

<https://doi.org/10.30853/filnauki.2019.6.55>

Москин Николай Дмитриевич

ОПРЕДЕЛЕНИЕ ЗНАЧИМОСТИ ОБЪЕКТОВ ФОЛЬКЛОРНЫХ ПЕСЕН С ПОМОЩЬЮ ТЕОРЕТИКО-ГРАФОВЫХ МОДЕЛЕЙ

В статье рассматривается задача определения значимости объектов в сюжете фольклорной песни при помощи математических методов. На примере бесёдных песен Заонежья XIX - начала XX века показано, как строится теоретико-графовая модель, отражающая семантическую структуру текста, и вычисляются количественные характеристики этой модели. По ним можно судить о локальной и общетекстовой значимости объектов фольклорной песни. Процедура построения теоретико-графовых моделей и подсчета характеристик реализована в информационной системе "Фольклор".

Адрес статьи: www.gramota.net/materials/2/2019/6/55.html

Источник

Филологические науки. Вопросы теории и практики

Тамбов: Грамота, 2019. Том 12. Выпуск 6. С. 251-255. ISSN 1997-2911.

Адрес журнала: www.gramota.net/editions/2.html

Содержание данного номера журнала: www.gramota.net/materials/2/2019/6/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: phil@gramota.net

УДК 81'32; 519.766

Дата поступления рукописи: 07.04.2019

<https://doi.org/10.30853/filnauki.2019.6.55>

В статье рассматривается задача определения значимости объектов в сюжете фольклорной песни при помощи математических методов. На примере бесёдных песен Заонежья XIX – начала XX века показано, как строится теоретико-графовая модель, отражающая семантическую структуру текста, и вычисляются количественные характеристики этой модели. По ним можно судить о локальной и общетекстовой значимости объектов фольклорной песни. Процедура построения теоретико-графовых моделей и подсчета характеристик реализована в информационной системе «Фольклор».

Ключевые слова и фразы: лингвофольклористика; бесёдная песня; теоретико-графовая модель; значимость элементов текста; информационная система «Фольклор».

Москин Николай Дмитриевич, к.т.н., доцент
Петрозаводский государственный университет
moskin@petrsu.ru

ОПРЕДЕЛЕНИЕ ЗНАЧИМОСТИ ОБЪЕКТОВ ФОЛЬКЛОРНЫХ ПЕСЕН С ПОМОЩЬЮ ТЕОРЕТИКО-ГРАФОВЫХ МОДЕЛЕЙ

Лингвофольклористика как филологическая наука начала свое развитие в последней трети XX столетия. При этом в настоящее время можно говорить о трех сложившихся направлениях в изучении языка фольклора [12, с. 11]:

- 1) выяснение природы языка фольклора через его соотношение с диалектами;
- 2) изучение отдельных элементов структуры народно-поэтической речи;
- 3) функционально-стилистическое использование фактов языка в системе народной поэтики.

А. Т. Хроленко отмечает, что «каждый элемент, включаясь в систему, приобретает новое качество, новую значимость, т.е. системные свойства, поэтому логично предположить, что каждое опорное слово фольклорного текста получает дополнительный семантический груз, определяемый структурными свойствами “фольклорной картины мира”» [14, с. 19]. Определение значимости текстовых объектов является **актуальной** задачей в области лингвофольклористики [4, с. 8-10].

Специалисты, которые анализируют фольклорные тексты, все чаще применяют в своих исследованиях математические методы и современные компьютерные технологии. Одним из таких методов является контент-анализ, который сводится к подсчету частот встречаемости в тексте определенных словосочетаний (индикаторов). Например, подобным образом проводилось исследование Н. Э. Шишковой [15]. Другой подход заключается в том, чтобы представить текст в виде типологической формулы, похожей на формулу библиотечной классификации УДК. Ее использовал, например, И. Г. Левин при создании типологии таджикских сказок [10, с. 76]. Однако такие методы, заменяющие текст набором из нескольких чисел или символов (вектором), вряд ли достаточны для отражения его содержания. **Целью** данной работы является разработка новых технологий анализа фольклорных песен для исследования локальной и общетекстовой значимости их объектов. Достижение этой цели достигалось путем решения ряда **задач**, а именно: разработка теоретико-графовой модели сюжета фольклорных песен, определение числовых характеристик, отражающих значимость текстовых объектов, и реализация процедуры автоматизированного подсчета этих характеристик в информационной системе «Фольклор». **Научная новизна** работы заключается в применении математических моделей и методов, которые ранее не использовались для решения подобных задач.

Исследуя семантику произведений фольклора, А. Т. Хроленко отмечает: «Нет сомнения в том, что перед нами очень устойчивая система, состоящая, как и любая система, из определяющих константных элементов и отношений между элементами» [14, с. 18]. Математическим представлением подобной системы является граф, который состоит из множества вершин и множества ребер, соединяющих эти вершины. При этом вершинам и ребрам можно поставить в соответствие дополнительную смысловую информацию (в этом случае следует говорить о теоретико-графовых моделях). В работе «Теоретико-графовые модели фольклорных текстов и методы их анализа» [7, с. 22-29] описывается подобная модель на материале бесёдных песен Заонежья XIX – начала XX века. Бесёдными назывались песни, которые исполнялись в закрытом помещении (избе) во время молодежных вечеринок в Олонецкой губернии (в частности, в Заонежье) в осенне-зимний период [2, с. 18].

Покажем, как строится теоретико-графовая модель на примере фольклорной песни «Во горенке во новой, во новой» (записана в 1841 году) из сборника «Народные песни Вологодской и Олонецкой губерний, собранные Ф. Студитским» [8, с. 69]:

*Во горенке во новой, во новой,
Стоял столик дубовой, дубовой.
Стоит чаша медвяна, медвяна,
Полна меду налита, налита.
Это кто наливал, наливал?
Наливал-то Иванушко,
Подносил-то Марьюшке.*

Ты испей, испей, Марьюшка, душа,
 Поискушай, Федоровна,
 Роди сына во меня, во меня,
 Роди дочь во себя, во себя.
 Сына выучим грамоте,
 Дочь выучим шелком шить,
 Сына женим у попа, у попа,
 Дочку выдаем за дьяка, за дьяка.

В этой песне выделим следующие объекты: *новая горенка*, *дубовый столик*, *чаша медвяна*, *мед*, *Иванушко*, *Марьюшка*, *сын*, *дочь*, *поп*, *дьяк*. Объекты *Иванушко*, *Марьюшка*, *сын*, *дочь*, *поп*, *дьяк* следует отнести к группе «персонажи», *новая горенка*, *дубовый столик* и *чаша медвяна* – к группе «жилище и домашнее имущество», а *мед* – к группе «пища, питье, деньги». Подробное описание групп (лексико-семантических полей) представлено в работе А. Т. Хроленко [13, с. 41-42]. Между объектами устанавливаются локальные связи, которые подкреплены в тексте глаголом или глагольной формой. Каждой локальной связи можно поставить в соответствие ее порядковый номер появления в тексте. В результате получается теоретико-графовая модель сюжета песни, где объектам сопоставляются вершины графа, а дугам (ориентированным ребрам) – локальные связи (Рис. 1).

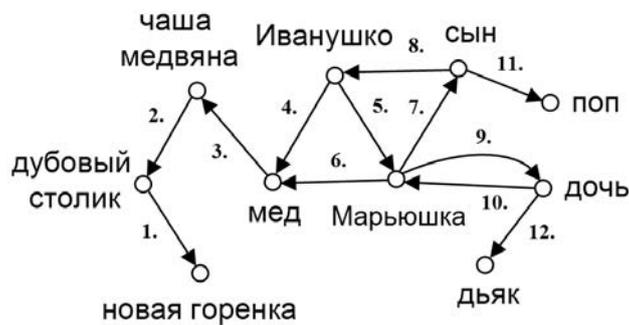


Рисунок 1. Теоретико-графовая модель песни «Во горенке во новой, во новой»

Покажем, как с помощью математических методов можно количественно измерить значимость объекта:

1. Вычисление степеней вершин

Степенью вершины v в графе G (обозначается $\deg v$) называется число дуг, инцидентных v (инцидентность означает, что вершина является началом или концом дуги). Например, для графа песни «Во горенке во новой, во новой» распределение степеней вершин $\{5, 3, 3, 3, 3, 2, 2, 1, 1, 1\}$ изображено на Рис. 2 (числа отсортированы по убыванию).

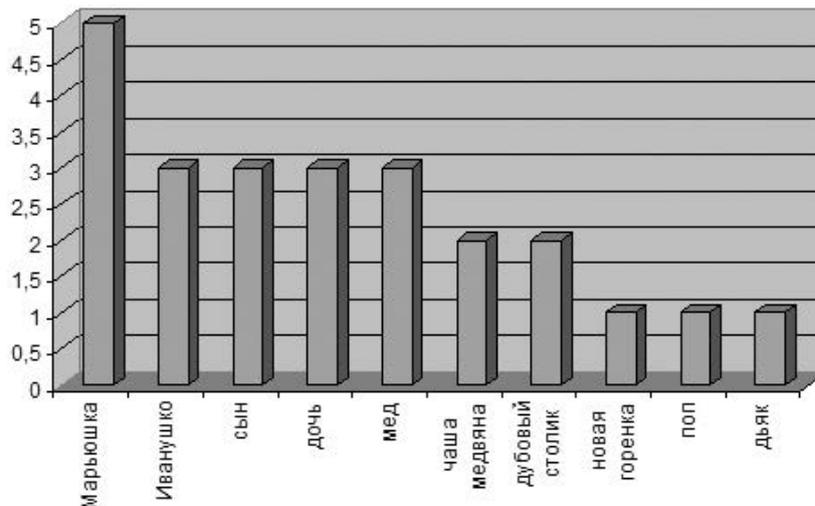


Рисунок 2. Распределение степеней вершин для графа песни «Во горенке во новой, во новой»

Подобное распределение можно интерполировать гиперболической кривой вида $y = \frac{a}{x} + b$, вычислив коэффициенты a и b с помощью метода наименьших квадратов [7, с. 64]. Данные коэффициенты можно использовать для решения задачи разграничения фольклорных текстов и текстов, стилизованных под фольклор (которые сочинили такие известные поэты, как Н. А. Клюев, А. К. Толстой и С. А. Есенин) [16].

В статье «Применение метода дискриминантного анализа для сравнения групп фольклорных песен разных территорий» [3] на примере фольклорных песен Заонежья и деревень Лужского уезда показано, что наиболее значимым фактором, разграничивающим тексты разных регионов, также является распределение их объектов по числу связей.

2. *Вычисление функциональных весов вершин*

Функциональный вес характеризует локальную и общетекстовую значимость соответствующей единицы в данном тексте [11, с. 128]. Он находится по формуле

$$\phi_v = \text{deg } v \cdot (m - m_v^{\max}),$$

где $\text{deg } v$ – степень вершины v , m – число вершин в графе, m_v^{\max} – число вершин в максимальном по величине связном подграфе, полученном после удаления вершины v . Составляющая $\text{deg } v$ представляет собой локальную, а $m - m_v^{\max}$ – общетекстовую значимость.

Например, для графа песни «Во горенке во новой, во новой» распределение функциональных весов вершин {15, 12, 6, 6, 6, 4, 3, 1, 1, 1} изображено на Рис. 3. Наибольший функциональный вес имеет объект *Марьюшка* (15), а наименьший – объекты *новая горенка*, *поп* и *дьяк* (1). Если сравнивать с предыдущей характеристикой, то объект *Иванушко* сдвигается в диаграмме на седьмое место, а объект *мед*, напротив, демонстрирует большую общетекстовую значимость.

Отметим также, что для данного текста распределение функциональных весов коррелирует с распределением степеней соответствующих вершин (коэффициент корреляции Пирсона [1, с. 407] $r \approx 0,87$, что, согласно шкале Чеддока [9, с. 181], считается высокой связью).

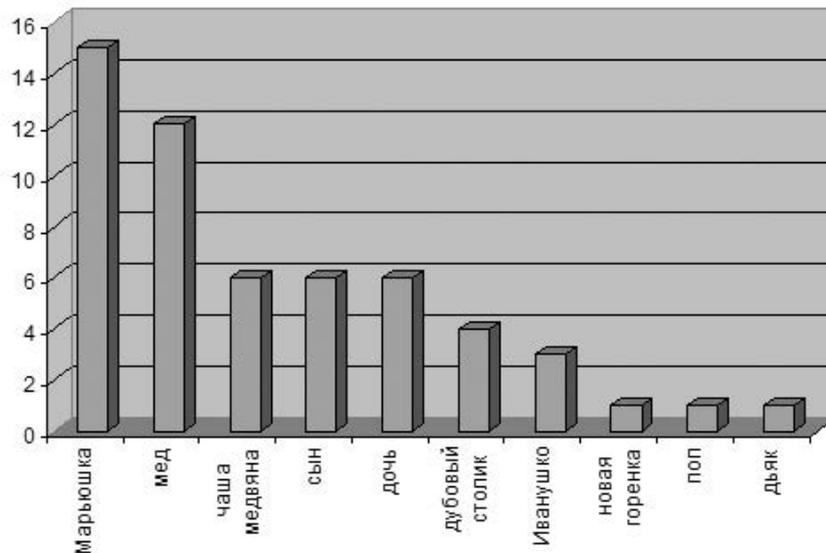


Рисунок 3. Распределение функциональных весов вершин графа песни «Во горенке во новой, во новой»

3. *Вычисление весового индекса, индекса экспансивности и индекса статуса*

Подобные характеристики применяются в социометрии [5] для изучения внутригрупповых и межгрупповых отношений в малых социальных группах. Рассмотрим следующие индексы:

- Индекс веса вершины v . Вычисляется как отношение числа дуг графа, для которых v – конечная вершина, к числу вершин графа без единицы:

$$P_v = \frac{n_v}{m-1},$$

где n_v – число дуг, для которых v – конечная вершина, m – число вершин графа.

- Индекс экспансивности вершины v . Вычисляется как отношение числа дуг графа, для которых v – начальная вершина, к числу вершин графа без единицы:

$$A_v = \frac{n_v^*}{m-1},$$

где n_v^* – число дуг, для которых v – начальная вершина. В Таблице 1 представлены округленные значения P_v и A_v для каждого объекта песни. Видно, что во всех случаях разница этих показателей не превышает 0,11.

Таблица 1. Индексы веса и экспансивности объектов песни «Во горенке во новой, во новой»

	новая горенка	дубовый столик	чаша медвяна	мед	Иванушко	Марьюшка	сын	дочь	поп	дьяк
P_v	0,11	0,11	0,11	0,22	0,11	0,22	0,11	0,11	0,11	0,11
A_v	0	0,11	0,11	0,11	0,22	0,33	0,22	0,22	0	0

• Статус вершины v . Определяется с помощью матрицы смежности M графа G . Статус вершины v равен сумме элементов соответствующей строки матрицы, которая представляет собой нормированную сумму матрицы смежности с удвоенным квадратом этой матрицы:

$$St_v = \sum_{j=1}^m s_{vj},$$

где s_{vj} – элемент матрицы $S = \frac{2}{m(m-1)}(M + 2M^2)$. Например, для графа песни «Во горенке во новой, во новой» распределение статуса вершин изображено на Рис. 4. Заметим, что у персонажа *Иванушко* статус увеличился, если сравнивать с предыдущей диаграммой, а у объекта *мед* – уменьшился.

Для данного текста распределение статусов больше всего коррелирует с распределением степеней соответствующих вершин (коэффициент корреляции Пирсона $r \approx 0,92$) и распределением индексов экспансивности (коэффициент корреляции Пирсона $r \approx 0,95$). Согласно шкале Чеддока, такая связь считается весьма высокой.

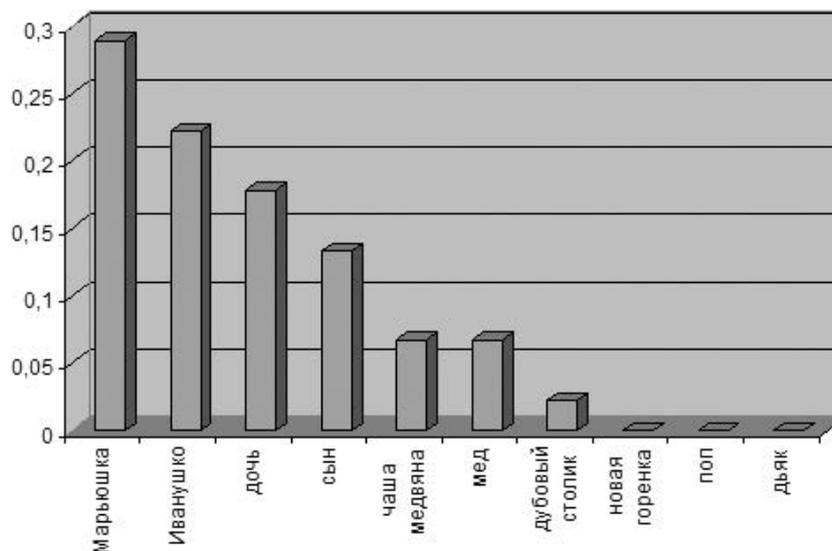


Рисунок 4. Распределение статуса объектов песни «Во горенке во новой, во новой»

В сводной Таблице 2 представлены все вышеперечисленные характеристики (в процентах).

Таблица 2. Числовые характеристики объектов песни «Во горенке во новой, во новой»

№	Объект	Степень вершины $\deg v$	Функциональный вес ϕ_v	Индекс веса P_v	Индекс экспансивности A_v	Индекс статуса St_v
1	новая горенка	4,17%	1,82%	8,33%	0%	0%
2	дубовый столик	8,33%	7,27%	8,33%	8,33%	2,25%
3	чаша медвяна	8,33%	10,91%	8,33%	8,33%	6,85%
4	мед	12,5%	21,82%	16,67%	8,33%	6,85%
5	Иванушко	12,5%	5,45%	8,33%	16,67%	22,7%
6	Марьюшка	20,83%	27,27%	16,67%	25%	29,55%
7	сын	12,5%	10,91%	8,33%	16,67%	13,6%
8	дочь	12,5%	10,91%	8,33%	16,67%	18,2%
9	поп	4,17%	1,82%	8,33%	0%	0%
10	дьяк	4,17%	1,82%	8,33%	0%	0%

Из таблицы видно, что наиболее значимым персонажем фольклорной песни «Во горенке во новой, во новой» по всем показателям является объект *Марьюшка*, а наименее значимыми – объекты *новая горенка*,

поп и *дьяк*. Объект *мед* имеет достаточно большую общетекстовую и локальную значимость, однако уступает объекту *Иванушко* по статусу.

Таким образом, можно сделать следующие **выводы**. В статье рассмотрен новый подход для определения значимости объектов фольклорных песен, основанный на построении теоретико-графовых моделей, отражающих семантическую структуру текстов, и вычислении пяти числовых характеристик (на примере беседных песен Заонежья XIX – начала XX века). По ним можно судить о локальной и общетекстовой значимости объектов фольклорной песни. Процедура построения теоретико-графовых моделей и подсчета пяти характеристик была реализована в информационной системе «Фольклор» [6]. Пользовательский интерфейс этой программы разработан таким образом, что позволяет специалисту в области филологии без какой-либо дополнительной подготовки вводить тексты, строить графы и выполнять их обработку. Отметим, что система создана на кафедре информатики и математического обеспечения Петрозаводского государственного университета в среде визуального программирования Delphi 7.0.

В дальнейшем одним из возможных направлений данной работы может стать количественное измерение значимости более крупных единиц фольклорного текста (например, песенных мотивов [4]). Также на основе вышеперечисленных характеристик можно выявлять закономерности при сравнении текстов разных видов, жанров, регионов и т.д.

Список источников

1. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики: в 2-х т. М.: ЮНИТИ-ДАНА, 2001. Т. 1. 656 с.
2. Калашникова Р. Б. Беседы и беседные песни Заонежья второй половины XIX века. Петрозаводск: Изд-во ПетрГУ, 1999. 162 с.
3. Лебедев А. А., Москин Н. Д. Применение метода дискриминантного анализа для сравнения групп фольклорных песен разных территорий // Лингвофольклористика. 2018. № 27. С. 141-148.
4. Мальцев Г. И. Традиционные формулы русской народной необрядовой лирики (исследования по эстетике устно-поэтического канона). Л.: Наука, 1989. 165 с.
5. Морено Я. Л. Социометрия. Экспериментальный метод и наука об обществе. М.: Академический проект, 2004. 320 с.
6. Москин Н. Д. Инструменты исследования текстовых коллекций на основе теоретико-графовых моделей в информационной системе «Фольклор» // Прикладная информатика. 2010. № 4 (28). С. 48-62.
7. Москин Н. Д. Теоретико-графовые модели фольклорных текстов и методы их анализа. Петрозаводск: Изд-во ПетрГУ, 2013. 148 с.
8. Народные песни Вологодской и Олонецкой губерний, собранные Ф. Студитским: в 2-х ч. СПб.: Тип. А. Бородина и К^о, 1841. Ч. 2. Народные песни Олонецкой губернии. 116 с.
9. Рой О. М. Исследования социально-экономических и политических процессов. СПб.: Питер, 2004. 364 с.
10. Свод таджикского фольклора: в 4-х т. / под ред. И. Г. Левина. М.: Наука, 1981. Т. 1. Басни и сказки о животных. 389 с.
11. Скороходько Э. Ф. Семантические сети и автоматическая обработка текста. К.: Наукова думка, 1983. 218 с.
12. Хроленко А. Т. Введение в лингвофольклористику: учебное пособие. Изд-е 2-е. М.: Флинта, 2016. 194 с.
13. Хроленко А. Т. Поэтическая фразеология русской народной лирической песни. Воронеж: Изд-во Воронежского университета, 1981. 163 с.
14. Хроленко А. Т. Семантика фольклорного слова. Воронеж: Изд-во Воронежского университета, 1992. 139 с.
15. Шишкова Н. Э. Частеречное соотношение частотной лексики в текстах обрядовых и необрядовых песен (по сборникам П. В. Киреевского и М. Г. Халанского) // Лингвофольклористика: сборник научных статей. Курск: Издательство КГПУ, 1999. Вып. 1. С. 36-42.
16. Shchegoleva L., Lebedev A., Moskin N. Recognition of Folklore Texts and Author's Poems Using Classification Trees and Neural Networks // Proceedings of the 22nd Conference of Open Innovations Association FRUCT. Jyväskylä: University of Jyväskylä, 2018. P. 418-420.

IDENTIFYING THE IMPORTANCE OF FOLK SONG OBJECTS USING GRAPH-THEORETICAL MODELS

Moskin Nikolai Dmitrievich, Ph. D. in Technical Sciences, Associate Professor
Petrozavodsk State University
moskin@petsu.ru

The article focuses on identifying the importance of folk song objects with mathematical methods. Using Zaonezhye get-together songs of the XIX – early XX century as the research material the author develops a graph-theoretical model representing the semantic structure of a text and calculates the quantitative characteristics of this model. They allow evaluating the importance of folk song objects at the local and textual levels. The research was conducted using the “Folklore” informational system.

Key words and phrases: linguo-folkloristics; get-together song; graph-theoretical model; importance of text elements; “Folklore” informational system.