Objectivity and Reproducibility of Proppian Narrative Annotations

Rens Bod¹, Bernhard Fisseni¹,²,³,⁴, Aadil Kurji¹, Benedikt Löwe¹,³,⁴,⁵

¹ Institute for Logic, Language and Computation, Universiteit van Amsterdam
Postbus 94242, 1090 GE Amsterdam, The Netherlands
{l.w.m.bod, b.loewe}@uva.nl, aadil.kurji@student.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 Fachbereich Mathematik, Universität Hamburg
Bundesstraße 55, 20146 Hamburg, Germany
⁵ Corpus Christi College, University of Cambridge
Cambridge CB2 1RH, United Kingdom

Abstract

A formal narrative representation is a procedure assigning a formal description to a natural language narrative. One of the goals of the computational models of narrative community is to understand this procedure better in order to automatize it. A formal framework fit for automatization should allow for objective and reproducible representations. In this paper, we present empirical work focussing on objectivity and reproducibility of the formal framework by Vladimir Propp (1928). The experiments consider Propp’s formalization of Russian fairy tales and formalizations done by test subjects in the same formal framework; 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 not easy to reproduce.

1. Introduction & Motivation

The formal study of narratives goes back to the Russian structuralist school, paradigmatically represented by Vladimir Propp’s 1928 study *Morphology of the Folktale* (Propp, 1958). Researchers in the field of computational models of narrative have developed the general Proppian methodology into formal and computational frameworks for the analysis, automated understanding and generation of narratives.¹

In recent years, there has been an increased interest in the methodological and conceptual issues involved. The enterprise of representing a narrative by a formal 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 Λ, a class of mathematical structures Λ, and a description of a procedure (called formalization in (Löwe, 2011)) of assigning to each natural language narrative N a structure Σ(N) ∈ Λ. Note that this procedure is not a function in the mathematical sense, but an activity by expert formalizers who follow given guidelines.

In this paper, we 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 Σ(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 automatized system. The existence or non-existence of formal frameworks Λ with property Obj(Λ) is closely related to Assumption O. In (Bod et al., 2011), we described the investigation of Obj(Λ) 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 (e.g., (Marcus et al., 1993; Brants, 2000; Passonneau et al., 2006)), the formalization or annotation of a narrative is at the next level of complexity. At the sentence or discourse level, inter-annotator agreement has been studied (e.g. (Carletta et al., 1997; Marcus et al., 1999)), but no such analysis has been done for the formalization of narratives, not even for the oldest and best-known formal approach, Propp’s *Morphology of the folktale*, first published in 1928.

We focus on this formal framework, not because it is a particularly good candidate for a framework close to the *stable structural core*, but due to its prominent place in the history of formal representations of narratives. In §2., we describe the Proppian formal framework and discuss two empirical studies pertaining to it, referred to as *Propp I* and *Propp II*, performed at the *Universiteit van Amsterdam*; in §3., we discuss the results and future work.

¹Lehnert’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) are some examples; cf. (Lehnert, 1981; Rumelhart, 1980; Schank, 1982; Dyer, 1983; Turner, 1994).
Table 1: The assignment of the *dramatis personae* for the three folktales in Propp I.

### 2. Propp’s formal system

#### 2.1. Overview of Propp

Working with a corpus of 100 Russian folktales from Afanas’ev’s collection *Narodnye Russkie Skazki*, 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, 1958).

Propp identified seven² *dramatis personae* representing the roles characters may play within the tales. They 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, 1958, § 3).

The actions of the *dramatis personae* are described by a set of thirty-one functions described in (Propp, 1958, § 3) by means of examples and more specified subfunctions. These 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 *preliminary functions*: β Absentation; γ Interdiction; δ Violation, ε Reconnaissance, ζ Delivery, η Trickery, θ Complicity. The *preliminary functions* are not fully developed in (Propp, 1958) and are not included in Propp’s own annotation strings. The main functions are: A Villainy, a Lack, B Mediation, C Beginning counteraction, D Departure, E First function of the Donor, F Hero’s reaction, G Provision or receipt of magical agent, H Spatial transference between two kingdoms, I Struggle, J Branding, K Victory, L Liquidation, M Pursuit, N Rescue, O Unrecognized Arrival, P Unfounded Claims, Q Difficult Task, R Solution, S Recognition, T Exposure, U Transfiguration, V Punishment, W Wedding. These functions, instantiated by *subfunctions* marked by superscripts, occur in strict sequential order, i.e., functions have to occur in the folktale in the order they are given in the list above. In the full Proppian system, there are a few specific ways to break strict sequentiality (Propp, 1958, § IX.A): The most important one is that some folktales contain a series of individual tale units, called *moves*. Examples are *retribing*, 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. None of the tales we used had moves (according to Propp’s own annotations), so we did not include this option in our experiment.

#### 2.2. Description of Propp I

Test subjects were trained in the Proppian framework and then asked to annotate four of the folktales formalized in (Propp, 1958). We used the folktales *The Seven Semyons*, *Shabarsha*, *Ivan the Bear’s Son*, and *Ivan the Bear’s Son, 152*; in the following, we refer to these folktales as Semyons, Shabarsha, and Ivanko.³ We chose tales that were available in English translation, and in Propp’s annotation had no moves.

²One of these, the Princess/Princess’s Father, can be split into two with a slightly difficult delineation. In our experiment, we presented the resulting list of eight *dramatis personae*.

³In Propp I, we also used the folktale *The Enchanted Princess*, but it was too long and omitted in Propp II. Due to an
(i.e., retained strict sequential ordering) and used few functions (Ivanko uses eight functions, Shabarsha six). An annotation of 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 the functions with corresponding 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. We split them into two groups: Test subjects 1 to 5 were group 1 (no subfunction marking) and test subjects 6 to 9 were group 2 (subfunction marking). Test subjects were instructed that the experiment would last three hours and received a moderate financial compensation for participation.

The experiment started with a 45-minute 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. A condensed version of the dramatis personae and functions was distributed as a leaflet for use during the annotation.

Results. Propp’s annotation for Shabarsha was $A^0 B^3 C \uparrow H^2 I^3 K^3 \downarrow$: his annotation for Ivanko was $A^5 \uparrow H^3 I^5 K^1 \downarrow$. These consist of the function strings alone and do not include the preliminary functions.\(^5\)

We give the results of the assignments of dramatis personae in Table 1. 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, H, V, and P (see below for a methodological remark).

The annotation strings vary widely and are given in Table 2 (subfunctions are marked by superscripts, with a missing subfunction marked by \(^2\)). Since no two strings are the same, comparison would have to be per function; calculations of statistics per function are not useful because of the variation in the assignment of dramatis personae and the small amount of data. The strings are longer than Propp’s strings (compare an average of 14.2, 13.2, and 12.8 functions with the Propp’s of 6 and 8 for Ivanko and Shabarsha, respectively).

Methodological Conclusion. Four out of nine test subjects reported that the example story from the presentation was considerably simpler than the 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 used the words “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.\(^5\)

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; 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.

2.3. Description of Propp II

The experiment Propp II was a modified version of Propp I, taking the problems discussed in § 2.1 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 dramatis personae; subfunctions were not discussed at all; the example story was from Propp’s own corpus. It should be noted that Propp only recorded the annotation strings, so that his choice of dramatis personae was extrapolated from (Propp, 1958).\(^6\)

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.

The experiment started with a 45-minute 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. 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

\(^5\)It is conceivable that the designator “devil” created a connotation in the original audience 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”.

\(^6\)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 $D$; finally, in Shabarsha, we assigned Shabarsha to $H$ and the Little Devil and the Grandfather jointly to $V$.\(^4\)
text. We analyzed the folktale (Ivan Popyalov, 135) from the Propp corpus. Again the condensed version of the dramatis personae and functions was distributed as a leaflet to for use during the annotation. Test subjects were assigned an assignment of characters to the dramatis personae together with each folktale.

Results. We give the results of the function annotation in Table 3. 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 6 and 8 functions in the original Propp strings for Ivanko and Shabarsha, respectively),

7 and more similar to Propp’s original strings, but we still do not have matching strings among the test subjects.

It is again not possible to do a serious statistical analysis on the basis of six annotations; we therefore do a qualitative analysis instead. We say that a function occurs stably in Propp II if it is in at least four of the six annotations. We further distinguish strong stability when the marked text of the annotators overlaps, and weak otherwise. In Ivanko, β, I and ↓ were strongly stable and ↑ and H 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).

3. Discussion & Future Work

The difference 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: 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. Making these vague descriptions understandable for the test subjects may require considerably more time and training than we gave the test subjects in our experiment.

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.
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 second, and fourth 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 Ekaterina Abramova and Sanchit Saraf (Amsterdam) for their work in the early set-up phase of the research.

4. References


