Narrative summarization and its correspondence to Proppian functions

Alexander Block[1,5], **Bernhard Fisseni**[1,2,3,5], **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@student.uva.nl, 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

Some formal systems of narrative aim to represent those features that belong to faithful summaries of the narrative. In this paper, we compare data obtained in two experiments by means of an *event mapping*: summaries by untrained test subjects and annotations of test subjects trained in the narrative framework of Vladimir Propp. The method used is adequate for the task and highlights differences between summaries and our Proppian data.

1. Introduction

This paper is part of the larger project of determining adequate formal models of narrative for the *natural human notions of story representation*.

In this paper, we compare the narrative components included in a summary by untrained human readers with the narrative components corresponding to formal annotations based on a given formal framework (Propp, 1958): one reason for this is that the linear annotation strings of Propp's system (with each symbol standing for a narrateme that can be expressed by a natural language sentence) could be seen as summaries in their own right; the other reason is that due to our work on inter-annotator agreement of the Proppian annotation system discussed in the sibling paper (Bod et al., 2012), we had direct access to data on the use of text passages for Proppian annotations.

In § 2., we give an overview of the theoretical background of formal frameworks and summarization. In §§ 3. and 4., we describe the two experiments and the comparison metric we are using (*event mapping*), and give the results in § 5. Finally, in § 6., we discuss our results and comment on possible future work.

2. Background

Formal frameworks. A formal framework Λ for representing narratives consists of a formal language $\mathscr{L}_\Lambda$, a class of mathematical structures $\mathscr{M}_\Lambda$, and a description of a procedure (called *formalization* in (Löwe, 2011)) of assigning to each natural language narrative N a structure $\Sigma_\Lambda(N) \in \mathscr{M}_\Lambda$. We are aiming to design formal frameworks Λ that match well with human judgments of *story equivalence* (i.e., human judges consider narratives N and N^* *the same story* if and only if $\Sigma_\Lambda(N)$ and $\Sigma_\Lambda(N^*)$ are isomorphic) or the cognitive representation of narratives in the human mind (i.e., a feature is relevant for the mental representation of a narrative in human memory if an only if it is part of Λ).

Some formal approaches to narrative explicitly highlight this latter aim of designing a formal representation that matches the mental representation in human minds used for storing and retelling the narrative (cf. (Gentner, 1983; Falkenhainer et al., 1989)). As a motivation for her *Plot Units*, Lehnert connects them to the cognitive representation of memories in the mind of the reader:

> When a person reads a narrative story, an internal representation of that story is constructed in memory. ... [V]ast amounts of information within the memory representation are selectively ignored, in order to produce a distilled version of the original narrative. (Lehnert, 1981, pp. 293–294)

Summarization. A summary of a text is an account "containing ... the chief points or the ... substance of the matter" (OED 1989, *summary, adj.*), i.e., a "distilled version of the original" in Lehnert's words. What the *chief points* or the *substance of the matter* are depends on the task and the context, and thus writing summaries is not a natural but a trained skill. This skill features prominently in reading and

writing education: in different education cultures, we find very different definitions of what constitutes a good summary.[1]

Formal annotations and natural language summaries can be compared on many different levels. Since the formal frameworks we are aiming at are considering narratives as sequences of events and their relations,[2] we aimed at a comparison in terms of events represented by the formal framework and the summary, respectively. Our notion of events is the same as (von Stutterheim, 2004, 328f) used, which regroups all 'dynamic' aspects of (van Valin and LaPolla, 1997) and hence *accomplishment*, *achievement* and *activity* (Vendler, 1957). In order to evaluate the results of our experiments, we created an *event mapping* (cf. § 4.2.).

The relationship between formal representations of natural language text and summarization is of course particularly relevant for the field of *automatic text summarization* (e.g., (Alterman, 1991; Mani and Maybury, 1997; Hahn and Harman, 2002; Moens and Szpakowicz, 2004)). Marcu's work using the Rhetorical Structure Theory (RST) (Marcu, 1998) is more closely related to our approach: RST proposes the assignment of different levels of importance among parts of the text and the possible relations between them. There are also various evaluation metrics for the quality of automatic summaries (e.g., (Lin and Hovy, 2002; Lin, 2004; Papineni et al., 2002)), but they are not directly applicable to our case.

3. Experiment *Propp*

Propp's formal system. 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 Proppian system, the events and actions of the narrative are categorized using a set of thirty-one *functions*, described in (Propp, 1958, § 3); examples are: **a** Lack, **B** Mediation, **C** Beginning counteraction, ↑ Departure, **G** Spatial transference between two kingdoms, **H** Struggle, **I** Victory, **K** Liquidation, **M** Difficult Task, **N** Solution, **U** Punishment, **W** Wedding. For more details, we refer the reader to (Propp, 1958) and the our short account in (Bod et al., 2012).

The experiment. The experiment **Propp** is described in detail in (Bod et al., 2012, § 2.3) where it is denoted by **Propp II**. We used the folktales *The Seven Semyons, 147*, *Shabarsha, 151*, and *Ivan the Bear's Son, 152* from Afanas'ev's corpus; in the following, we refer to these folktales as *Semyons*, *Shabarsha*, and *Ivanko*.[3] Six test subjects, all students of the *Universiteit van Amsterdam*, and all with native or near-native competence of English, read the three folktales and produced a list of the Proppian functions occurring in the folktale, and marked text passages corresponding to each of the functions that occurred.

4. Experiment *Summarization*

4.1. Setup.

The experiment was conducted with six students of the *Universität Hamburg*; they all had native or near-native competence of German. In order to be able to compare the results with the Proppian annotations from (Bod et al., 2012), we used German translations of the Russian folktales used in the experiment **Propp** (cf. § 3.).

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 *recognizable* from the summary, 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. To facilitate the mapping of sentences to events, test subjects were instructed to use "simple sentences" (*einfache Sätze*); simple sentences were explained to "normally consist of up to 14 words" and examples were given, which contained at most one level of subordination or coordination. No example of a summary was given.

Test subjects were then given 2½ hours to write the summaries, and were given a modest financial compensation.

4.2. Event mapping

As mentioned above, we compared the summaries and the Proppian annotations in terms of the represented events. The decision to focus on events was confirmed by the fact that the majority of references in the summaries refer to *events* rather than situational descriptions (out 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%).

In order to compare the represented events, we created event mappings both for the text passages marked in the Proppian annotations and the sentences in the summaries. Here, we counted a sentence or fragment of a sentence as an *event description* if it grammatically describes a change of

[1] As an example for quality measures, consider the *Writing Framework for the 2011 National Assessment of Educational Progress* of the National Assessment Governing Board of the U.S. Department of Education (pp. 10–11). Cf. also the discussion of the differences in "holistic encoding" (*holistische Kodierung*) of texts between Germany and the United States in (Bremerich-Vos and Possmayer, 2011, pp. 38–39).

[2] Cf. (Genette, 2007, p. 13): "[D]ans un premier sens – qui est aujourd'hui, dans l'usage commun, le plus évident et le plus central –, *récit* désigne l'énoncé narratif, le discours oral ou écrit qui assume la relation d'un événement ou d'une série d'événements." Also, in text linguistics (cf., e.g., (Heinemann and Viehweger, 1991, 238ff)), *narrative* is used as a label for texts that relate events in temporal order.

[3] 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.

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. In addition to that, we included non-events if they were expressed as a full clause.

Based on this, we created a master list of the events mentioned in at least one summary or one Proppian annotation. Figure 1 shows the event mapping for *Shabarsha* where the columns S1 to S6 correspond to the six test subjects in the **Summarization** experiment and columns P1 to P6 correspond to the six test subjects in **Propp**.

We say that a event occurs *stably* in one of the two experiments if at least four of the six test subjects list it. For the Proppian functions, we say that a function occurs *stably* if at least four of the six annotators list it. In *Ivanko*, the functions ↑, **H**, **I**, and ↓ occur stably; in *Semyons*, the functions **a**, **B**, ↑, **G**, **K**, and **W** occur stably; in *Shabarsha*, the functions ↑, **H**, and **I** occur stably. A given Proppian function can be assigned to different text passages (and even different events) by different annotators: we call a stable function *strongly stable* if there is a text overlap in the assigned text passages of at least four of the six annotators.

5. Results

5.1. Agreement between the summaries

In *Semyons*, test subjects agree least about which details to include in their summary. However, all agree to mention some events before the theft, in particular 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 to obtain the gold is completely omitted in *Ivanko* summaries but mentioned by four for *Shabarsha*.

5.2. Comparison with the *Propp* experiment

Comparison is difficult since summaries regroup events (something not allowed in a Proppian annotation). In the following, we focus on three qualitative examples of differences: storylines omitted from function assignment, *stable events*, some *stable functions* that are not present in the summaries, and an example of a *stable passage* which we find in all Propp annotations but in no summary.

Denseness. Summary descriptions tend to be very dense, making use of certain implicatures (Grice, 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 but implicates that he actually arrives at the pond. Repetitive events are often combined in the summaries: 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 **Summarization**, the majority of test subjects mentions these globally, not as individual events; in **Propp**, test subjects tend to pick out single elements of these and assign functions to them, so that at most two of about 24 events are *stable* in the summaries.

Storylines. Both *Ivanko* and the *Semyons* have a prelude storyline that leads up to the central story line of the competition with the devil (*Ivanko*) and the theft of the princess (*Semyons*). This does not fit Propp's system which essentially models one storyline; consequently, the initial storylines are free from Propp functions, with one exception: Nearly all test subjects mark *Absentation* (six for *Ivanko*, three for *Semyons*) or *Lack* (two for *Ivanko*, one also marked *Absentation*).[4] While these are motivations for the further development in the stories, the summaries only mention the events resulting from them: *Shabarsha*'s plan to earn money by fishing (six) and the order/permission to steal Helena the *Semyons* receive or their journey (three and two, respectively).

Shabarsha and the *Semyons* also have a final commentary by the narrator stepping out of the storyline into the narrative frame. In *Semyons*, only one summary and no Proppian annotation represents this (surreal) third story line. In *Shabarsha*, part of the final commentary relates to the main storyline and is represented by three summaries and three Propp annotators.

Stable Events. A count of *stable* events is given in Table 1. Nearly always (three exceptions), the events that are stable in one group are reported by at most half as many people in the other group. For each of the stories, there is exactly one stable event in common between the Proppian annotations and the summaries: the *Semyons* are rewarded; *Shabarsha* goes to the pond; and *Ivanko* is assigned the task to go to the lake.

Story	Summary	Propp	Common
Semyons	7	5	1
Shabarsha	5	5	1
Ivanko	6	7	1

Table 1: Stable events for the three stories

Stable Functions. In the Semyons, four Propp annotators agree that the Tsar's love for Helena constitutes a *Lack* (*strongly stable function*), while none of the summaries mentions this condition. Similarly so for the *Lack* that affects *Shabarsha*'s master right at the beginning of the story, which is, however, not mentioned in any summary.

Stable Passages. Conversely, near the end of *Ivanko*, all Propp annotators have (a) one function for Ivanko's journey to the lake and (b) at least one labelled event after the competitions and before the transfer of money and labourer to Ivanko's father; none of the summaries mentions these

[4] The correctness of these with respect to Propp's system may be questionable, as the only affect minor characters.

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 }	Overall Challenge			4	5	4								[H]
	Overall Trick					5a	5a	7a						
	Overall Success			5	6	5b	5b	7b						
	Wrestling	Challenge	4					6						
		Trick	5a											
		Success	5b											
	Running	Challenge									[H]	[H]		
		Trick	6a								[H/I]			
		Success	6b								[I]			
	Whistling	Challenge								[M]	[M]	[M]	[H]	
		Trick	8b							[N]		[N]		
		Success	8a											
	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]	[I]	[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 1: Synoptic annotations of *Shabarsha* from **Summarization** and **Propp**. 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.

explicitly, as the first is implicated by his assignment; the second ones are relatively minor details.[5]

6. Discussion & Future Work

In this paper, we compared summaries and annotations according to Propp's formal system. We observed that the events represented in Proppian annotations are not the same as in summaries, and that the differences reflect properties of the formal system.

Methodologically, the event mapping approach was adequate for the task and we expect that it will prove to be a useful tool for further, similar investigations in the "larger project" mentioned at the beginning of the paper. The event mapping approach could be extended in two directions. First, as seen in the discussion of the results, it can be helpful to extend the list of events into a hierarchy to represent repetitive events and storylines. Secondly, relations between events should be considered (cf. the Genette quotation in fn. 2).

Taking into account the results on trainability of Propp formalizations (cf. (Bod et al., 2012)), it will be interesting to apply our methodology to more modern and abstract formalizations.

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 (both Amsterdam) for their work in the early set-up phase of the research.

7. References

Richard Alterman. 1991. Understanding and summarization. *Artificial Intelligence Review*, 5:239–254.

[5] A similar case that illustrates the variety of the **Propp** data is the end of the *Semyons*: three test subjects assign functions (one, two and four, respectively) between the theft of the princess and her delivery to the Tsar, while no summary mentions these details.

Rens Bod, Bernhard Fisseni, Aadil Kurji, and Benedikt Löwe. 2012. Objectivity and reproducibility of proppian narrative annotations. this volume.

Albert Bremerich-Vos and Miriam Possmayer. 2011. Zur Reliabilität eines Modells der Entwicklung von Textkompetenz im Grundschulalter. *Didaktik Deutsch*, 31:30–49.

Brian Falkenhainer, Kenneth Forbus, and Dedre Gentner. 1989. The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 20:1–63.

Gérard Genette. 2007. *Discours du récit*. Seuil, Paris.

Dedre Gentner. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2):155–170.

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

Udo Hahn and Donna Harman, editors. 2002. *Proceedings of the ACL-02 Workshop on Automatic Summarization, vol. 4*, AS '02. Association for Computational Linguistics.

Wolfgang Heinemann and Dieter Viehweger. 1991. *Textlinguistik*. Niemeyer, Tübingen.

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

Chin-Yew Lin and Eduard Hovy. 2002. Manual and automatic evaluation of summaries. In (Hahn and Harman, 2002), pp. 45–51.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In (Moens and Szpakowicz, 2004), pp. 74–81.

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.

Inderjeet Mani and Mark Maybury, editors. 1997. *Intelligent Scalable Text Summarization. Proceedings of a Workshop sponsored by the ACL*. Association for Computational Linguistics.

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.

Marie-Francine Moens and Stan Szpakowicz, editors. 2004. *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, Barcelona, Spain. Association for Computational Linguistics.

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. Association for Computational Linguistics.

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

Vladimir Yakovlevich Propp. 1958. *Morphology of the folktale*. University of Texas Press, Austin, Texas, 2nd edition. English translation of (Propp, 1928).

Robert D. van Valin and Randy J. LaPolla. 1997. *Syntax. Structure, meaning and function*. Cambridge University Press, Cambridge.

Zeno Vendler. 1957. Verbs and times. *Philosophical Review*, 56:143–160.

Christiane von Stutterheim. 2004. Makrostrukturelle Planungsprozesse in Erzählungen. *Zeitschrift für germanistische Linguistik*, 32:325–356.