \zeta\eta\theta$ **a B C**↑ **F G** **K** ↓
6	$\beta^{\varnothing}\gamma^1\ \delta\ \zeta$ $\mathbf{a}^1\mathbf{B}^2$ ↑ $\mathbf{G}^{\varnothing}$ $\mathbf{K}^1$ **Pr N** $\mathbf{W}^6$
7	$\mathbf{a}^5\mathbf{B}^1$**C**↑ $\mathbf{F}^3\mathbf{G}^3$ $\mathbf{K}^1\downarrow$ $\mathbf{W}^6$
8	$\beta^2\gamma^1$ $\mathbf{a}^1\mathbf{B}^1$**C**↑ **G**^{3} $\mathbf{K}^2\downarrow$ $\mathbf{W}^6$
9	$\beta^2\gamma^{\varnothing}\delta\ \zeta$ $\mathbf{a}^5\mathbf{B}^2$ ↑ $\mathbf{G}^{\varnothing}$ $\mathbf{K}^2\downarrow$**Pr N** $\mathbf{W}^6$

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

Table 2: The annotation strings for the three folktales in **Propp I** (cf. Footnote 4).

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 in general more similar to the Proppian strings, but we still do not have matching strings among the test subjects.

We say that a function occurs *stably* in **Propp II** if at least four of the six annotators list it. In Table 4, we list the stable functions (five for *Ivanko*, six for *Semyons*, and four for *Shabarsha*). Of these 15 stable functions, 12 are listed in Propp's original annotations; this is the majority of the functions listed by Propp. It is interesting to note that each of the three folktales has one stable function that is not annotated by Propp.

The areas of text assigned to the functions by the test subjects varied considerably, both in length and place. We call a stable function *strongly stable* if at least four of the six annotators list the function and the text marked by these four annotators overlaps. We call it *weakly stable* otherwise. We list strong and weak stability for the 15 stable functions in Table 4.

As an example, in *Semyons*, all annotators marked the same stretch of text as the function **W** (Wedding):

> He generously rewarded the Semyons, exempted them from land rent and head taxes...

In contrast, consider the stable function **H** (Struggle) in *Shabarsha*: five out of the six subjects annotated this function, but none of the text areas match with each other. One of the annotations covered approximately half the text of the entire narrative. The other four marked stretches were:

Text passage 1. While he was working a little boy in a black jacket and a red cap jumped out of the water onto the bank. "What are you doing, uncle?" he asked. "Making some twine." "What for?" "I'm going to clean up the pond and pull you devils out of the water." "Oh no! Wait a moment, I'll go and tell my grandad."

Text passage 2. "Shabarsha! Hey, Shabarsha! Grandad says we must see who can whistle the loudest." "Alright, you whistle first." The devil boy whistled so loudly that Shabarsha could hardly keep on his feet, and the leaves fell off the trees. "Not

Subj.	*Ivanko*
Propp	**A ↑ HIK↓**
1	**β a ↑GHIK↓**
2	**$\beta\gamma$ ↑ MN W**
3	**β B↑ HI U**
4	**β ↑ HI ↓ U**
5	**β aB↑ HI ↓**
6	**β aB↑ HIK↓ W**

Subj.	*Semyons*
Propp	**aBC↑FGK↓**
1	**aB ↑ GK W**
2	**β aB ↑ K↓ W**
3	**β aB G oNW**
4	**ηa ↑ G Pr W**
5	**aB ↑ K↓PrRs W**
6	**β aBC↑ GK↓PrRs W**

Subj.	*Shabarsha*
Propp	**A BC↑HIK↓**
1	**a HIK N**
2	**aB ↑ MN W**
3	**a C↑HI M UW**
4	**a ↑HI MN**
5	**a ↑H K**
6	**aBC HIK W**

Table 3: The annotation strings for the three folktales in **Propp II** (cf. Footnote 4).

Story	Function	Occurrences	Stability
Ivanko	**β**	6	strong
	↑	6	weak
	H	5	weak
	I	5	strong
	↓	4	strong
Semyons	**a**	6	strong
	B	5	strong
	↑	5	weak
	G	4	strong
	K	4	weak
	W	6	strong
Shabarsha	**a**	6	strong
	↑	4	strong
	H	5	weak
	I	4	weak

Table 4: Stable functions in **Propp II**, marked as weakly and strongly stable

> bad," said Shabarsha, "but not as good as me! When I whistle you'll be knocked off your feet and your eardrums will split. So lie face down on the ground and put your hands over your ears."

Text passage 3. So lie face down on the ground and put your hands over your ears." The devil boy lay face down and covered his ears with his hands. Shabarsha took a heavy stick, brought it down with all his might on the devil boy's neck, and whistled.

Text passage 4. "Alright, you toss first and I'll watch."

2.4. Conclusion

The assignment of the characters to the *dramatis personae* has an important effect on the assignment of the functions. Different Proppian functions are handled very differently by the annotators: some are stable, others aren't; some are typically assigned to the same text passages, others not. It is particularly striking that some of the stably annotated functions do not show up in Propp's own annotation strings. As an illustration, we mention that subfunction 6 of **W** is listed as "Other form of compensation like a monetary reward". This vague description fits in much more general situations than Propp apparently intended.

3. Experiment Summarization

3.1. Overview of automatic summarization

In the experiment **Summarization**, we aimed at testing the hypothesis that Proppian functions formalise core events of the story which will be replicated in summaries.

Automatic text summarization has been tackled by the Natural Language Processing community from the perspectives of *machine learning* and *deep natural language analysis*. Machine learning approaches apply statistical techniques to produce document extracts (Lin and Hovy, 1997; Kupiec et al., 1995; Conroy and O'Leary, 2001). This statistical approach has proven to be successful in a wide range of domains, reaching acceptable quality levels. On the other hand, deep natural language analysis uses linguistic knowledge to process and summarize texts (Barzilay and Elhadad, 1997; Miller, 1995; Marcu, 1998). By extracting the chunks of text conveying the main message (*nuclei*), a summary can be constructed.

Human assessment of evaluation of summaries has proven to be rather unstable (Lin and Hovy, 2002), so automatic metrics like ROUGE (Lin, 2004) and BLEU (Papineni et al., 2002) have been proposed. In our comparison of summaries to the results of **Propp II**, we are mostly interested in natural human summaries.

3.2. Description of the experiment

The experiment was conducted with six students of the *Universität Hamburg*; they all had native or near-native competence of German. We used the same stories used in the **Propp II** experiment in German translations. The test subjects were given a sheet of instructions which was read to them by a native speaker of German. The instructions highlighted that the story should be *recognisable* from the summary, and that the summary should not just retell the story, and that it should not give comments on style or order of events. No precise algorithm was given how to determine the important events to mention in the summary. Test subjects were instructed to use "simple sentences" (*einfache Sätze*) to facilitate the mapping of sentences to events; simple sentences were explained to "normally consist of up to 14 words" and examples were given, which contained at most one level of sub- or co-ordination. No example of a summary was given.

Test subjects were then given 2 1/2 hours to write the summaries, and were given a modest financial compensation.

3.3. Results of the experiment

On the basis of six summaries, no quantitative analysis can be done, but certain features emerge from the data. First of all, the majority of references in the summaries refer to *events* rather than situational descriptions (our of 241 individual facts mentioned in the summaries as a summary sentence or part of a summary sentence, 22 are not events, i.e., only 9.1%). Of these events, some show up as a *core* that is mentioned by many or all test subjects, others form a cluster of events possibly mentioned.

Major Agreement. In *Semyons*, test subjects agree least which details to take into their summary. However, all agree to mention some events before the theft, esp. meeting the Tsar, presenting their plans what trade to learn and the test cases (four test subjects each). All agree in naming theft and reward, and all except one mention the wedding of Tsar and Princess. The fact that a trick was performed is only mentioned by four.
In *Ivanko* and *Shabarsha*, test subjects agree on the central events: All mention some event leading up to the competition between the hero and the little devil (Ivanko's blunders and assignment to go to the lake; Shabarsha's fishing plans or presence at the lake). The competitions are always mentioned, so is receiving the gold. The trickery is only mentioned by four (one test subject fails to mention trickery in all three tales). The final trick with gold is completely omitted in *Ivanko* summaries but mentioned by four for *Shabarsha*.

Event mapping. Since the majority of marked features of the narrative were events, we decided to construct an event mapping as follows: we identified the event descriptions that occurred in all of the summaries[9] and created a master list of the events mentioned in at least one summary. Figure 2 shows the event mapping for *Shabarsha* where the columns S1 to S6 correspond to the six test subjects in the **Summarization** experiment. In order to allow a comparison with the annotations from **Propp II**, we extended the event mapping to account for the events marked in **Pr! opp II**, as not all text passages marked by test subjects in **Propp II** were represented in the summaries. The six test subjects from **Propp II** are listed as columns P1 to P6 in Table 2.
Comparison is made more difficult by the fact that summaries may regroup events (something not allowed in a Proppian annotation due to the strict ordering of the functions). Furthermore, summary descriptions tend to be very dense, making use of certain implicatures (Grice, 1968 1989): for instance, "er geht zu einem Teich, um zu angeln" (he goes to a pond in order to catch fish), does not formally imply that he actually arrives at the pond, but this is certainly implicated in the summary.

Trebling. In two of the stories, *trebling* occurs: There are different tasks that are structurally similar and occur in order: In *Shabarsha*, there are four competitions, each consisting of a challenge, a trick and the success); in *Ivanko*, both the mistakes made and the competitions with the devil occur in several forms. In the summaries, we see that the majority of test subjects mentions these as a block and not as individual events. This corresponds to the fact that the full Proppian system has a special symbol for *trebling*.

3.4. Comparison of Propp II and Summarization

An overall comparison is difficult, because already the Propp II results on their own are so diverse. We therefore focus on three qualitative examples of differences: storylines that are omitted from function assignment, some 'stable functions' that are not present in the summaries, an example of a 'stable passage', where we find in all Propp annotations but in no summary.

Storylines. Both *Ivanko* and the *Semyons* have a *prehistory* storyline that leads up to the central story line of the competition with the devil and the theft of the princess. While Propp's full system allows for *moves* with several storylines, Propp did not annotate the pre-history events in his own annotation strings (and we did not mention *moves* as a possibility to the test subjects in **Propp I** and **Propp II**. The pre-history storylines are largely not annotated in **Propp II**, with one exception: nearly all test subjects mark β (Absentation) or **a** (lack).[10] The summaries only mention the events resulting from them: Shabarsha's plan to earn money by fishing and the order/permission to steal Elena of the Semyon's journey.
Shabarsha and the *Semyons* also have a final commentary by the narrator, which steps out of the main storyline into the narration context. These final commentaries are left out by nearly all test subjects in both experiments (except for one summary) in the *Semyons*. In *Shabarsha*, the part of the commentary relating to the main storyline is referenced by three summaries and three Propp annotators, while the final 'morale' is referenced only by three Summaries.

Stable Functions. As mentioned above, *weakly stable* functions are those which are not annotated at the same passage by Propp test subjects. We take up the discussion of the **H** (Struggle) function from the end of § 2.3.: The annotators do not agree, but mostly annotate single events in the course of the competition; the majority of the summaries regroup the single events, as discussed above (*trebling*).
In *Semyons*, four Propp annotators mark the Tsar's love for Elena as **a** (lack; *strongly stable function*), while none of the summaries mentions this condition. Similarly, the lack that affects Shabarsha's master at the beginning of the story is marked in the Propp annotations, but not mentioned in any summary.

Stable Passages. Conversely, near the end of *Ivanko*, all Propp annotators have one function for Ivanko's journey to the lake and at least one labelled event after the competitions and before the transfer of money and servant to Ivanko's father; none of the summaries mentions these explicitly. Similarly, in *Semyons*, three Propp annotators assign various functions between the theft of the princess and

[9] Here, we counted a sentence or fragment of a sentence as an *event description* if it grammatically describes a change of the state of affairs. So, for instance, "Kurz darauf verwaisen die sieben Brüder" (Shortly after that, the seven brothers are orphaned) is an event description, whereas "Sieben Waisen namens Simeon..." (Seven orphans named Semyon...) is not.

[10] The correctness of these with respect to Propp's system may be questionable, as the only affect characters not representing *dramatis personae*.

Event	Sub-Event	Sub-Sub-Event	S1	S2	S3	S4	S5	S6	P1	P2	P3	P4	P5	P6
{Sit} (Contemplation)			1	1	1			1	[a]	[a]	[a]		[a]	[a]
Free Day						1				[B/↑]	[C]	[a/↑]		[B]
§	Pond		2a	2a	2a	3a	1a	2a			[↑]		[↑]	
	Fishing Plan		2b	2b	2b	2		2b						[C]
Twine							1b							
Meet the Devil				3	3†	3b	2a	3	[H]					
Threat			3b			3c	2b	4						[C]
Protest					4									
Demand Rent †			3a†					5†						
Talk to Grandfather (= Gf)							3	6						
Assignment by Gf							4							
Competition, consisting of: { challenge, trick, success }					5									[H]
	Overall Challenge			4		4								
	Overall Trick					5a	5a	7a						
	Overall Success			5	6	5b	5b	7b						
	Wrestling	Challenge	4					6						
		Trick	5a											
		Success	5b											
	Running	Challenge												
		Trick	6a											
		Success	6b											
	Whistling	Challenge									[H]	[H]		
		Trick	7a								[H/I]			
		Success	7b								[I]			
	Throwing	Challenge								[M]	[M]	[M]	[H]	
		Trick	8b							[N]		[N]		
		Success	8a											
{additional detail: near-loss of cudgel }				6								[I]		[I]
Intervention Gf				7			6		[I]		[U]		[K]	
Demand for Gold (by Sh) †						5c†								
Gold Received			9a	8	7	6	7a	8						
	Filling Cap								[K]	[W]	[U]			
	Last Chest/Finish										[W]			[K]
Gold Trick#			9b		8		7b	9						
{Sit:Clover}									[N]		[W]			[W]
{Sit:Mead}														

Figure 2: Synoptic annotations of *Shabarsha* from **Propp II** and **Summarization**. We used the following markers: – for deficient or imcomplete descriptions, † for mistakes or errors, § for anonymous super-events, {Sit ...} for non-event (situation descriptions), * for events occurring only in Propp, and # for auxiliary events only occurring in the summaries.

her delivery to the Tsar, while no summary mentions these details.

4. Discussion & Future Work

The detailed study of human annotations of Propp's framework highlights some weaknesses of its description, and in general, points out some important obstacles for an automatisation of the process of formalisation in a computational setting. Key weaknesses are that descriptions of some of the *dramatis personae* and functions are vague and require a large amount of interpretation. We also observed that the Proppian framework encourages the marking of minor events that do not naturally occur in summaries of the same folktales. It would also be preferable accommodate more than one storyline in a framework.

In (Bod et al., 2011), we suggested to follow up the studies **Propp I** and **Propp II** with a large-scale inter-annotator study: the results of our experiments suggest that this is not worthwhile. Instead, we should distill the lessons learned from this Proppian case study into studies dealing with other formal representation systems, possibly designed and documented on the basis of the results of this study.

Acknowledgements

The research in this paper was funded by the *John Templeton Foundation* (JTF) via the project *What makes stories similar?* (grant id 20565) and the *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (NWO) via the projects *Integrating Cognition* in the VICI programme (DN 277-70-006) and *Dialogical Foundations of Semantics* in the ESF EuroCoRes programme LogICCC (LogICCC-FP004; DN 231-80-002; CN 2008/08314/GW). The third, fifth and sixth author acknowledge the financial support and the kind hospitality of the *Isaac Newton Institute for Mathematical Sciences* (programme *Semantics & Syntax*). All authors would like to thank Ekatarina Abramova and Sanchit Saraf (Amsterdam) for their work in the early set-up phase of the research.

5. References

Aleksandr Nikolayevich Afanas'ev. 1973. *Russian Fairy Tales*. Pantheon. Translation by Norbert Guterman from the collections of Aleksandr Afanas'ev. Folkloristic commentary by Roman Jakobson.

Aleksandr Nikolayevich Afanas'ev. 1985. Shabarsha, translated by Kathleen Cook. In *The Three Kingdoms. Russian Fairy Tales From Alexander Afanasiev's Collection, illustrated by Alexander Kurkin*. Raduga Publisher, Moscow.

Regina Barzilay and Michael Elhadad. 1997. Using lexical chains for text summarization. In Inderjeet Mani and Mark Maybury, editors, *Intelligent Scalable Text Summarization. Proceedings of a Workshop sponsored by the ACL*, pages 10–17, Somerset, NJ. Association for Computational Linguistics.

Rens Bod, Benedikt Löwe, and Sanchit Saraf. 2011. How much do formal narrative annotations differ? A Proppian case study. In Charles Ess and Ruth Hagengruber, editors, *The computational turn: Past, presents, futures?, Aarhus University, July 4–6, 2011*, pages 242–245, Münster. MV-Wissenschaft.

Thorsten Brants. 2000. Inter-annotator agreement for a German newspaper corpus. In *Proceedings Second International Conference on Language Resources and Evaluation LREC-2000*.

Jean C. Carletta, Amy Isard, Steven Isard, Jacqueline C. Kowtko, Gwyneth Doherty-Sneddon, and Anne H. Anderson. 1997. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31.

John M. Conroy and Dianne P. O'Leary. 2001. Text summarization via hidden markov models. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '01, pages 406–407, New York, NY, USA. ACM.

Michael G. Dyer. 1983. *In-depth understanding: A computer model of integrated processing for narrative comprehension*. Artificial Intelligence Series. MIT Press, Cambridge MA.

Herbert Paul Grice. [1968] 1989. Logic and conversation. In *Studies in the Way of Words*, pages 22–40. Harvard University Press, Cambridge MA.

Julan Kupiec, Jan Pedersen, and Francine Chen. 1995. A trainable document summarizer. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 68–73, New York, NY, USA. ACM Press.

Wendy G. Lehnert. 1981. Plot units and narrative summarization. *Cognitive Science*, 4:293–331.

Chin-Yew Lin and Eduard Hovy. 1997. Identifying topics by position. In *Proceedings of the fifth conference on Applied natural language processing*, ANLC '97, pages 283–290. Association for Computational Linguistics, Stroudsburg, PA, USA.

Chin-Yew Lin and Eduard Hovy. 2002. Manual and automatic evaluation of summaries. In *Proceedings of the ACL-02 Workshop on Automatic Summarization, vol. 4*, AS '02, pages 45–51, Stroudsburg, PA, USA. Association for Computational Linguistics.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In Stan Szpakowicz Marie-Francine Moens, editor, *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 74–81, Barcelona, Spain, July. Association for Computational Linguistics.

Benedikt Löwe. 2011. Methodological remarks about comparing formal frameworks for narratives. In Patrick Allo and Giuseppe Primiero, editors, *Third Workshop in the Philosophy of Information, Contactforum van de Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten*, pages 10–28, Brussel. KVAB.

Daniel Marcu, Magdalena Romera, and Estibaliz Amorrortu. 1999. Experiments in constructing a corpus of discourse trees: Problems, annotation choices, issues. In Marilyn A. Walker, editor, *Towards Standards and Tools for Discourse Tagging. Proceedings of the Workshop*, pages 71–78.

Daniel Marcu. 1998. Improving summarization through rhetorical parsing tuning. In Eugene Charniak, editor, *Proceedings of the Sixth Workshop on Very Large Corpora*, Montréal. Université de Montréal.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19:302–330.

George A. Miller. 1995. Wordnet: a lexical database for english. *Commun. ACM*, 38:39–41, November.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the ACL*, pages 311–318, Stroudsburg, PA, USA. Association for Computational Linguistics.

Rebecca Passonneau, Nizar Habash, and Owen Rambow. 2006. Inter-annotator agreement on a multilingual semantic annotation task. In *Proceedings LREC-2006*, pages 1951–1956.

Vladimir Yakovlevich Propp. 1928. *Morfologiya skazki*. Akademija, Leningrad.

David E. Rumelhart. 1980. On evaluating story grammars. *Cognitive Science*, 4:313–316.

Roger C. Schank. 1982. *Dynamic memory: A theory of reminding and learning in computers and people*. Cambridge University Press.

Scott Turner. 1994. *The creative process. A computer model of storytelling*. Lawrence Erlbaum Associates, Hillsdale, NJ.