

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СТРУКТУРЫ ТЕКСТОВ АВТОРСКИХ СТИХОТВОРЕНИЙ И ФОЛЬКЛОРНЫХ ПЕСЕН С ПОМОЩЬЮ ТЕОРЕТИКО-ГРАФОВЫХ МОДЕЛЕЙ

Москин Н. Д. (moskin@petsu.ru),
Варфоломеев А. Г. (avarf@petsu.ru),
Лебедев А. А. (perevodchik88@yandex.ru)
ПетрГУ, Петрозаводск

Фольклорные и литературные тексты

Многообразие связей между фольклорной и литературной традицией едва ли может быть поставлено кем-то под сомнение. Многие из прозаиков и поэтов черпали свое вдохновение именно в фольклорных текстах, принося в свои произведения как народную мудрость в целом, так и систему присущих народным произведениям образов. Однако сможет ли обыватель отличить фольклорный текст от литературного произведения, стилизованного под фольклор, да и всегда ли по силам аналогичная задача подготовленному филологу?

Бесёдные песни Заонежья XIX – начала XX века и стихотворения Н.А.Клюева

Возникает интересная задача выделения критериев, позволяющих отличить собственно фольклорные тексты от стилизованных под фольклор, а также возможной автоматизации процесса разделения текстов. Для сравнения нами были выбраны две коллекции:

- *бесёдные песни Заонежья XIX – начала XX века,*
- *стихотворные тексты Н. А. Клюева.*

Н. А. Клюева называли народным поэтом и «провозвестником народной культуры», «духом добра и красоты, исходящим от потаенной народной России».

Объекты текстов

Согласно А. Т. Хроленко в текстах можно выделить следующие группы объектов (лексико-семантических полей):

- *персонажи*: красна девица, добрый молодец, родной батюшка;
- *части человеческого тела*: белые руки, желтые кудри;
- *проявление качеств человека, его эмоций, этикет*: низкий поклон, громкий голос, худая слава;
- *одежда и украшения*: золотой перстень, алая лента;
- *жилище*: новая горенка, высокий терем;
- *пища, питье*: зеленое вино, сладка водочка;
- *животный мир*: сизый голубь, белая лебедь, вороной конь;
- *растительный мир*: ракитов куст, белая береза, алый цветок;
- *земля и воды*: зеленый сад, быстрая речка, высокие горы;
- *явления природы*: темная ночь, белый снег, красное солнце;
- *разные предметы*: острая сабля, шелковая плетть.

Действия

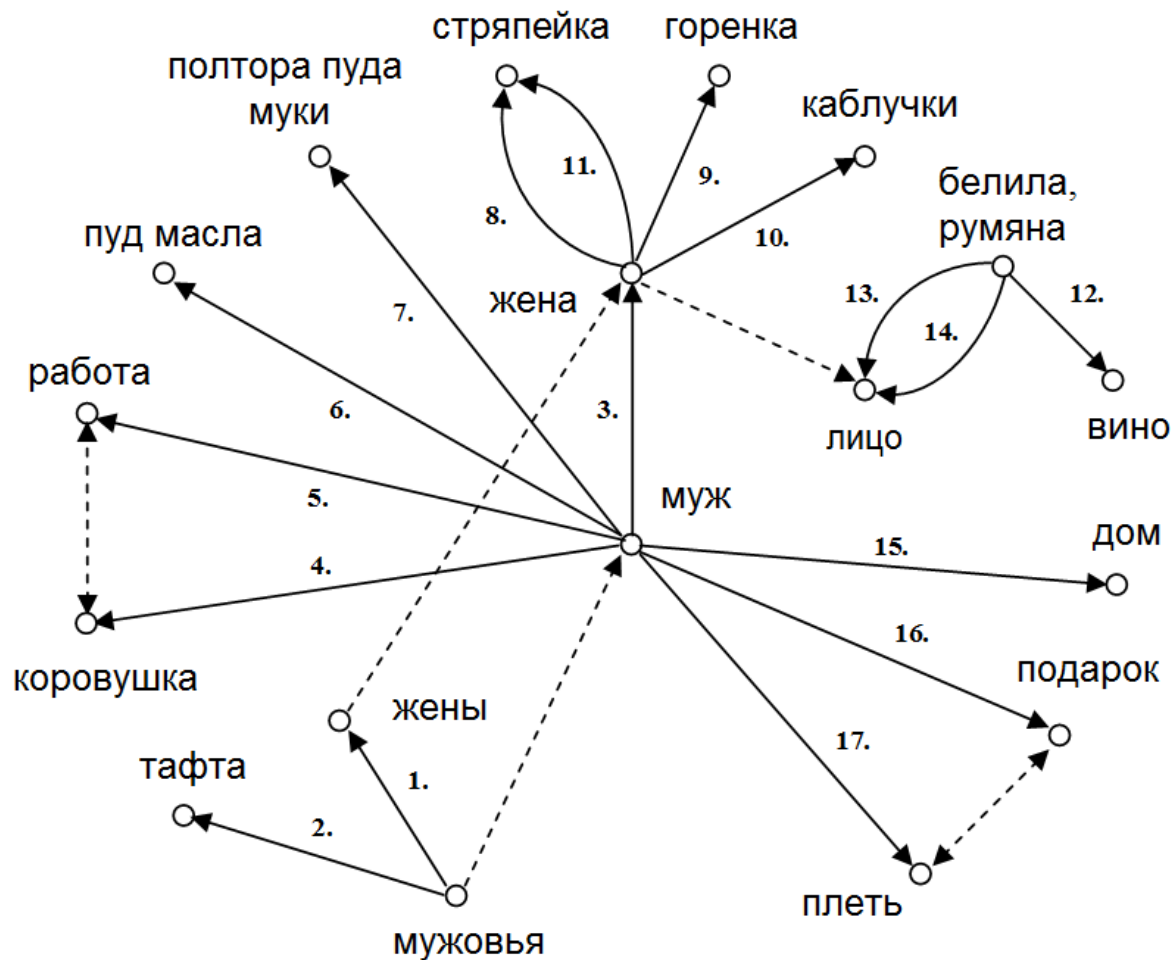
Содержательную основу композиционного фрагмента (мотива) составляет серия типовых действий персонажей («функций» – по аналогии с пропповскими). Вот примеры таких действий (согласно Е. Б. Артеменко):

- где-то кто-то идет (сидит, лежит, находится, живет), на что-то садится, ложится; у кого-то кто-то имеется;
- кто-то кого-то ведет (что-то несет, везет, держит, кладет);
- что-то воспринимает (на что-то смотрит, что-то видит, слышит);
- что-то делает, на кого-то (на что-то) воздействует;
- испытывает или побуждает кого-то испытывать определенные чувства;
- о чем-то думает;
- о чем-то говорит.

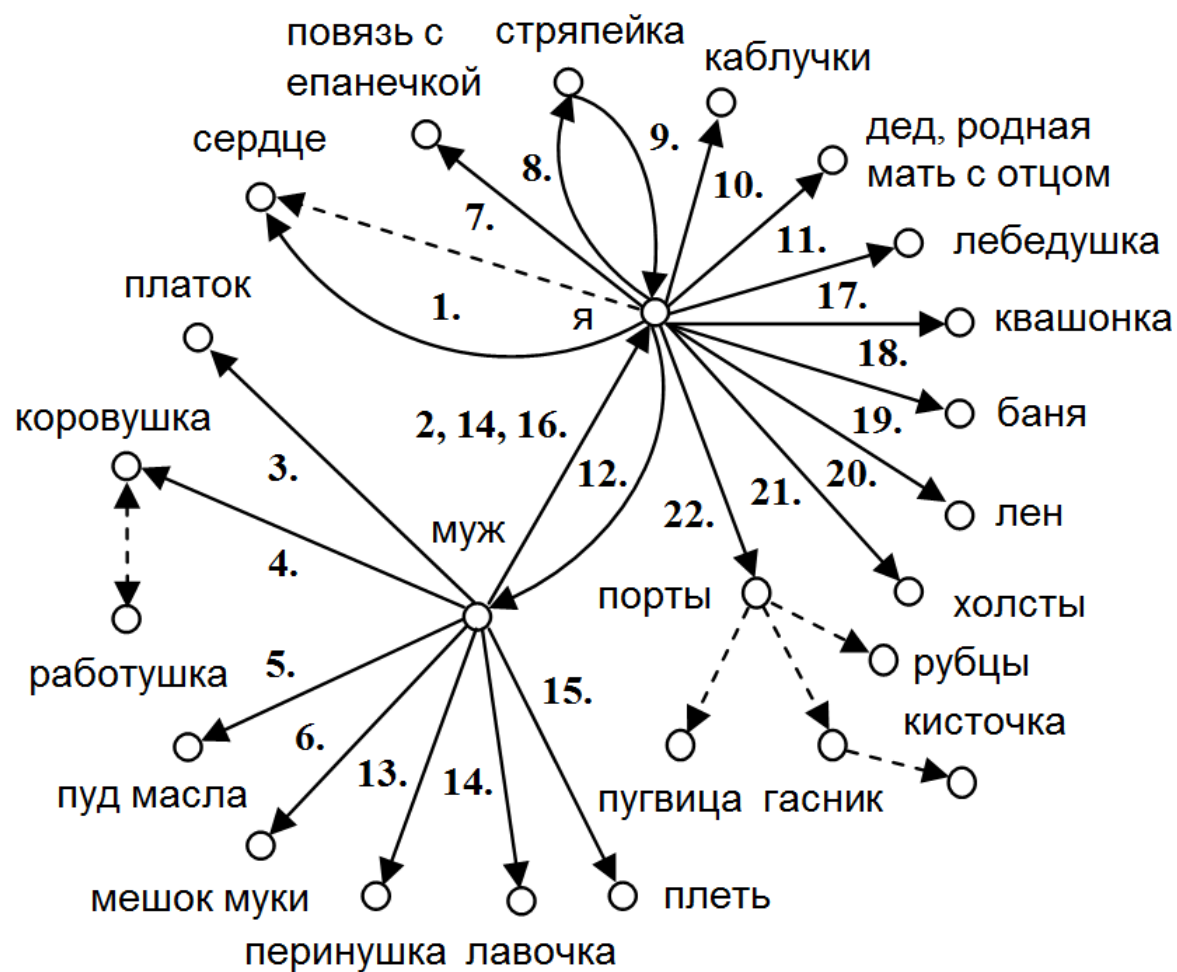
Локальные и глобальные связи

- Действия, подкрепленные в тексте глаголами или отглагольными формами, будем называть *локальными*. Каждой локальной связи ставится в соответствие ее порядковый номер появления в тексте. В ситуации, когда глагол пропущен, но предполагается, что он есть, связь существует.
- *Глобальные связи* распространяются на всю песню и, как правило, никак не выражены в тексте. В фольклорных песнях встречаются отношения равенства (например, коровка – лишняя работа), принадлежности (мужовья – муж, девушка – коса) и месторасположения (горенка – стол).

Теоретико-графовая модель фольклор-ной песни «Все мужовья до жон добры»



Теоретико-графовая модель стихотворения «Я сгорела молоденька без огня»



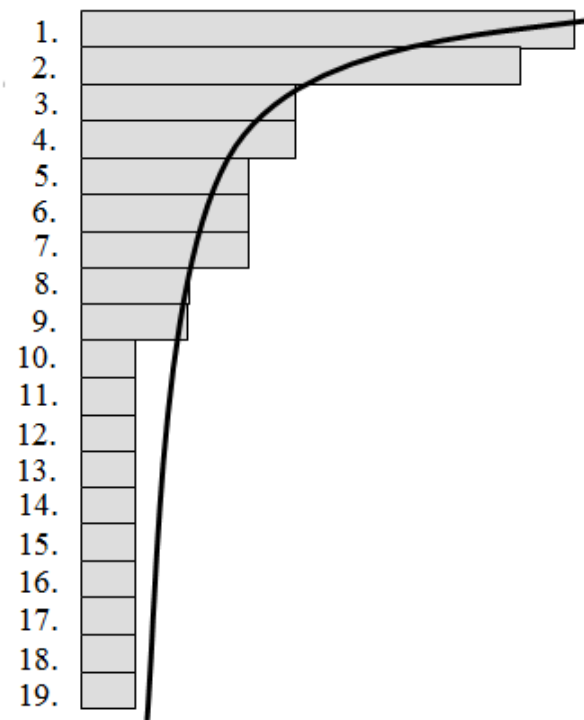
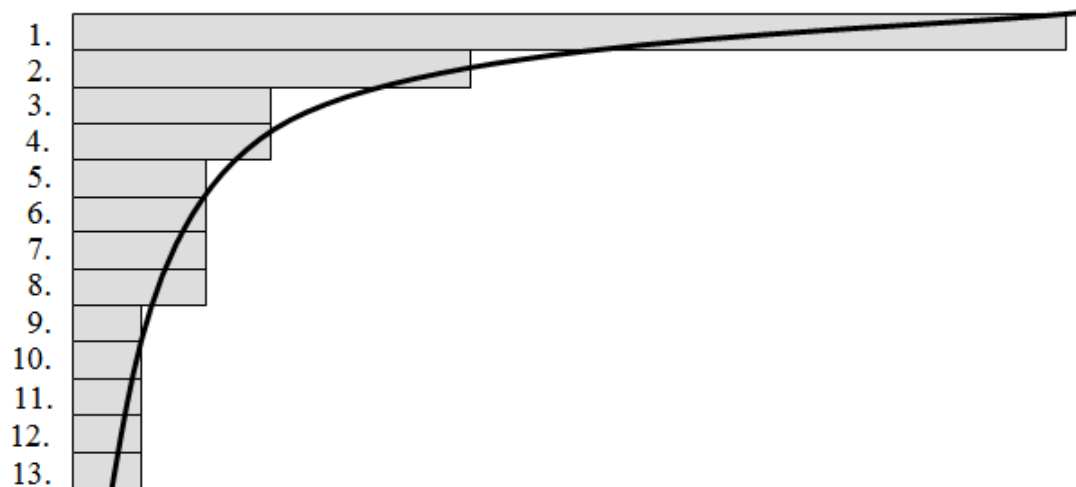
Признаки

Для сравнения теоретико-графовых моделей используем набор числовых характеристик:

- x_1 – число слов в тексте (word);
- x_2 – число строк в тексте (string);
- x_3 – число вершин в графе (vertex);
- x_4 – число дуг в графе (edge);
- x_5 – максимальная степень вершины графа (max);
- x_6 – параметр связности графа (link);
- x_7 – коэффициент гиперболической регрессии a (a);
- x_8 – коэффициент гиперболической регрессии b (b);
- x_9 – процент глобальных связей (global).

Распределение степеней вершин по убыванию

Подобную зависимость можно интерполировать гиперболической кривой вида $y=a/x+b$.



Структура информационной системы «Фольклор»



Дискриминантная функция

Для проведения исследования была взята обучающая выборка из 30 стихотворений Н. А. Клюева и 50 бесёдных песен Заонежья. С помощью статистического пакета SPSS была выполнена процедура дискриминантного анализа для обнаружения тех признаков, которые лучше всего различают (дискриминируют) две группы текстов. Была получена дискриминантная функция:

$$d = 0,006*x_1 - 0,041*x_2 + 0,332*x_3 - 0,359*x_4 + 0,384*x_5 + 4,521*x_6 - 0,116*x_7 - 0,301*x_8 + 3,753*x_9 - 0,518$$

Дискриминантная функция

Подставляя в эту функцию характеристики нового текста, можно вычислить значение дискриминантной функции d . Полученные групповые центроиды показывают, что

- если значение d ближе к 1,878, то объект классифицируется как принадлежащий к классу «стихотворения Н. А. Клюева»,
- если ближе к -1,127, то к классу «бесёдные песни Заонежья».

Результаты анализа

- Из 30 текстов Н. А. Клюева были неверно классифицированы 4 текста (около 13%). Из 50 бесёдных песен Заонежья неверно классифицированы 3 текста, что составляет 6%. Таким образом, в итоге из 80 текстов правильно были классифицированы 73, т. е. 91,25%.
- Для подтверждения правильности построения дискриминантной функции была проанализирована *тестовая выборка*, состоящая из 20 текстов. В результате ее применения 19 из 20 объектов были классифицированы правильно, что составляет 95%.



Спасибо за внимание!