

Motives and Characters in Folklore Indices and Russian Folktales

Anna Rafeeva

Moscow State University, Moscow, Russia

anna_raf@rambler.ru

ABSTRACT

In this paper, some experiments on automated search for motives and folktale characters will be discussed. First experiments were carried out in 1997-2002, and simple tools for automatic search for some motives and characters in the text of famous Aarne-Thompson index [1] were developed. The second experiment (2002) was an attempt to automatically find one motif in number of folklore indices in Russian. Since 2009 the program SKAZKA-2 (FOLKTALE-2) is developing by author. SKAZKA-2 now contains a corpus of Russian folktales and their descriptions, dictionaries and some tools for automatic and automated proceeding of folktales. Some preliminary results of the work will also be discussed.

1. SKAZKA (FOLKTALE) DATABASE

1.1 Description of Program Tool

SKAZKA (FOLKTALE) program tool was developed by author to deal with text of Aarne-Thompson index [1]. For example, the tasks were to find tale types, where *stepdaughter* is heroine, or to count all the tale types, which are present in Russian folklore. The database contains a small part of the index (some types of tales of magic only), so it is working model of possible system.

SKAZKA was made in Starling, powerful linguistic database software, developed by S. A. Starostin (see <http://starling.rinet.ru/program.php?lan=en> for details). Starling has a lot of useful possibilities, among them are:

- Build-in dictionaries and morphological processor (Russian and English).
- Capability of creating and processing powerful databases specifically suited for linguistic purposes.
- Capability to deal with text fields of variable length.
- Build-in program interpreter for simple user-defined programs.

1.2 Archetypical Motives

There is a variety of definitions for motif now. The aim of the first experiment was to search for *archetypical motives (AM)* in the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

First International AMICUS Workshop, October 21, 2010, Vienna, Austria.
Copyright 2010 AMICUS project at <http://amicus.uvt.nl>.

text of [1]. The term *archetypical motif (or plot archetypes)* was suggested by E.M. Meletinsky in [9], examples are *rescue from monster* and *unpromising hero*. Each AM has variants, for example, *unpromising hero* (or heroine) may be youngest son, orphan or stepdaughter. Hero also may be (or become by some reasons) dirty, covered with ash, speechless and so on.

According to Meletinsky, archetypical motives can be found in different folklore and literature genres, from myth to novel. Archetypical motives are changing and developing from myth to folklore and literature, for example, mythical “totemic wife” transforms into “animal bride” (cat, mouse, frog and so on, see type 402 in [1]) in tales of magic. Plots of some myths can be treated as realization of only one AM (for example, *creation*), narrative structure of folktales is more complicated, and *creation* (in the form of *getting or stealing magic objects*) is only one motif among others.

Full description of archetypical motives is impossible or very difficult. We can, however, make partial description of AM system in some genres. Narrative structure of magic folktales can be described by Propp’s formula [12], so we supposed, that archetypical motives could also be described and formalized. So, the formal system of AM in tales of magic was developed to describe narrative structure of folktales.

In this system, each archetypical motif can be treated as predicate with a fixed number of terms (characters, action, condition and object). Every term can be word or expression of natural language (for example, *prince* and *bear* (characters), *fight* (action) in index, *Prince Ivan* (role: hero) in folktale) or predicate itself. So, folktales can be described at abstract level or more concrete when terms are defined or not. The system is described in [13].

So, on the most abstract level archetypical motives are similar to Propp’s functions [12].

1.3 Experiment

The first task was to find tale types with given archetypical motives in [1]. The sets of motives, provided by S. Thompson, couldn’t be used directly, but motif definitions were very useful and were treated as text, such as descriptions of tale types. So, search rules were derived for some archetypical motives:

- First, keywords were extracted manually for each archetypical motif or some concrete forms, if possible. Sometimes, the Russian-English and English-Russian dictionaries were used to find possible synonymous for given word (for example, *herd*, *shepherd* and *rabbit-herd* in [1]).
- For each keyword, its concordances were made automatically.
- After that, the rules themselves were defined.
- The last step was testing and writing additional rules.

1.4 Parents in Tales of Magic

According to E. Novik [11], relate terms and parents' relations are very important in Russian magic folktales. Besides, automatic search for characters and even their roles (in Propp's terms) can be easier in index and folklore texts, than that of motives or functions. So, the aim of experiment was to extract information about relate terms and roles (in Propp's terms) of relatives from given tale types (text fields of types 300–681 in database). First, programs to make frequency dictionary (for all text fields in database) and concordances for given word or expression was written. The dictionary was used to extract keywords manually, for example *grandmother*, *twin*, *daughter*, *stepchildren*, *niece*, *bride*, *father-in-law*, *godson*, *blood-brothers*, *relatives* and so on. After that, the table of concordances for all keywords was made automatically; it was used for further analysis. The experiment let me draw some interesting conclusions. For example, the roles of *blood-brothers* are similar to the roles of *twins*: they often act together as hero and helper, while elder brothers of hero are false heroes. More results are described in [14].

2. PROCEEDING FOLKLORE INDICES AS TEXT

The aim of next experiment was to compare indices and to find automatically single motif in descriptions of plots and characters. The motif "yearning after lost bridegroom or husband" (that causes appearance of monster or demon) was chosen, because it was present in tales and legends, so the indices of different genres [3–7] could be compared. All indices are in Russian; they describe East Slavic folktales and Russian and Lithuanian legends. Keywords for search were extracted manually with the use of frequency dictionary for the texts of all indices, which was made automatically. The list of keywords contained nouns, verbs and expressions. The most important results of the work seem to be the following:

- The language of indices is simple enough, so a single list of keywords can be used for automatic search in different indices. Thus, some keywords are special for only one index. The full description of keyword extraction and more detailed conclusions are given in [15].
- No one of indices [3–7] describes motives, so this information is not given directly. Automatic proceeding lets us extract this information from plot descriptions.
- For scholar, the structure of index is very important, while automatic proceeding needs large and detail descriptions. For example, descriptions in [4] are short, so it is difficult to use in for automated motif identification without any additional data. This data is provided by SKAZKA-2 software.

3. SKAZKA-2 SOFTWARE

SKAZKA-2 (FOLKTALE-2) is developing for automatic and automated analysis of Russian folktales. Now it contains a corpus of about 1000 tales, their descriptions, frequency dictionary, some tools to deal with text, for example, to make frequency dictionaries and concordances for given word or expression and the program to construct semantic networks in automated way. The index of East Slavic tale types [4], additional dictionaries and morphological processor should be added later.

Tales are stored in plain text format (txt); the resulting dictionaries and tables of concordances are in text format too, but can be imported into e-table or even some database format for further work. All programs are written in C++ language.

The existing corpus of folktales and simple program tools, however, can be used to carry out a lot of experiments and to conclude, what program tools have to be developed and what data (dictionaries, tale descriptions, indices or semantic nets) is to be added.

3.1 Simple Tasks

First task, where SKAZKA-2 was used, was study of numbers in Russian folktales from Afanasjev's textbook [2]. SKAZKA-2 was used to select keywords, find all numbers in corpus and then make the table of concordances. The resulting table was imported into MS Excel for further analysis. Some results of this work were very interesting; for example, 13 is happy or neutral number in Russian tales of magic, as it is often treated as $12 + 1$, where the last object is the best one (for example, 12 swan maidens and the last is hero's bride).

An attempt to find motives and Propp's functions in the text of folktales failed, because extracting keywords for search manually was too difficult. For example, there is a strong correlation between motif "Hero of supernatural origin/strength" and appearance of childless couple at the beginning of the tale; but there exist many ways to describe childless couple in text. This result can be explained easily, if we remember, that archetypical motives and Propp's functions are very abstract: if we take word or expression as first level of abstraction, functions and archetypical motives will be third level or even higher. So, it is necessary to find keywords in some automated way (for example with any statistical method) or other elements for automatic search have to be chosen.

3.2 Characters and Semantic Network

Study of folktale characters and their formal definition is interesting by itself, and an index of characters can be used for a number of purposes, for example, for automated motif definition. Now an index of characters in the form of semantic network is under constructions. It connects characters, as they are given in folktale, and their roles in fairy tale (the modified system of roles described in [10] is used). To fix all characters in Russian tales of magic it is necessary to add new roles, additional to Propp's ones. For example, *prince* can be *hero*, *helper* (rarely) or *false hero*, but he can also be *watcher* or *informer*; *cat* can be *helper*, *object of difficult task* or one of *hero's forms*, if hero is magically transformed into cat. Cat can also be one of *witch's helpers* or *servants*. The list of roles is open now and is developing together with network.

To build semantic network in automated way ProSeCa (PROgram for SEmantic Classification) is used. The program is developing to make semantic dictionary of Russian language in the form of directed graph. In this tool, semantic descriptions for every word is performed like chain in digraph, where the beginning vertex is text example (optionally) or word of natural language, and the end vertex is one of user-defined semantic "primitives". ProSeCa is used to store dictionary and to build chains in automated way (see [8]). In next version of ProSeCa not only vertices, but edges too will allow marking to show different relations between characters in Russian folktales.

In SKAZKA-2, semantic network connects text examples, characters, as they appear in text (words or expressions of natural language), abstract descriptions (roles and others) and other characters too. Here some short chains for *bear* are shown. Text examples and their descriptions are omitted.

Mishka (name) – medved' (bear) – wild animal – animal

medved' (bear) – monster (location: forest) – monster

medved' (bear) – animal helper – helper

medvediha (she-bear) – bear – enchantment: transformation into bear

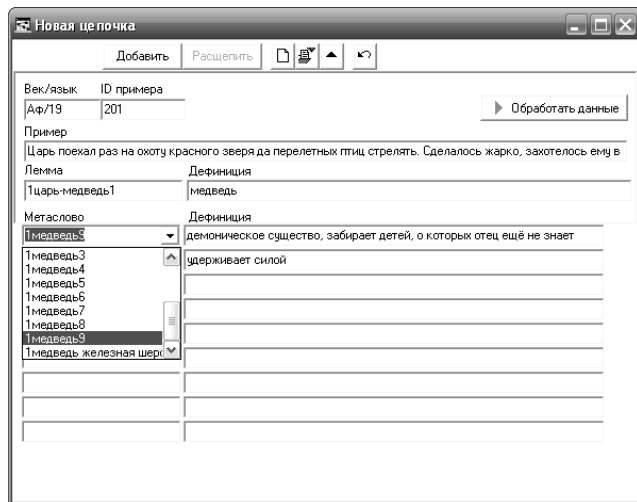


Figure 1. Selecting definition for *tsar-medved'* (bear king)

If given word or semantic description is present in the network, Proseca lets user to select definition for it and then builds the rest of chain automatically (see Figure 1).

Resulting chains can be viewed in text form, where vertices (characters and roles), their definitions in dictionary and examples are placed in different windows (see Figure 2).

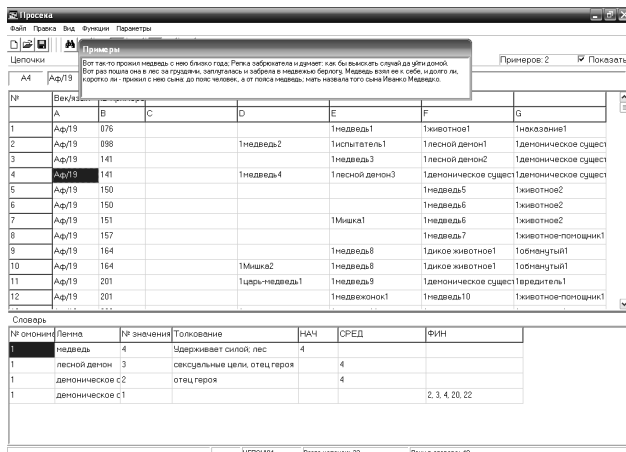


Figure 2. Chains for *bear* with examples and dictionary

Only few characters – *medved'(bear)*, *kot (cat)*, *baba-jaga (witch)*, *tsarevich (prince)* and *lisa (fox)* – are partly described at this time. But even now the network is useful for studies in Russian folktales. Later it will be used not only for automated

proceeding in SKAZKA-2, but to make an index of characters in Russian tales of magic too.

4. REFERENCES

- [1] Aarne, A. and Thompson, S. 1973. *The Types of the Folktale: A Classification and Bibliography*. Anti Aarne's *Verzeichnis der Märchentypen* (FF Communications 3). Translated and Enlarged by Stith Thompson (FF Communications 184). Academia Scientiarum Fennica, Helsinki.
- [2] Afanasjev, A. 1984-1985. *Narodnyje russkie skaski* A.N. Afanasjeva (Russian folk tales from A.N. Afanasjev). Nauka, Moscow. See also: <http://feb-web.ru/feb/skazki/default.asp>
- [3] Ajvazjan, S. and Jakimova O. 1975. *Ukazatel' sujetov russkikh bylichek i byval'chin o mifologicheskikh personagah* (Plot-Index of Legends about Mythical Characters). In: *Pomerantseva E. Mifologicheskije personaji v russkom folklore* (Mythical Characters in Russian Folklore). Moscow. See also: <http://ruthenia.ru/folklore/ayvazan1.htm>
- [4] Barag, L., Berezovsky, I., Kabashnikov, K and Novikov, N. 1979. *Sravnitel'nyj ukf=zetel' luzhetov. Vostochnoslavjanskija skazka* (Comparative Index of Plots. East Slavic Folktales). Nauka, Leningrad. See also: <http://ruthenia.ru/folklore/sus/index.htm>
- [5] Gordeeva, N. 1978-1984. *Ukazatel' sujetov bylichek i byvalshin Omskoj oblasti* (Index of Plots of Legends for Omskij Region). In *Folklore and Postfolklore* (Moscow). DOI= <http://ruthenia.ru/folklore/gordeeva1.htm>
- [6] Kerbelyte, B. 2001. *Tipy narodnyh skazanij: Strukturo-semanticheskaja klassifikacija litovskih etiologicheskikh, mifologicheskikh skazanij i predanij*. The Structural-semantic Classification of Lithuanian Aetiological, Mythological and Historical Legends. Sankt-Peterburg. See also: <http://ruthenia.ru/folklore/kerbelite1.htm>
- [7] Kozlova, N. 2000. *Vostochnoslavjanskije bylichki o zmeje i zmejah. Mificheskij l'ubovnik. Ukazatel' sujetov i teksty*. (East Slavic Legends about serpent. Mythical Lover. Index of Plots and Textes. Omsk. See also: <http://ruthenia.ru/folklore/kozlova1.htm>
- [8] Kretov, A. and Rafeeva, A. 2009. *Programma semanticheskoi klassifikacii leksiki – ProSeKa: teoreticheskije i prikladnye aspekty* (On semantic classification programm ProSeCa: theoretical and practical aspects). In: *Kompjuternaja lingvistika i intellektualnyje tehnologii: Po materialam ejegodnoj Mezhdunarodnoj konferencii "Dialog'2009"*. Russian State University for Humanities, Moscow.
- [9] Meletinsky, E. 1994 *O literaturnyh arhetipah* (About Literary Archetypes). Russian State University for Humanities, Moscow.
- [10] Meletinsky, E., Nekludov, S., Novik, E. and Segal, D. 2001. *Problemy struktornogo opisanija volshebnoj skazki* (Problems of structural description of fairy tale). In: *Structura volshebnoj skazki* (The Structure of Fairy Tale). Russian State University for Humanities, Moscow.

- [11] Novik, E. 2001. Sistema personagej ruskoj volshebnoj skazki (The System of Personages in Russian Fairy Tale). In: *Structura volshebnoj skazki (Structure of Fairy Tale)*. Russian State University for Humanities, Moscow.
- [12] Propp, V. 1969. *Morfologija skazki (The Morphology of Folktales)*. Nauka, Moscow.
- [13] Rafeeva, A. 1998. *Issledovanie semanticheskikh struktur traditsionnykh sujetov i motivov (Research for Semantic Structures of Traditional Plots and Motives)*. Dissertation. Russian State University for Humanities, Moscow.
- [14] Rafeeva, A. 2004. Analiz rodstvennykh otnoshenij s pomosh'ju sistemy SKAZKA (Analysis of Related Terms with SKAZKA Program). In: *Problemy kompjuternoj lingvistiki: Sbornik nauchnykh trudov*. Voronezh State University, Voronezh. See also: <http://www.ruthenia.ru/folklore/rafaeva4.htm>
- [15] Rafeeva, A. 2005. Nekotoryje vozmozhnosti kompjuternogo analiza folklornyx ukazatelej (Some Capabilities of Computer-Based Analysis of Folklore Indices). In: *Problemy kompjuternoj lingvistiki*. Voronezh State University, Voronezh. See also: <http://ruthenia.ru/folklore/rafaeva5.htm>