Tagging Narrative with Propp's Character Functions Using Large Language Models

Pablo Gervás[1,2], Gonzalo Méndez[1,2]

Facultad de Informática, Universidad Complutense de Madrid, Madrid, 28040 Spain
Instituto de Tecnología del Conocimiento, Universidad Complutense de Madrid, Madrid, 28223 Spain

Abstract
The character functions proposed by Vladimir Propp as abstraction of plot structure to understand Russian folk tales have been popular as means of analysing narrative. Several efforts have been carried out to annotate narratives in this way manually or to apply machine learning techniques over texts previously annotated with syntactic and semantic information. The present paper explores the feasibility of annotating directly from the bare text of synopses of the stories by relying on large language models.

Keywords
automated tagging, Propp's Morphology of the Folktale, character functions, large language models

1. Introduction

When trying to understand narrative, a very powerful tool used in the past has been the identification of abstractions of the meaning of the story that describe its plot in a way that is more generic than a summary of the story. A major tool to help in that aim was the Morphology of the Folktale proposed by Russian formalist Vladimir Propp [1]. The set of character functions proposed by Propp to describe the structure of Russian folktales has been subsequently applied to analyse French fairy tales [2], fantasy plays [3] and television science fiction [4]. Research efforts have focused on developing detailed annotation schemes [5, 6] or ontologies [7, 8, 9] that can be used to annotate stories manually. More recent efforts consider the possibility of developing automatic processes of annotation at the level of Proppian functions that operate over a prior layer of syntactic and semantic annotation of the stories [10, 11, 12, 13, 14].

The present paper explores the possibility of relying on the few-shot learning capabilities of large language models to annotate stories with Proppian functions by operating directly over the text of synopses of the stories.

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Litvak (eds.): Proceedings of the Text2Story'24 Workshop, Glasgow (United Kingdom), 24-March-2024.
*Corresponding author.
pgervas@ucm.es (P. Gervás); gmendez@ucm.es (G. Méndez)
http://nil.fdi.ucm.es (P. Gervás); http://nil.fdi.ucm.es (G. Méndez)
0000-0003-4906-9837 (P. Gervás); 0000-0001-7659-1482 (G. Méndez)

Table 1
Character functions proposed by Propp.

Absentation	The villain or a member of the family leaves the home	Branding	The hero is marked or identified
Interdiction	A prohibition is imposed on the hero	Victory	The hero defeats the villain.
Interdiction violation	The interdiction is violated	Liquidation of lack	Villainy or lack resolved
Reconnaissance	The villain attempts to obtain information	Return	The hero returns
Delivery	The villain gains information or an object	Pursuit	The villain chases the hero
Trickery	The villain deceives the hero to gain an advantage	Rescue from pursuit	The hero is saved from pursuit or danger
Complicity	The victim submits to deception.	Unrecognized arrival	The hero arrives unrecognized
Villainy or lack	The villain causes harm or injury or someone lacks something they want	Unfounded claims	A false hero claims hero status
Mediation	Misfortune is made known	Difficult task	The hero is given a difficult task
Counteraction	The hero reacts to the villain's actions	Solution	The task is accomplished or resolved
Departure	The hero leaves home	Recognition	The hero is recognized or acknowledged
1st donor function	The hero is tested by a potential donor	Exposure	The villain's identity or deception is exposed
Hero's reaction	The hero reacts to the test	Transfiguration	The hero undergoes a transformation.
Receipt of agent	The hero acquires a magical agent	Punishment	The villain receives punishment or consequences.
Transfer	The hero is hero directed or taken to new location	Wedding	The hero marries or is rewarded
Struggle	The hero and villain engage in direct conflict.		

2. Previous Work

The work relevant to this paper that needs reviewing is the representation of narrative using Propp's character functions, existing efforts to annotate narrative with them, and solutions for annotating text with large language models.

2.1. Narrative Structure and Propp's Character Functions

Based on a detailed study of 100 Russian folk tales from the anthology by Afanasiev, the Russian formalist Vladimir Propp proposed a set of abstractions of plot-relevant functions played by the characters of a tale, which he called *character functions* [1]. These character functions represent basic contributions to the plot such as going on a journey, having a conflict with another character, being sent on a mission, or being rewarded. The character functions proposed by Propp are shown in Table 1.

Certain character functions are associated with particular types of character, such as the hero or the villain. Propp also postulated seven such types of character: the villain, the donor (provider), the helper, the princess (or sought-for person), the dispatcher, the hero or victim and the false hero.

Early uptake of Propp's character functions as means of annotating a corpus of stories relied on XML [5]. Propp's character functions have been re-represented as ontologies defined in Description Logics [7]. The more relevant effort at annotating with character function information the very stories for which Propp designed his character functions was carried out by Finlayson as part of his PhD thesis [12], and later extended by Yarlott and Finlayson [6]. They propose ProppML, "an annotation scheme designed to capture all the components of a Proppian-style morphological analysis of narratives". Using this scheme, they annotated fifteen Russian folk tales from the original corpus, which amounts to 18,862 words. The corpus was annotated separately by two highly trained annotators, and the results then merged with supervision by a third expert annotator.

2.2. Automated Annotation of Propp Character Functions

Finlayson's efforts at deeply annotating a corpus of Russian folk tales at various levels–syntax, semantics, and narrative structure– made it possible to develop a machine-learning algorithm capable of extracting culturally-relevant plot patterns from sets of folktales. Such an algorithm takes as input a story annotated as an event timeline and it can predict for it abstract narrative structures such as Villainy, Struggle-Victory, and Reward [14]. This an instance of a computational system learning a theory of narrative structure directly from a corpus of data.

2.3. Text Annotation with Large Language Models

The recent advances demonstrated by applying Large Language Models (LLMs) to various Natural Language Processing (NLP) tasks [15] has opened for exploration a number of avenues that appeared to be impossible just a few years ago. The main idea is that a model pre-trained on very large amounts of textual data can be queried in very simple ways to carry out specific NLP tasks and perform surprisingly well either with no additional information (zero-shot learning) or with very simple queues that exemply how the task should be solved inserted in the prompt that poses the query (few-shot learning).

In particular, the task of annotating texts shows that tools such as ChatGPT show very high performance [16]. ChatGPT is an AI language model developed as part of the GPT (Generative Pre-trained Transformer) series of models and specifically designed for generating human-like text responses in conversational contexts. This type of approach has been applied to diverse tasks such as annotating legal text [17] or propagandistic text [18].

Gemini [19] (previously Bard [20]) is a family of generative AI models that lets developers generate content handling both text and images as input. These models can generate text, handle multi-turn conversations and also zero, one, and few-shot tasks. These models have been trained with data existing up to the beginning of 2023. The Gemini Pro 1.0 version is available for use through an open web interface or through a commercial web API. It allows for similar queries as those presented to ChatGPT [21].

3. Automated Tagging of Narrative Synopses with Proppian Character Functions

We want to explore the relative performance of large language models when used to annotate stories with Propp's character functions via zero-shot learning and few-shot learning approaches.

3.1. Zero-Shot Tagging Using ChatGPT

We are going to try to get ChatGPT to annotate the synopses of the tales that Propp includes as examples of annotations in his book. This should allow us to compare the annotation proposed by ChatGPT with the one indicated by Propp himself.

As a first step, we ask ChatGPT if it knows about Propp's functions. This is the reply:

> Vladimir Propp, a Russian folklorist and scholar, identified and analyzed common narrative elements in Russian folktales. One of his most notable contributions is his identification of "morphology of the folktale," which includes what he termed "character functions". Propp identified 31 character functions that frequently appear in folktales. Here's a brief overview of Propp's character functions:

It then provides a list with 31 items, containing 7 character types and 24 character functions, some of which are Propp's and others are not; there are also some missing Propp's character functions:

- Character types: hero, villain, donor, helper, princess/prize, false hero, dispatcher
- Propp's character functions: interdiction, violation, reconnaissance, delivery, trickery, complicity, mediation, counteraction, departure, first function of the donor, hero's reaction, receipt of a magical agent, struggle, branding, victory, return, pursuit, rescue, unrecognized arrival
- Other functions: helper's object, magical agent, villain's lack, guidance, liquidation (equivalent to victory)
- Missing Propp's functions: absentation, villainy or lack, transfer, resolution of trigger, unfounded claims, difficult task, solution, recognition, exposure, transfiguration, punishment, wedding

It is interesting to note that ChatGPT, although knowledgeable in general in a surprising way, mixes up the types of character identified by Propp with the character functions, gets the number right, but misses out a number of functions because it includes the types of characters instead of those functions.

To test the ability of the system to annotate a story with the characters functions, the system is given the following prompt:

```
Can you annotate this tale using Propp's character functions?
<story>
```

In response, the system outputs a text that describes its version of the annotation. For the following example (tale 104 as described in Propp's book):

> A girl called Vasilisa has a miraculous doll. The tsar does not have a wife. Vasilisa goes to town. Vasilisa settles in an old woman's house. The old woman buys Vasilisa some flax. Vasilisa spins unusually fine yarn from the flax. The doll makes a loom for Vasilisa in one night . Vasilisa weaves unusually fine linen on the loom. The old woman takes the linen to the tsar. The tsar orders that shirts be sewn by the one who has spun and woven the linen. The girl sews the shirts from the linen. The tsar sends for Vasilisa. Vasilisa marries the tsar.

The assignment of character functions that Propp suggests is:

> A girl called Vasilisa has a miraculous doll (*Acquisition of Magical Agent*). The tsar does not have a wife (*Lack*). Vasilisa goes to town. Vasilisa settles in an old woman's house (*Unrecognised Arrival*). The old woman buys Vasilisa some flax. Vasilisa spins unusually fine yarn from the flax. The doll makes a loom for Vasilisa in one night (*Acquisition of Magical Agent*). Vasilisa weaves unusually fine linen on the loom. The old woman takes the linen to the tsar. The tsar orders that shirts be sewn by the one who has spun and woven the linen (*Difficult Task*). The girl sews the shirts from the linen (*Task Resolved*). The tsar sends for Vasilisa. Vasilisa marries the tsar (*Wedding*).

In contrast, the annotation that ChatGPT suggests as a zero-shot solution is:

```
Lack: The tsar lacks a wife.
Dispatcher: Vasilisa goes to town.
Helper: Vasilisa settles in the old woman's house and receives aid from the miraculous doll.
Task: The old woman buys flax for Vasilisa, who spins unusually fine yarn and weaves linen.
Delivery: The old woman takes the linen to the tsar.
Task: The tsar orders shirts to be sewn from the linen spun and woven by the same person.
Performance of task: Vasilisa sews the shirts.
Arrival: The tsar sends for Vasilisa.
Final union: Vasilisa marries the tsar.
```

Zero-shot learning presents the difficulty of having the system provide the result in a format of its own devising, which may differ significantly from the one being used as reference.

3.2. Zero-Shot Tagging Using Gemini

Similarly to what we did using ChatGPT, we prompted Gemini for what it knew about Propp's character types and character functions. It provided a list of the seven character types and then another with 31 character functions:[1] initial situation, absentation, interdiction, violation, reconnaissance, delivery of information, trickery, complicity, mediation, beginning of the counteraction, hero's departure, first function of the donor, tests of the hero, acquisition of magic agent, hero's reaction, spatial translocation, struggle, branding, victory, liquidation of the lack, return, pursuit, rescue, unrecognized arrival, false hero's claims. recognition, reward, difficult task, exposure of the false hero, transfiguration, wedding. punishment, and resolution.

This response does in essence match Propp's set of character functions, but it presents an issue that adds complexity to the task. The system is sometimes not using the exact names that

[1] Gemini listed the character functions grouped by sphere of action, following Propp's description of them.

Table 2
Mapping between Propp's character function names and Bard's paraphrases.

Propp's name	Bard's paraphrase
Delivery	Delivery of information
Transfer	Spatial translocation
Exposure	Exposure of false hero
Solution	Resolution

Propp assigned to his character functions, but rather paraphrasing some of them. This happens for instance for the ones shown in Table 2.

Zero-shot learning approaches to this task will likely suffer from having to identify an appropriate mapping between the set of tags used by the system and the set of tags being used as reference. This presents a particular problem because it makes it very difficult to carry out automated processing of the tagging.

To test the ability of the system to annotate a story with the characters functions, the system was given the same prompt as in the previous case. For the same example considered above (tale 104 as described in Propp's book) the annotation that Gemini suggests as a zero-shot solution is:

```
Characters:
  Hero: Vasilisa
  Dispatcher: None explicitly identified (opportunity arises naturally)
  Donor: Doll (magical helper)
  Prize/Princess: Tsar (implicitly)
  False Hero: None explicitly identified
Narrative Functions:
  Sphere of Action 1: Preparation
    Initial Situation: Vasilisa has a miraculous doll and lives without a father figure.
    Lack: Tsar lacks a wife. (Note: not directly affecting Vasilisa, but creates initial context)
  Sphere of Action 2: Departure
    Beginning of the Counteraction: Vasilisa goes to town (implicitly seeking opportunity).
  Sphere of Action 3: The Hero's Actions
    Hero's Departure: Vasilisa leaves her home environment.
    First Function of the Donor: Old woman provides flax (indirect aid).
    Tests of the Hero: Task of spinning fine yarn (challenge).
    Acquisition of Magic Agent: Doll magically creates a loom for Vasilisa.
    Hero's Reaction: Uses the loom to weave fine linen.
    Spatial Translocation: Linen reaches the Tsar.
  Sphere of Action 4: Return
    Return: Tsar acknowledges the linen and initiates contact with Vasilisa.
    Return: Vasilisa arrives at the Tsar's court.
  Sphere of Action 5: Recognition and Reward
    Unrecognized Arrival: Vasilisa implicitly demonstrates her skills.
    Difficult Task: Sewing the shirts (implicit test of skill).
    Exposure of the False Hero: N/A
    Transfiguration: Vasilisa gains new status as Tsar's wife.
    Wedding: Marriage to the Tsar (reward).
  Sphere of Action 6: Resolution
    Resolution: Lack of the Tsar is addressed (gains a wife), and Vasilisa finds a new life.
```

It is interesting to note that, in the zero-shot learning, Gemini decides to provide information about how the content of the story relates to the Spheres of Action that it considers Propp to have defined.

It is also identifying a number of inferences of implicit content that is not mentioned in the story but is indeed inferable from the text. Examples of this are the fact that Vasilisa has left home (to go to town) or that Vasilisa gains new status as a result of having married the Tsar.

Another relevant detail is that Gemini assigns 11 character functions to this story, whereas Propp only assigned 7. Propp himself had assigned to this tale only 7 character functions, though not the same as ChatGPT.

3.3. Few-Shot Tagging Using ChatGPT

Finally, we wanted to test the performance of the system at a few-shot learning task. So we replicated the experiment above with an enriched prompt that explicitly included the list of character functions and their definitions.

```
I need to annotate some tales using Propp's character types and narrative functions.
These are the tags and definitions I need to use:

For the character types:
+ Hero: The role model of a story.
+ Villain: The negative character who creates struggles for the hero.
+ Donor: The character who provides some magical object to the hero.
+ Helper: The character who helps the hero.
+ Princess: The character who becomes a companion of the hero.
+ Dispatcher: The character who illustrates the need for the hero's quest and sends the hero off.
+ False Hero: The character who takes credit for the hero's actions
For the narrative functions:
+ Absentation: The villain or a member of the family leaves the home
+ Branding: The hero is marked or identified
+ Interdiction: A prohibition is imposed on the hero
+ Victory: The hero defeats the villain
+ Interdiction violation: The interdiction is violated
+ Resolution of trigger: Villainy or lack resolved
+ Reconnaissance: The villain attempts to obtain information
+ Return: The hero returns
+ Delivery: The villain gains information or an object
+ Pursuit: The villain chases the hero
+ Trickery: The villain deceives the hero to gain an advantage
+ Rescue from pursuit: The hero is saved from pursuit or danger
+ Complicity: The victim submits to deception
+ Unrecognized arrival: The hero arrives unrecognized
+ Villainy or lack: The villain causes harm or injury or someone lacks something they want
+ Unfounded claims: A false hero claims hero status
+ Mediation: Misfortune is made known
+ Difficult task: The hero is given a difficult task
+ Counteraction: The hero reacts to the villain's actions
+ Solution: The task is accomplished or resolved
+ Departure: The hero leaves home
+ Recognition: The hero is recognized or acknowledged
+ 1st donor function: The hero is tested by a potential donor
+ Exposure: The villain's identity or deception is exposed
+ Hero's reaction: The hero reacts to the test
+ Transfiguration: The hero undergoes a transformation
+ Receipt of agent: The hero acquires a magical agent
+ Punishment: The villain receives punishment or consequences
+ Transfer: The hero is  hero directed or taken to new location
+ Wedding: The hero marries or is rewarded
+ Struggle: The hero and villain engage in direct conflict

Annotate this tale: <story>
```

We tested this with ChatGPT and for the previous example (tale 104 as described in Propp's book), the annotation produced is:

```
Character Types:
  Hero: Vasilisa
  Villain: None explicitly mentioned
  Donor: The miraculous doll
  Helper: The old woman
  Princess: Vasilisa herself, who becomes the tsar's wife
  Dispatcher: The tsar
  False Hero: None explicitly mentioned
Narrative Functions:
  Dispatcher: The tsar sends for Vasilisa.
  1st donor function: The miraculous doll aids Vasilisa in spinning and weaving.
  Receipt of agent: The doll makes a loom for Vasilisa.
  Delivery: Vasilisa presents the fine linen to the tsar.
  Difficult task: The tsar orders shirts to be sewn from the fine linen.
  Solution: Vasilisa sews the shirts.
  Wedding: Vasilisa marries the tsar.
```

An important point to note here is that the system response does not actually annotate the text in the sense of respecting the text and assigning labels to specific segments of it. Instead it is declaring what labels might be applicable to the text in some sense, and for each one it provides a description of the events in the text that justify the assignment. For instance, the *1st donor function* label is justified with the sentence "The miraculous doll aids Vasilisa in spinning and weaving", which can be understood as a condensed description of the various tasks that Vasilisa carries out in the story, taken in the context that we have been told that the magical doll is helping her. The summary is actually incorrect, or maybe just not altogether precise, as the doll actually only provides the magical loom with which Vasilisa weaves, and it does not actually help with the spinning. This presents a problem in how this type of response should be annotated. First, because it is difficult to establish whether the system assignment of the label matches accurately Propp's assignment–the boundaries of the segment of text considered to justify the label are not defined. Second, because the assignment appears to be based on an incorrect interpretation of the facts in the story.

In this particular case, the *1st donor function* label is incorrectly assigned, because Propp defines the first action of the donor to be testing of the hero by the donor to decide whether he is worthy of gift or assistance. The system has also made an incorrect inference here. A similar case can be observed for the inference supporting the *Delivery* tag: the system justifies it by saying Vasilisa delivers the linen to the tsar, when the story says the old woman does.

A more refined option might be to include in the prompt an example of how the annotations should be presented in combination with the input.

3.4. Few-Shot Tagging Using Gemini

We tested the same prompt-enriched with the explicit list of character functions to use–with GeminiPro and for the previous example (tale 104 as described in Propp's book), the annotation produced is:

```
Annotating the Tale of Vasilisa using Propp's Framework
Characters:
```

```
Hero: Vasilisa
Donor: The Doll (magical helper)
Dispatcher (indirect): The Tsar's lack of wife creates an opportunity for Vasilisa.
Narrative Functions:

Absentation: The Tsar lacks a wife (lack).
Hero's possession of a magical agent: Vasilisa has the miraculous doll.
Departure: Vasilisa goes to town (possibly seeking a new life).
Helper: The old woman provides Vasilisa with flax.
Difficult Task (indirect): Vasilisa spins unusually fine yarn, implying exceptional skill.
Receipt of Magical Agent (indirect): The doll creates a loom for Vasilisa, demonstrating its
magical abilities.
Difficult Task: Vasilisa weaves unusually fine linen, again showcasing her skill.
Mediation: The old woman takes the linen to the Tsar, revealing Vasilisa's talent.
Transfer: The Tsar sends for Vasilisa, bringing her closer to her potential reward.
Marriage: Vasilisa marries the Tsar (fulfills the lack and resolves the story).
Notes:

This tale is a simplified version and omits some of the more complex elements of Propp's morphology.
The villain and some other functions (e.g., struggle, victory) are absent.
The focus is on Vasilisa's exceptional skills, aided by the magical doll, which ultimately lead to
her marriage with the Tsar.
```

It is interesting to see that Gemini has some difficulty in identifying the difference between character types and narrative functions as described in the prompt. This results in an annotation that mixes both kinds of labels indistincly.

4. Discussion

Regardless of the impressive nature at first glance of the replies provided by these large language models, it is very easy for them to be incorrect to a surprising degree. This is especially true for queries that require responses involving detailed recall of definitions. The examples reported above of how the different LLMs recall the actual list of Propp's character functions shows both the strengths and the weaknesses of these models.

The very nature of these systems implies that the text of any response is dynamically constructed on demand from a neural representation in an underlying latent space. For this reason, responses are very likely to be paraphrases rather than verbatim transcriptions. This has the advantage of allowing for a certain leeway in the interpretation of definitions. These systems are more likely to identify instances of a particular character function even when they are phrased in a significantly different way from the original definition. However, it also has the disadvantage of relying on potentially divergent understandings of specific function. The set of definitions provided for the character functions when prompted suggests that this is the case at least for ChatGPT. This suggests that the set of tags being used to annotate by ChatGPT in a zero-shot approach to the task is simply incorrect.

To provide some quantitative measure of the degree of divergence, the set of tags assigned to the sentences in the tale is compared with the assignment that Propp provides for it in his book. In appendices II and III, Propp provides examples of how the plots of a number of tales can be described in terms of his set of character functions. The tales so described are referred to only by number, and the assignment of character functions is given over a synopses of the

Table 3
Metrics for performance of different models and prompting strategies on assignment of character functions to tale synopses with respect to Propp's assignment. Metrics for different model + prompting strategy pairs is shown in a different column. For each model + prompting strategy pair, the first pair of rows in the table show: precision and recall against Propp's assignment, the second pair of rows shows percentage of labels that show some linguistically plausible relation to corresponding story content, and percentage of positions identified by Propp as involving a character function that the system offers a label for, and the third pair of rows shows percentage of incorrect assignments of Propp's categories, and percentage of labels assigned not equivalent to any Proppian character function.

	ChatGPT		Gemini	
	zero shot	few-shot	zero shot	few-shot
Precision	0.38	0.37	0.41	0.46
Recall	0.26	0.22	0.31	0.34
Assignment meaningful	70	66	77	78
Response required	33	33	74	72
Incorrect assignments	26	29	16	44
Invented labels	01	03	10	04

tale. These synopses have been used as input to the system. The tales considered correspond to numbers 93, 104, 123, 127, 131, 133, 139, 155, 198, 244 and 247.

The assignments provided of character functions to segments of the synopses required a detailed analysis. The responses by the various models include some kind of label together with a justification of which part of the story the label is assigned to, both labels and descriptions tend to be paraphrases rather than references to specific character function names or specific segments of the text. This makes simple measures like precision and recall against Propp's original assignment difficult to compute and even more difficult to interpret. To circumvent this problem, in all cases where the model suggested a label that could reasonably be considered to refer to the one assigned by Propp it has been considered valid. In the hope of making the analysis more informative, four additional metrics have been computed. First, the number of times that identifiable references to Propp's character functions have been assigned to segments of the story that they do not apply to, which measures correct match between label and text. Second, the number of times that labels have been used that cannot in be related in some way to the character functions defined by Propp, which measures familiarity with the set of character functions. Third, the number of times that labels assigned (whether identifiable as Propp's character functions or not) can be seen to relate to the corresponding story segment by some valid inference, which measures the language understanding ability of the system. Finally, we report a metric to measure the number of times that the system has attempted to assign a label to segments of the text to which Propp had assigned one, which measure the ability of the system to identify plot-relevant events, regardless of whether it can tag them correctly. The results for these set of metrics are shown in Table 3.

The first observation to make is that the first four metrics show relatively low values. This indicates poor performance overall.

It is interesting to note that for several of the metrics the zero-shot approach performs better than the few-shot approach. This happens to ChatGPT for all metrics save the identification

of plot-relevant segments, and to Gemini at least for the number of incorrect assignments (44 from 16). This is consistent with observed behaviour of LLMs, with quality of results dropping as the size of prompts increases.

The fact that ChatGPT zero-shot had a very low recall is not surprising given that it was not really using the full set of character functions. The zero-shot approach using Gemini had both higher precision and higher recall than the other options, which is consistent with the fact that it is not confused about the character functions. Nevertheless, the scores are surprisingly low.

The additional metrics reported provide some insight into the reasons for these low values. We notice very poor performance (33%) for ChatGPt at the identification of plot-relevant segments, and significant percentages of incorrect .assignments of existing labels (26% for ChatGPT and 16 for Gemini%). The number of labels assigned that do not correspond to Propp's set is surprisingly high (10%) for Gemini zero shot given that it supposedly knows the correct set, but it drops significantly (to 4%) when prompted with the set of names. In contrast, ChatGPT reacts badly to the enriched prompt, with percentages for both incorrect assignments and invented labels.

5. Conclusions

The experiments reported suggest that the use of large language models for automated tagging of stories with Propp character function may yield very different results depending on the level of familiarity of the particular model employed with the set of character functions in question. They also suggest that, if the model is not sufficiently familiar with the definitions of the character functions, a few-shot approach in which brief versions of the definitions are provided may not be enough to compensate for this.

Further experiments are needed both in terms of exploration of a broader range of variations in the prompting and in the consideration of different models or versions of the models. Important aspects to consider in this sense are the differences between public web interfaces and publicly available sources for the models, and diffences in model size.

Acknowledgments

This paper has been partially supported by the CANTOR project (PID2019-108927RB-I00) funded by the Spanish Ministry of Science and Innovation.

References

[1] V. Propp, Morphology of the Folktale, University of texas Press, 1968.
[2] E. D. Williams, The Fairy Tales by Madamme d'Aulnoy, Ph.D. thesis, Rice University, Houston, Texas, 1982.
[3] M. A. McGowan, An analysis of the fantasy plays of James M. Barrie utilizing Vladimir Propp's structural model of the fairy tale, Ph.D. thesis, New York University, 1984.
[4] D. J. Latourette, Doctor Who meets Vladimir Propp: A comparative narrative analysis of myth/folktale and the television science fiction genre, Ph.D. thesis, Northwestern University, 1990.

[5] S. A. Malec, Proppian structural analysis and xml modeling, Proc. of Computers, Literature and Philology (CLiP 2001) (2001).
[6] W. V. H. Yarlott, M. A. Finlayson, Proppml: A complete annotation scheme for proppian morphologies, in: 7th Workshop on Computational Models of Narrative (CMN 2016), Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.
[7] F. Peinado, P. Gervás, B. Díaz-Agudo, A description logic ontology for fairy tale generation, in: Procs. of the Workshop on Language Resources for Linguistic Creativity, LREC, volume 4, 2004, pp. 56–61.
[8] T. Declerck, A. Aman, M. Banzer, D. Macháček, L. Schäfer, N. Skachkova, Multilingual ontologies for the representation and processing of folktales, Proceedings of the LT4DHCSEE in conjunction with RANLP (2017) 20–23.
[9] F. Pannach, An ontology-driven information system based on Vladimir Propp's morphology of the folktale for southern African folktales, Ph.D. thesis, Master's thesis, University of Göttingen, Germany, 2019.
[10] M. A. Finlayson, Deriving narrative morphologies via analogical story merging, New Frontiers in Analogy Research (2009) 127–136.
[11] M. Despontin, L. Sbattella, R. Tedesco, Natural language processing for storytelling and role playing: a training system based on the propp model, in: ICERI2010 Proceedings, IATED, 2010, pp. 5036–5045.
[12] M. M. A. Finlayson, Learning narrative structure from annotated folktales, Ph.D. thesis, Massachusetts Institute of Technology, 2012.
[13] M. A. Finlayson, Inferring propp's functions from semantically annotated text, Journal of American Folklore 129 (2016) 55–77.
[14] M. A. Finlayson, Propplearner: Deeply annotating a corpus of russian folktales to enable the machine learning of a russian formalist theory, Digital Scholarship in the Humanities 32 (2017) 284–300.
[15] B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz, E. Agirre, I. Heintz, D. Roth, Recent advances in natural language processing via large pre-trained language models: A survey, ACM Computing Surveys 56 (2023) 1–40.
[16] M. Alizadeh, M. Kubli, Z. Samei, S. Dehghani, J. D. Bermeo, M. Korobeynikova, F. Gilardi, Open-source large language models outperform crowd workers and approach chatgpt in text-annotation tasks, arXiv preprint arXiv:2307.02179 (2023).
[17] J. Savelka, K. D. Ashley, The unreasonable effectiveness of large language models in zero-shot semantic annotation of legal texts, Frontiers in Artificial Intelligence 6 (2023).
[18] M. Hasanain, F. Ahmed, F. Alam, Large language models for propaganda span annotation, arXiv preprint arXiv:2311.09812 (2023).
[19] G. Team, Gemini: A family of highly capable multimodal models, 2023. arXiv:2312.11805.
[20] J. Manyika, S. Hsiao, An overview of bard: an early experiment with generative ai, AI. Google Static Documents 2 (2023).
[21] S. Mandvikar, Factors to consider when selecting a large language model: A comparative analysis, International Journal of Intelligent Automation and Computing 6 (2023) 37–40.