ProppLearner:
Deeply annotating a corpus of Russian folktales to enable the machine learning of a Russian formalist theory

Mark A. Finlayson

Подготовила:
Ксения Григорьева

Предмет:

   В. Пропп «Морфология волшебной сказки»

   Сказки, разобранные Проппом

Цель:

   Компьютерное изучение теории нарратива русских формалистов

Выбор текстов

15 текстов, 18 862 слова

Тексты на английском языке

Single-move tales

Table 5 Tales in the corpus

Tale number	Russian title	English title	Number of words	Number of events
148	Никита кожемяка	Nikita the Tanner	646	104
113	Гуси-лебеди	The Magic Swan Geese	696	132
145	Семь симеонов	The Seven Simeons	725	121
163	Бухтан Бухтанович	Bukhtan Bukhtanovich	888	150
162	Хрустальная гора	The Crystal Mountain	989	150
151	Шабарша	Sharbarsha the Laborer	1,202	236
152	Иванко Медведко	Ivanko the Bear's Son	1,210	223
149	Змей и цыган	The Serpent and the Gypsy	1,210	250
135	Иван Попялов	Ivan Popyalov	1,228	220
131	Фролка-сидень	Frolka Stay-at-Home	1,388	248
108	Ивашко и ведьма	Ivashko and The Witch	1,448	276
154	Беглый солдат и черт	The Runaway Soldier and the Devil	1,698	317
114	Князь Данила-Говорила	Prince Danila Govorila	1,774	341
127	Купеческая дочь и служанка	The Merchant's Daughter and the Maidservant	1,794	331
140	Зорька, вечорка и полуночка	Dawn, Evening, and Midnight	1,934	339
Average			1,258	229
Sum			18,862	3,438

Аннотирование

- $125,000 USD
- 12 исследователей => 8 команд по

3 человека (2+1)

- 3000 слов за 1-3 недели
- 2 аннотации => 1
- 10 месяцев
- The Story Workbench
- 18 слоев аннотации

Table 6 Teams and the layers for which they were responsible

Team number	Layers
1	Word Senses, Part of Speech Tags, Lemmas, MWEs
2	Referring Expressions, Co-Reference Bundles
3	Time Expressions, Events
4	Semantic Roles
5	Temporal Links
6	Referent Properties, Context Relations
7	Event Valences
8	Dramatis Personae, Functions

О чем пишет Пропп?

«Грамматика сюжета»

- Функции (типы и порядок сюжетов)

- Повествовательная структура (нарратив)

- Зависимость функций друг от друга

- Типы героев

Raw data (необработанные данные)

Surface semantics (поверхностная семантика)

Слои информации

Events and the Timeline

TimeML

1. События (events)
2. Временное выражение (time expressions)
3. Временные связи (time links)

Слои информации

Referential structure

1. Объект -- Субъект

2. Семантические роли

Table 1 Generalized meanings of PropBank frame arguments

Role label	Generalized meaning
ARG0	Subject, agent, or theme
ARG1	Object or patient
ARG2	Instrument
ARG3	Start state or starting point
ARG4	Benefactive, end state, or ending point
ARG5	Direction or attribute
ARGM	Modifying argument, usually augmented with a feature; all verbs may take ARGMs regardless of their frame
ARGA	Agentive argument where the agent is not ARG0; see example

Слои информации

Semantics

Word sense

disambiguation (WSD)

Table 2 Categories of referent attributes and their meaning

Type	Description
Physical	Visible or measurable characteristics such as size, height, and weight
Material	What a referent is made or composed of, or one of its ingredients
Location	Identifying spatial position of a referent, e.g. 'His front teeth'
Personality	Nonphysical character traits of characters
Name/Title	Nicknames, proper names, titles, and other terms of address
Class	Answers the question 'What kind?'
Origin	Whence an object comes, e.g. 'Cockroach milk'
Whole	What the referent is (or was formerly) a part of
Ordinal	Indicates the order or position of the referent in a set
Quantification	Answers the question 'Which one(s)?'
Mass amount	Answers the question 'How much?'
Countable amount	Specific numbers that answer the question 'How many?'
Descriptive	Catch-all for attributes that do not fall into another category

Слои информации

Валентность события (Event valences)

От -3 до 3

Table 3 Event valences and their meaning

Valence	Description	Example
−3	Immediately bad for the hero or his allies	The princess is kidnapped; the hero is banished
−2	May lead directly to a−3 event	The hero and the dragon fight
−1	Someone threatens a −2 or −3 event	The witch threatens death to, or chases, the hero
0	Neither good nor bad	
+1	Someone promises a +2 or +3 event	An old man promises help someday when most needed
+2	May lead directly to a +3 event	Someone hides the hero from pursuit
+3	Immediately good for the hero or his allies	The hero marries the princess; the hero is given gold

Слои информации

Sintax

1. Слова (tokens)
2. Выражения (multi-word expressions)
3. Части речи (part-of-speech tags)
4. Слово в словарной форме (lemmas)
5. Предложения (sentences)

Слои информации

Propp's morphology

Персонажи

Table 4 Propp's 'dramatis personae' and their meanings

Role	Description
Hero	main character of the story
Villain	perpetrator of the villainy; struggles with the Hero
Helper	accompanies and assists the Hero
Donor	prepares and provides the magical agent to the Hero
Princess	sought-for person, not necessarily female
Dispatcher	sends the Hero on his adventure
False Hero	someone who pretends to be the Hero to gain the promised reward

The label 'False Hero' did not occur in the corpus.

Слои информации

Функции

Проблемы:

1. Нечеткое размещение (unclear placement)

2. Неявные функции (implicit functions)

3. Несоответствие разметки (inconsistent marking of trebling)

4. Несоответствие функций, обозначенных Проппом и обнаруженных в тексте

Резюме

+

- Подробное аннотирование каждой сказки
- Структурированность
- Проверка теории Проппа

-

- Отсутствие результатов в статье (это оговаривается)
- Небольшой объем данных
- Большая часть работы сделана вручную
- Вероятность применить такую схему к другому материалу невысока