Фольклористика больших данных

Первопричиной этой статьи (и других работ) стала практическая задача упорядочения текстов русских народных песен, создания их систематизированного собрания. Результаты изложены в издании «Синергетическая фольклористика». В нем две части. Первая - печатная, теоретическая («Порядок в хаосе фольклорного микромира»). Вторая («Дети – родители. Систематизированное собрание текстов русских народных песен») - на CD. Все анализы, выводы первой части издания можно проверить на материалах второй.

Текст – объективная данность, конкретное высказывание, имеющее начало и конец, место, время, исполнителей. Для упорядочения текстов оказались необходимыми два условия. Во-первых, наличие технических средств. Компьютеры появились вовремя. Во-вторых, переосмысление собственно научных фольклористических установок, подходов. В первую очередь - изменение объемов исследуемого материала. Переход от жанров (много текстов, произведений, которые обусловлены социальными связями, нишами текстов) к произведениям, т.е. к ближайшим отдельным текстам их общностям. Акцентирование особого качества этих общностей. В нашей фольклористике произведения – научный фантом, она их почти не видит. Между тем - это богатейший ресурс для понимания сущности и фольклора, и многих других явлений культуры.

Три важнейших обстоятельства. 1. Каждая песня – это множество текстов, отличающихся между собой, вариантов одного произведения. Нам известна (была записана) малая часть вариантов, кооперативно творивших песни, поддерживавших их существование. Но и записей немало. В собрании учтено 11280 текстов. Это примерно 2450 произведений.

2. Фольклор – устное творчество. Его материальный носитель – создаваемый человеком звук. Он пластичен, легко отзывается на усилия, потребности исполнителя, осознаваемые и неосознаваемые. У кого что болит, тот о том и говорит.

3. Тексты песен поются, т.е. звук выносит в мир не только вербальную, но и музыкальную информацию. Это область музыковедов. Для филолога важно, что кооперация с музыкой, безусловно, влияет на строение текстов, увеличивает их отзывчивость на неочевидное, скрытое внутри человека, содержание.

У филологов-фольклористов есть свои способы, инструменты адаптации материала к научному анализу. Особо важны два. Тексты относятся к большой общности (жанрам). Далее исследуется их жанровая специфика. Тексты пересказываются или фрагментизируются, т.е. сокращаются. Дальше описываются какие-то составляющие текстов (зачины, сюжеты пр.).

Оба способа плохо справляются с песнями собрания. Песен очень много. Они исключительно разнообразны, но в пересказах похожи, разнообразие трудно передать. Наконец, они очень изменчивы. Кроме того – оба способа не дают доступа к ценнейшему ресурсу песен, их вариативности. К тем отличиям и общему, которые есть в вариантах той или иной песни, во всем корпусе песен. Фольклористы регулярно подчеркивают значимость вариативности. Но эти упоминания во многом декларативны.

От общих деклараций к конкретным данным, к цифрам позволил перейти компьютер. Стало можно каждый текст учесть, описать, что-то подсчитать. Это относится и к фрагментам текстов.

Принципиально важно то, что за большими данными, вносимыми в компьютер, стоят большие и очень сложные реальности. В собрании учтены записи 18, 19 и 20 веков. Это тексты русского, т.е. большого народа, расселившегося на большой территории. Данных не просто много. Они разнообразны. В устройстве песен заложены значительные возможности трансформаций, реализуемые через варианты. Своего рода «вербальный солярис». Одна и та

же песня может петься про мужчину или женщину, про людей разного социального положения, События могут трактоваться как радостные или печальные. И т.д. Такие отличия в содержании впечатляют сами по себе. Но дело не только в том, что песни, живя столетия, откликались на разнообразный, меняющийся мир.

Большие данные текстов – это также своего рода «вербальный томограф», глубоко проникающий в человека, обнаруживающий то, что в человеке устойчиво. Для упорядочения текстов в собрании надо было найти что-то более или менее постоянное, т.е. классификаторы. В текстах варьируется все. Но оказалось, что в них есть части – смыслы, выражающие их слова, которые встречаются чаще. Кроме того эти же немногочисленные части регулярно заменяют друг друга в вариантах, т.е. образуют парадигмы. Принципиально то, что все они относятся к человеку. Несколько упрощая, можно сказать, что они отражают активности человека (например, видит, в руке есть предмет), а также его признаки (например, старый – молодой).

Доступ к информации про внешний мир легче, чем про человека. Реалии внешнего мира можно учесть техническими средствами через употребляемые слова (например, географические названия, наименования тканей, архитектурных деталей и пр.). Для выявления классификаторов необходим исследователь. Очень сложна и нуждается в дальнейшем изучении связь между их смыслом и вербальным выражением. Смысл может выражаться словами разных частей речи (гневается, гнев, гневливый). Задействуется мощный диалектный потенциал (не велят, претят, заказывают, заговаривают и пр.). А соответствующего диалектного словаря может не быть в электронной форме, его может не быть вообще, в словаре не отражено нужное слово. Смысл может быть выражен синтаксически. И пр., и пр..

И общая во всех случаях сложность. Многие тексты напечатаны в старых книгах. Они даже сканированию почти не поддаются. Тем более преобразованию в данные. Тем не менее, когда трудности преодолены, открываются новые нежданные перспективы. О собственно фольклористических новациях отчасти сказано. Но дело не только в них.

Материалы собрания подталкивают к необходимости преобразования филологии в целом. Классификаторы как высокочастотные части текстов – это части управляющей системы, т.е. грамматики. Она вариативна той, что есть в языке, но другая. Фольклор, как и литература, при таком подходе выглядят формами речи наравне с языком, а не вторичными образованиями на его базе. Об этом подробнее см. http://lrybakova.ru/issledovaniya-i-proektyi/folkloristika-poznavatelnyie-perspektivyi/ В этом случае, при таком подходе очевидно, что речь как важнейшая характеристика жизнедеятельности человека может быть понята только через изучение во всех ее формах.

Классификаторы относятся к человеку. Они буквально называют то, что организует поведение человека, его больших сообществ, но плохо улавливается при простом наблюдении, а возможно, и в экспериментах. Это область этологии человека, биологии в разных ее частях. Область многих гуманитарных наук. Об этом подробнее см. СФ, с. 193-245.

Но и это не все. Тексты демонстрируют, что в объекте культуры действуют значимые общие законы, принципы, традиционно подведомственные естественным наукам.

Тексты песен в собрании делятся на две части в зависимости от активностей. Последние названы связками. Они связывают, организуют комплексы слов (фольклорные предложения), которые являются вербальными моделями ситуаций. Связки есть всегда. Тексты двух частей **асимметричны**. Отличаются формами. Тексты первых частей состоят из существенно отличающихся фольклорных предложений, связанных подчинительно (если…то). Тексты вторых частей состоят из похожих фольклорных предложений, связанных через сочинение, нанизывание. Подобное есть не только в песнях. Носителю русской культуры обя-

зательно знакомы, с одной стороны, волшебные сказки, где герои странствуют с приключениями и подвигами, а с другой стороны, сказки о животных (например, «Колобок»), где эпизоды повторяются почти дословно, а волк не предопределяет появление лисы.

Тексты двух частей не только асимметричны. Они отражают разные виды симметрии: поворотную и зеркальную. В тех же, всем известных, волшебных сказках герои уходят в чужой мир, а возвращаются либо с богатством и высоким социальным статусом, либо с потерями и позором. В сказках о животных всегда есть объекты, которые делают ситуацию противоположной (нос лисы и песенка колобка).

Зеркальную симметрию можно визуализировать через одно из многих изображений всемирно известного мирового дерева. Растения – одна из связок в текстах песен вторых частей. Это объект, который либо в/на руке (другой части тела), либо от нее отделился. Такого рода содержание хорошо визуализируется. Разнообразных традиционных изображений много. Кроме того - это содержание легко абстрагируется. Оно много исследовалось. Связки текстов первых частей визуализируются, обобщаются плохо. Наукой недооцениваются (об этом см. СФ, с. 201). Но визуализировать их можно. Например, через спираль ДНК. Естественники говорят: «Есть явления природы, есть законы природы. Но превыше всего принципы симметрии». Это верно и для культуры.

Кроме связок в текстах есть высокочастотные персонажи. Все они – носители комплекса качеств: дети – родители, женщина – мужчина, свой – чужой. Мать – родитель, женщина, своя. Зять – дитя, мужчина, чужой. Все основные персонажи – части парадигмы. Заменяя качества, найдем батюшку, дочь, сына, невестку и др. Принципиально важно, что качества обнаруживают себя только через противопоставление. Дочь есть там, где есть батюшка и пр.. В собрании довольно много текстов, в которых на месте основных персонажей появляются малочастотные, скажем, вместо свекра – соловей. Невестка в таком случае тоже теряет часть своих признаков. Она обозначается словами – я, молода. А дальше? Слова есть. Глубины смысла нет.

Много реже встречаются тексты с гораздо более радикальным преобразованием.

1а) «А моя ли кривошлыкая **свекровь**
Посылает и туды **меня**, сюды…
Посылает часто в погреб за питьем;..
Я немножко там замешкалася,
Зелена вина натрескалася.

Кир., II,I, № 1587 (AI – 13а, 2) (**21**)

1б) **Я** в анбар за мукой ходила,
Я в анбаре призамешкалася,
Аржаной муки натрескалася,
А пшеничной напоролася, -
С ребятишкам поборолася.
Красной булочки накушалася,
Соловьюшечки наслушалася.

ТФНО, № 234 (AI – 13в, 1) (**1**)

2а) Его **матушка** родима,..
Положила мать **княгиню** со младенцем
В белодубову колоду;..
Опустила мать колоду
В сине море Волынско.

Соб., I, № 65 (AI – 22а, 6) (**12**)

2б) Молода моя **княгиня** нездорова:
Ума-разума лишилася,
Во сине море бросилася».

Соб., I, № 69 (AI – 22в, 1) (**1**)

3а) **Меня** девки кликали
На улицу поиграть,..
Меня **свекор** не пустил;
Хотя пустил, - пригрозил:
«Гуляй, сноха, не долго!»

Соб., II, № 510 (AI – 30а, 1) (**10**)

3б) Что меня младу кликали…
Еще **мне младе** недосуг,..

НППК, I, № 291 (AI – 30д, 1) (**2**)

Примеры первого столбца обычны. В примерах второго – дети есть, родителей, каких либо подробностей, свидетельствующих о них – нет. Т.е. если сказать филологически и опять с опорой на общеизвестное, вместо обычных для фольклорных предложений двух подлежащих осталось одно. Такие варианты редки. Под текстами указаны книги, где напечатаны тексты, дальше в скобках – место в собрании, а в следующих скобках – число текстов в редакциях. В первом столбце текстов (обычных) в редакциях больше. В собрании все тексты с одним персонажем выделены в отдельные редакции и проиндексированы.

Тексты с одним подлежащим есть в первых частях собрания, т.е. связаны с определенными активностями. Активностям трудно дать обобщающую характеристику. Для гуманитариев вопрос станет понятнее, если активности обозначить как медиаторы, предикаты. Хотя использование этих слов потребует развернутых объяснений. Активности можно назвать и так, чтобы было понятно естественникам. Это силы. Не те силы, которые изучают физики. Тем не менее у них есть характеристика, которая понятна и физикам, и всем прочим людям. Человека (в первых частях), с одной стороны, видят, слышат, посылают подальше, а с другой стороны, ударяют, убивают (может быть, расчленят) или одаряют ценными подарками (во вторых частях). Очевидно, что в первом случае сила воздействия слабая, а во втором – сильная. Потеря второго подлежащего в фольклорных предложениях – это культурный аналог **несохранения четности** элементарных частиц, известного физикам (см. подробнее в СФ, с. 131-134; 226-229). Однако событие происходит в масштабе человека. Важно, что в его пространстве одновременно есть и слабые, и сильные силы. Представляется, что культурный вариант несохранения четности (а за ним, естественно, скрываются события, происходящие в теле человека) есть не только в текстах песен. Этот феномен чрезвычайно важен для культуры в целом. Во многом определил особенности ее составляющих и их соотношения, связи.

Тексты русских народных песен, видимо, в значительной мере уникальный объект. Будучи оцифрованным и понятым, открывает новые познавательные перспективы. Он является модельным объектом при изучении разных феноменов культуры. Способствует сближению, взаимопониманию гуманитарных и естественных наук. Он, наконец, отчасти отвечает на те мировоззренческие вопросы, которые появляются в цифровую эпоху. Когда в 19 веке появились социальные (классовые) теории, возник вопрос о свободе воли человека, протест против ее ограничений (см. СФ, с. 242-243). Этот вопрос встает вновь во времена больших данных. Тексты как раз и показывают, как в мире больших величин один (каждый) человек остается свободным. При этом через культуру органично взаимодействует с закономерным биологическим наследием, устраивает, организует свою жизнедеятельность.

Литература

Кир. - Песни, собранные П.В. Киреевским: Новая серия: В 2-х в. - М., 1911-I929.

НППК - Народные песни Пермского края: В 2-х т. /Сост. Ганина М.А., Комина Р.В., Пирожкова Т.Ф., Спивак Р.С. - Пермь: Б.и., 1966-1968.

СФ - Рыбакова, Л. В. Синергетическая фольклористика. Порядок в хаосе фольклорного микромира / Л. В. Рыбакова. - М. : Прогресс-Традиция : Технологии управляемого хаоса, 2012. - 335 с. ; 1 CD-ROM (Дети - Родители. Систематизированное собрание текстов русских народных песен). - Доп. тит. л. англ.

Соб. - Соболевекий А.Н. Великорусские народные песни: В 7-ми т. – Спб., 1895-1902.

ТФНО - Традиционный фольклор Новгородской области: (По записям 1963-1976 гг.): Песни: Причитания /Изд. подгот. Жекулина В.И. и др. - Л.: Наука, 1979. - 350 с., нот.