

Графовая визуализация терминологического тезауруса евангельского текста Ф.М. Достоевского (на примере биграммных моделей)

В.В. Борисова¹

Башкирский государственный педагогический университет им. М. Акмуллы, Московский государственный лингвистический университет, Государственный музей истории российской литературы им. В. И. Даля (Музейный центр «Московский дом Достоевского»), Уфа, Россия

¹ ORCID: 0000-0002-9011-0160, vvb1604@gmail.com

Аннотация

Методы и средства научной визуализации посредством графов сегодня активно используются во многих областях, в том числе в гуманитарных науках. Эффективность их использования очевидна и в филологических исследованиях. В статье описана графовая модель терминологического тезауруса евангельского текста Ф.М. Достоевского, основанная на кластеризации биграмм с учетом их принадлежности к разным тематическим дискурсам. Выявление и анализ биграмм как концептуальных точек терминологического тезауруса имеет важное значение для его сетевой визуализации, благодаря которой видно, в каком понятийном сопряжении «работают» ключевые термины, как они распределяются по определенным кластерам. В одном кластере могут быть термины из разных дискурсов. Это важный показатель того, как сопрягаются термины, насколько целостен в данном случае метаязык исследований евангельского текста Ф.М. Достоевского. Т.е. терминологический тезаурус выполняет в данном случае диагностирующую и прогностическую роль.

Ключевые слова: визуализация научных данных, графовая модель, биграммы, терминологический тезаурус, евангельский текст Ф.М. Достоевского.

1. Введение

В современных междисциплинарных теоретических и прикладных исследованиях широко реализуются различные формы компьютерной визуализации научных данных, машинный анализ которых с необходимостью сопрягается с так называемой «человеческой интерпретацией». Показательно и то, что утвердилось новое понятие «визуалистика» как сфера гуманитарного знания [1, 2].

Универсальным средством визуализации научной информации являются графовые модели [3, 4]. Сегодня они активно используются во многих областях, в том числе в гуманитарных науках. Эффективность их использования очевидна и в филологических исследованиях [5, 6, 7, 8, 9, 10]. Так, Б.В. Орехов обратил внимание на методы решения некоторых задач, связанных, в частности, с анализом литературоведческой терминологии, выделив среди них создание графов как эффективного способа ее моделирования [11].

На наш взгляд, использование такого инструментария целесообразно, например, в описании тезауруса, создаваемого на основе цифровой обработки максимально полного корпуса текстов, что в данном случае позволяет выстроить сетевую модель терминологического словника в виде системы графов, учитывающих совместную встречаемость базовых терминов, и на этой основе описать тезаурус в его основных концептах, вы-

явить их обрастание важными определениями, раскрывающими особенности словоупотребления.

Специфика нашего исследования обусловлена интердискурсивным характером терминологии, сегодня используемой в изучении «евангельского текста» Ф.М. Достоевского (имеется в виду отражение образов, сюжетов, мотивов Евангелия в текстах писателя), приоритетного предмета нового научного направления – этнопоэтики, в центре внимания которой – национальные и конфессиональные особенности русской литературы [12, 13],

Важнейшей научной задачей этнопоэтики является создание понятийного аппарата, непротиворечиво отражающего связи национальной художественной культуры и литературы с православной традицией, необходимо «сделать филологический инструментарий, обращенный <в том числе и к Достоевскому. – В.Б.>, более созвучным онтологическим и аксиологическим ценностям самого писателя» [14, с. 7]. По сути, речь идет о метаязыке русской словесности, в основе которого предполагается тезаурус как структурированный на ценностном основании комплекс терминологически оформленного знания.

В нынешнем виде терминологический словарь «евангельского текста» включает в себя группы базовых терминов и понятий, которые, относясь к сферам общей (теоретической), исторической поэтики и этнопоэтики, богословия и философии, все больше используются во взаимном сопряжении.

Накопленный опыт позволяет произвести полномасштабную экспликацию тезауруса, функционирующего в исследованиях евангельского текста автора «Братьев Карамазовых», дать актуальное описание ключевых понятий и терминов, используемых в достоевковедении, провести их сопоставительный анализ, что необходимо для новой систематизации теоретического знания о Достоевском, определенной «ревизии» его категориального аппарата на соответствие предмету исследования. Имеется в виду адекватность терминологической матрицы изучаемому тексту, в данном случае, евангельскому. Решению данной задачи призван способствовать анализ терминологического тезауруса через его графовую модель.

2. Источники и методы исследования

В качестве исходных данных нами использовался корпус научных текстов в журнале «Проблемы исторической поэтики», с 1990 по 2021 год, (всего 333 статьи, в которых упоминается имя писателя, количество словоупотреблений 1,4 миллиона)¹. Представленный в этом издании объем текстов достаточно репрезентативен для евангельского текста Достоевского и русской классической литературы в целом. Через его тезаурус открывается национальная картина мира, ориентированная на главные концепты и константы христианской традиции.

Чтобы наглядно показать взаимоотношения между терминами в форме, доступной для анализа и интерпретации, мы воспользовались методами автоматической обработки исходных текстов посредством языка программирования Python. Они были очищены от нерелевантного шума (ссылки, цифры, символы переноса строк и т.д.) и автоматически приведены к начальной форме при помощи консольного лемматизатора *Mystem* [15], также из текстов были убраны служебные слова (союзы, предлоги), на следующем этапе составлены частотные списки слов текста, из которых были выделены термины, распределенные по четырем дискурсам (разным предметным областям): общая (теоретическая) поэтика; историческая поэтика и этнопоэтика; богословие; философский и др. дискурсы. Для составления словаря мы выделили из всего списка наиболее частотные термины-униграммы (последовательность из одного слова,

¹ Архивы статей собраны и подсчеты произведены Д. Д. Бучневой (ПетрГУ).

например, «евангелие») и термины-биграмы (последовательности из двух слов, «евангельский текст»)².

Пример анализа графовой модели униграмм представлен нами в предыдущей работе [16]. В данной статье мы концентрируем внимание на визуализации биграмм как наиболее устойчивых терминологических сочетаний, закрепленных традицией научного словоупотребления. Если при построении графов из униграмм мы пользовались ограничением в 50, 40, 30 и 25 совместных словупотреблений двух терминов-узлов в окнах 3 и 5 (совместных употреблений терминов на расстоянии друг от друга до 3 и 5 слов соответственно), то в случае с биграмами их минимальная частота снижена до 5 и 3 совместных упоминаний в окне 5.

Для визуализации графа мы использовали программу Gephi [17]. В Gephi встроено алгоритм максимизации модулярности Louvian [18], который выделяет сообщества автоматически, без предварительно заданного числа сообществ, поэтому мы выбрали именно его как подходящий для задачи визуализации распределения терминов. Граф был уложен алгоритмом Yifan Hu, оптимальным для нашего объема данных (95 и 176 ребёр в графах с ограничением совместной встречаемости 5 и 3 соответственно) [19]. Также были подсчитаны коэффициенты степени узла (количества ребер, исходящих из узла) и степени посредничества узла (коэффициента, показывающего, насколько данная вершина важна для целостности графа). От него же зависел размер узлов.

Ключевые термины («вершины» или «узлы»), фигурирующие в графах как центральные понятийные точки, имеют концептуальное значение в исследованиях «евангельского текста» Достоевского. «Ребра», соединяющие линиями «узлы», наглядно отражают совместную встречаемость этих терминов в тексте. Чем их больше, тем сильнее связь и толще «ребра». При такой сетевой визуализации хорошо видно, в каком понятийном сопряжении «работают», например, термины-биграмы как самые устойчивые словосочетания, как они распределяются по определенным тематическим кластерам. В одном кластере могут быть понятия из разных дискурсов. Это важный показатель того, как они сопрягаются, насколько целостен в данном случае метаязык исследований евангельского текста Ф.М. Достоевского. Т.е. терминологический тезаурус выполняет диагностирующую и прогнозирующую роль.

Анализ терминологического словника

Анализ отсортированных по частоте биграмм позволил выявить в исследованиях евангельского текста Достоевского устойчивые терминологические словосочетания, закрепленные как литературоведческой, так и религиозно-философской традицией. Публикуем пятнадцать биграмм, наиболее показательных для каждого дискурса.

Общая (теоретическая) поэтика	Историческая поэтика и этнопоэтика	Богословие	Философский и др. дискурсы
русский литература 1470 главный герой 256 русский словесность 233 художественный мир 159 поэтика достоевский 119 древнерусский	евангельский текст 301 христианский традиция 147 святочный рассказ 134 христианский реализм 94 цитата реминисценция 91 мотив сюжет 87 православный традиция	блудный сын 156 иисус христос 130 священный писание 119 образ христос 99 православный церковь 81 ветхий завет 78 божий мать 77 святой отец 72	русский народ 226 русский человек 204 религиозный философский 96 русский культура 332 христианский культура 74 культурный традиция 62 православный тради-

² Полный список терминов доступен в репозитории нашего проекта, подготовленном А.В. Панасюк (НИУ ВШЭ): <https://thesaurus-dostoevsky.github.io/slovniki> Она же занималась созданием графовых моделей, представленных ниже.

литература 100 русский классика 97 художественный текст 92 русский роман 80 классический литера- тура 79 картина мир 68 русский поэзия 68 литературный жанр 68	83 сюжет жанр 82 русский православный 80 реминисценция мотив 79 пасхальный рассказ 72 канонический текст 71 пасхальный архетип 70 категория соборность 64	духовный жизнь 68 великий пост 56 евангельский сло- во 56 царство божий 53 святой дух 53 евангельский притча 50 страшный суд 48	ция 83 социальный христиан- ство 55 славянский культура 55 русский мир 51 русский религиозный 44 православный культу- ра 43 русский мысль 43 русский земля 43 христианский идея
--	--	--	---

В итоговую версию словника вошло более 200 биграмм, анализ которого с тематической точки зрения выявил следующие показательные особенности его структуры и содержания: во-первых, очевидна синонимическая соотнесенность используемых понятий, свидетельствующая о расширении терминологического поля, дополнении литературоведческого инструментария категориями из других дискурсов.

Так, в группу ключевых терминов, обозначающих предмет этнопоэтики, наряду с его традиционными формулировками входят и новые, обусловленные аксиологией и методологией научного направления, в рамках которого изучается евангельский текст Достоевского: вместе с биграммami «русская литература», «русская культура», «славянская культура», «национальная культура», «народная культура» исследователями активно используются, например, терминологические словосочетания «русская словесность», «отечественная словесность», «христианская культура», «православная культура». Сам переход от понятия «русская литература» к понятию «русская словесность» и аналогично от определения «древнерусская литература» к терминологическому словосочетанию «древнерусская словесность», например, имеет принципиальное концептуальное значение [20, 21]. Термин «словесность» шире понятия «литература». Словесность, согласно Евангелию от Иоанна, лежит в основе жизни: «В начале было Слово...».

Статистические данные терминологического словоупотребления свидетельствуют также о преимущественном исследовательском внимании к классическому периоду русской литературы (об этом свидетельствуют биграммы «классическая литература», «русская классика»), для которого в наибольшей степени характерна связь с православной традицией. Не меньшим этноконфессиональным своеобразием отмечена и «литература русского зарубежья». Конкретизируется предмет изучения художественной литературы благодаря таким биграммам как «русская поэзия», «русская проза», «русский роман», в которых акцентированно подчеркивается определение «русский» в рамках изучения «литературного процесса» и «творческого пути» писателя.

Аналогичным образом выглядит в терминологическом словнике система понятий, центральными в которой являются соотносимые категории «художественный мир», «картина мира», «внутренний мир» произведения. Синонимичны по отношению к ним новые биграммы, обозначающие аксиологический поворот в изучении русской литературы: «русский мир», «православный мир», «духовный мир», «божий мир», «христианский мир». Традиционными атрибутами «образа мира» по-прежнему выступают «художественное время» и «художественное пространство», также активно изучается «усадебный топос» и используются характерные топонимы «русская земля», «святое место», «святая земля», «сионская высота» и др.

Не менее репрезентативна тематическая группа, в центре которой – категории «литературный жанр» и, в частности «жанр Достоевского». Им, наряду с понятием «русский роман», сопутствуют такие жанровые формулы как «святочный рассказ», «пас-

хальный рассказ», «рождественский рассказ», «евангельская притча», «духовный стих», «послание апостола», «житие святого», «Нагорная проповедь», «откровение Иоанна» и др., что также свидетельствует о расширении терминологического аппарата в исследованиях евангельского текста в русской словесности в целом и в творчестве Достоевского в частности.

Эта тенденция очевидна и в системе терминов, связанных с базовыми категориями «художественный образ», «национальный образ», которая активно дополняется богословскими понятиями «христианский образ», «библейский образ», «евангельский образ», «образ Христа», «Иисус Христос», «русский Христос», «божья мать», «пресвятая богородица», перифразами «сын человеческий», «образ божий», «сын божий», а также устойчивыми словосочетаниями «святой отец», «преподобный отец», «блудный сын», «блудная дочь», «злой дух» и др. О том, что дискурс современного литературоведения расширяется, свидетельствует «гнездо» терминов, в которое наряду с традиционными категориями общей поэтики «художественный текст», «поэтический текст», «литературный текст» входят новые биграммы «евангельский текст», «канонический текст», «библейский текст», также как понятия «литературная традиция», «культурная традиция», «античная традиция» дополняются словосочетаниями «христианская традиция», «православная традиция».

Новое понимание отечественной словесности в ее связях с христианским преданием проявляется и в отношении сюжетологии, в терминологический аппарат которой помимо традиционных обозначений «сюжетная линия», «сюжетная ситуация», «сюжет испытания», а также терминологических пар «мотив сюжет», «сюжет жанр» входят биграммы «евангельский сюжет», «библейский сюжет», «Рождество Христово», «воскрешение Лазаря», «светлое Воскресение», «конец света», «преображение человека», «спасение души» и т.п. Аналогично выглядит тематическая группа терминов, в которой традиционные определения «литературная цитата», «литературная реминисценция» дополняются их новыми вариантами «евангельская цитата» и «библейская цитата».

Таким образом, уже на этапе анализа и интерпретации визуализированных данных терминологического словника на материале биграмм возможно сделать следующие выводы относительно современного состояния метаязыка этнопоэтики как приоритетного направления отечественного литературоведения: налицо его интердискурсивный характер, обусловленный динамично развивающимся сопряжением поэтологических, богословских и философских категорий в силу генетических и типологических связей между ними, актуализированных в новом контексте изучения русской словесности.

Что показывают графовые модели терминологического тезауруса евангельского текста Достоевского

Перейдем к анализу и интерпретации графовых моделей терминологического тезауруса евангельского текста Достоевского. Для их построения из корпуса статей были извлечены случаи совместной встречаемости терминов-биграмм на расстояниях до 5 словоупотреблений друг от друга (включительно). Затем были построены два графа, в которых термины автоматически сгруппированы по кластерам (они выделены цветом). Размер узла в графе зависит от значения коэффициента степени посредничества: чем важнее термин для сети, тем больше по размеру его узел. Для определения ключевых узлов в сетевой модели подсчитывалась центральность по степени посредничества, позволяющая выявить наиболее важные термины, которые способствуют взаимосвязи кластеров. Ключевым узлом (термином) в обоих графах является биграмма «русская литература», имеющая наибольший коэффициент *betweenness centrality*.

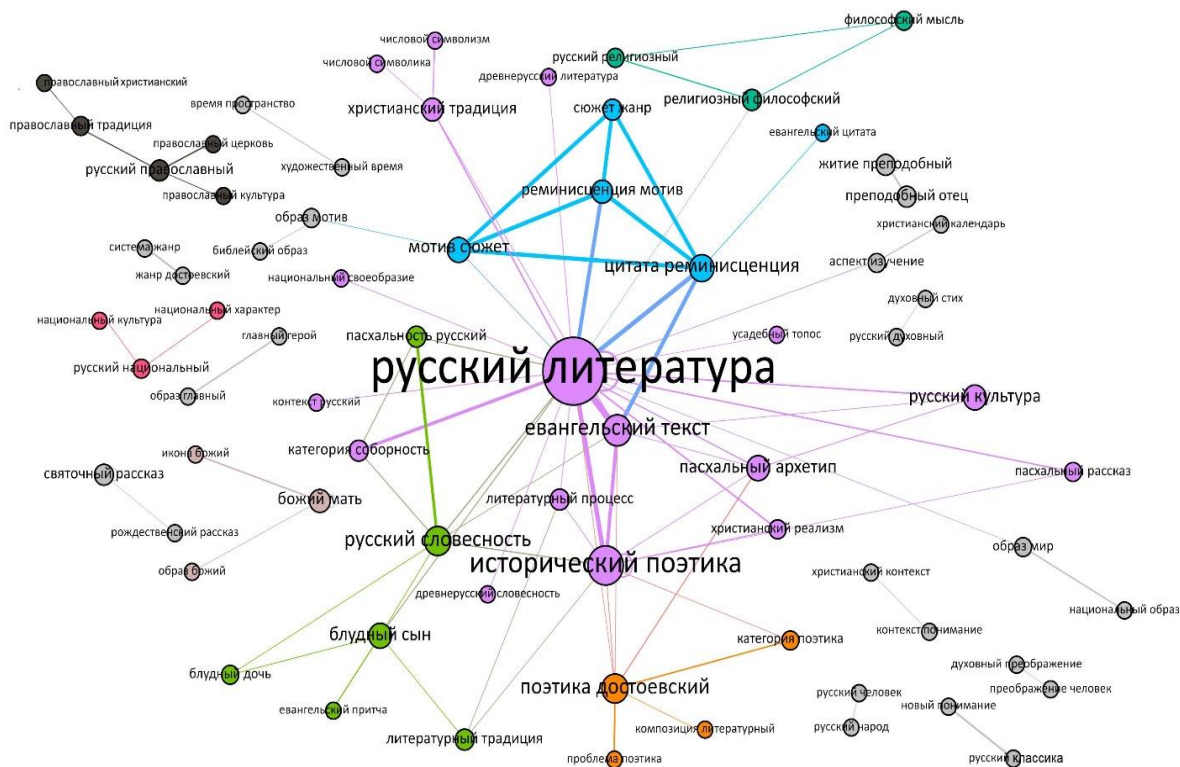


Рисунок 1. Граф терминов с совместной встречаемостью не меньше 5, расстояние 5.

Визуальный анализ данного графа показывает, на какие кластеры он разделяется. Это, во-первых, центральный кластер (фиолетовый цвет), в котором ключевые узлы «историческая поэтика», «евангельский текст» связаны с биграммами «пасхальный архетип», «христианский реализм», «христианская традиция», «категория соборности», «национальное своеобразие» и т.п., что свидетельствует в данном случае о взаимодействии дискурсов исторической поэтики и этнопоэтики.

Наибольшее количество ребер исходит из узла «русская словесность», главного в кластере (зеленый цвет), в котором наряду с понятием «литературная традиция» фигурируют такие устойчивые словосочетания как «блудный сын», «блудная дочь», «евангельская притча». В свою очередь вершина «евангельский текст» достаточно сильно связана с парными терминами исторической поэтики «сюжет жанр», «реминисценция мотив», «мотив сюжет», «цитата реминисценция» и с категорией этнопоэтики «евангельская цитата». Здесь также налицо большая степень сопряжения поэтологических терминов.

Меньшей степенью посредничества в данной сетевой модели отличаются термины-биграммы «философская мысль», «религиозно-философский» и т.п. Визуализация тезауруса показывает также, что ряд биграмм, в частности философские, культурологические и богословские понятия, находятся на периферии графа, несмотря на тенденцию к кластеризации. Это, например, отдельно стоящий монотематический кластер (черный цвет), в который входят биграммы «православная традиция», «православная церковь», «православная культура». Аналогично особняком, без связи с самой большой компонентой, функционирует кластер с терминами (красный цвет) «национальная культура», «национальный характер», «русский национальный» и т.п., что свидетельствует о его локальности.

В целом же, сопряжение терминов-биграмм, относящихся к дискурсам богословия, философии и этнопоэтики, с категориями общей и исторической поэтики, судя по дискретному характеру графа, в современном достоевковедении еще недостаточно полное. Только в центральном кластере с главным узлом «русская литература» проявляется

ся тенденция к совместному словоупотреблению категорий, имеющих отношение к разным дискурсам. Именно с ним связан потенциал развития терминологического инструментария евангельского текста Достоевского.

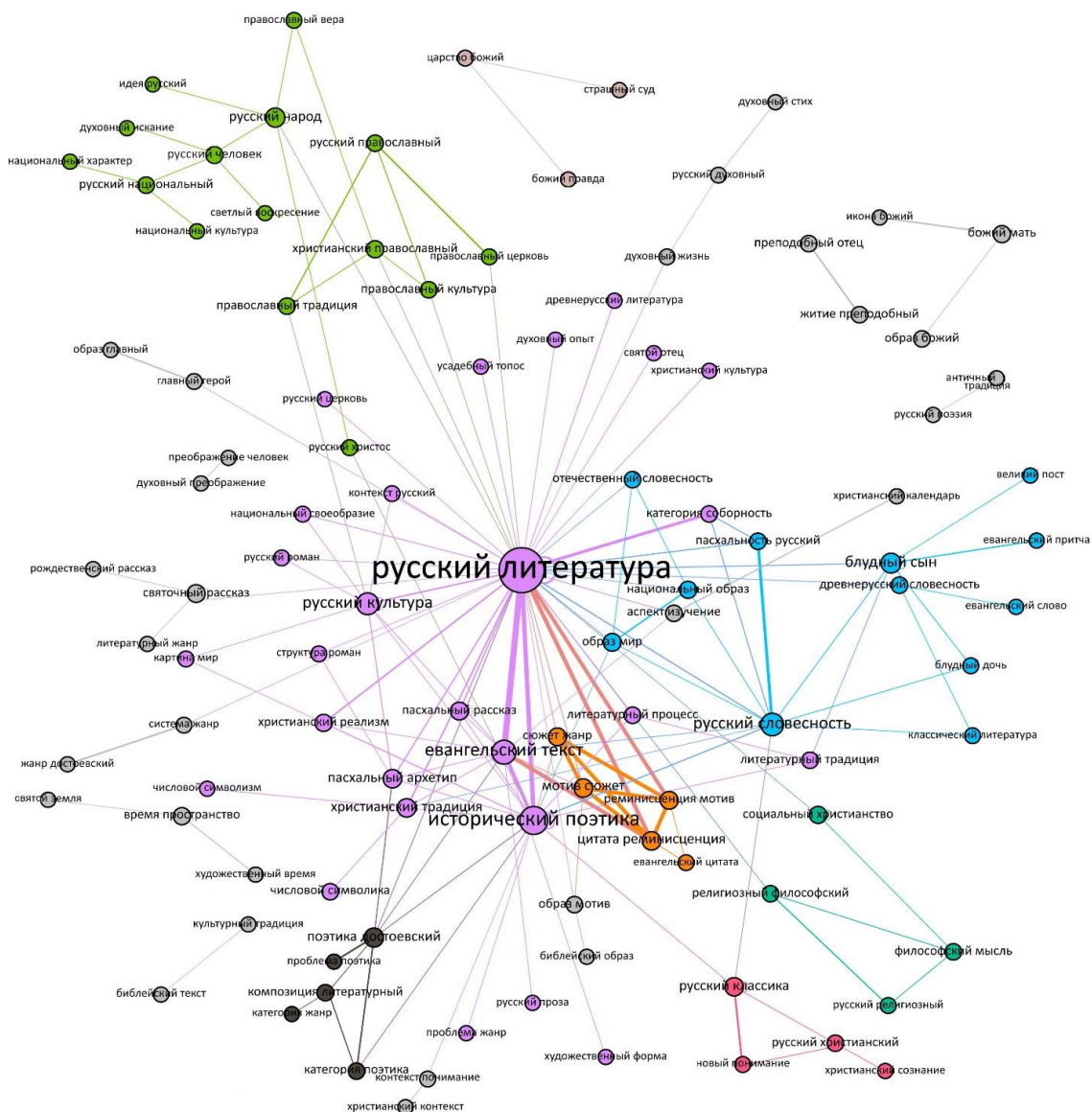


Рис. 2. Граф терминов с совместной встречаемостью не меньше 3, расстояние 5.

В целом данный граф выглядит менее дискретным, но его отличает преимущественная монокластеризация, выявляются несвязанные компоненты, что свидетельствует о неоднородности тезауруса евангельского текста Достоевского. Хотя от главной биграммы «русская литература» исходит наибольшее количество ребер ко всем кластерам, по-прежнему обособленно на периферии графа представлен ряд богословских понятий («преображение человека», «духовное преображение», «вечная жизнь», «образ божий», «царство божие» и др.). Не случайно они имеют нулевой коэффициент посредничества с другими терминами. Т.е. степень сопряжения терминов, в том числе биграмм, относящихся к разным дискурсам, в современном достоевковедении остается невысокой, поскольку некоторые термины фигурируют вне главной компоненты графа.

Тем не менее, в нем очевидна тенденция к совместному словоупотреблению разнородных понятий, к их семантическому расширению, когда, например, изначально богословские категории в литературоведческом контексте обрастают новыми значениями.

ми. В частности, в биграмме «блудный сын» выделяются два значения: «блудный сын» как персонаж библейской притчи и «блудный сын» как тип литературного героя, характерный для русской словесности, что свидетельствует о наложении и взаимопроникновении дискурсов.

5. Заключение

Наше исследование позволяет сделать вывод о том, что графы как универсальное средство визуального моделирования научной терминологии вполне эффективны в филологических исследованиях. Так, анализ терминологического тезауруса евангельского текста Достоевского через его графовую модель показывает, насколько целостен в данном случае метаязык нового направления – этнопоэтики – в отечественном литературоведении.

С одной стороны, налицо его интердискурсивный характер, обусловленный динамично развивающимся сопряжением поэтологических, богословских и философских категорий в силу генетических и типологических связей между ними, актуализированных в новом контексте изучения русской словесности.

С другой стороны, степень сопряжения терминов-биграмм, относящихся к разным дискурсам, судя по дискретному характеру графовых моделей, в современном достоевковедении недостаточно высокая. Однако очевидна и тенденция к совместному словопотреблению разнородных понятий, к взаимопроникновению дискурсов.

Таким образом, графовая визуализация, в данном случае терминологического тезауруса евангельского текста Достоевского, выполняет важную диагностирующую и прогностическую функцию, необходимую для дальнейшего развития научного аппарата в современных исследованиях русской литературы.

Благодарности

Статья подготовлена в рамках проекта «Терминологический словарь-тезаурус "евангельского текста" Ф.М. Достоевского» при финансовой поддержке РФФ (грант № 22-28-00833).

Список литературы

1. Дроздова А.В. Специфика визуальных исследований в современном гуманитарном знании // Ярославский педагогический вестник. 2015. № 3. С. 254-259.
2. Мочалова К.А., Ефремов А.В. Специфика визуализации данных в области цифровых гуманитарных наук // Современная наука: вопросы теории и практики. Сборник статей по материалам всероссийской научно-практической конференции. Йошкар-Ола, 2019. С. 100-110.
3. Касьянов В. Н., Касьянова Е. В. Визуализация информации на основе графовых моделей // Научная визуализация. 2014. Т. 6, № 1. С. 31-50.
4. Касьянов В.Н. Visual Graph – система визуализации сложно структурированной информации большого объема на основе графовых моделей // Труды Юбилейной 25-й Международной научной конференции ГРАФИКОН'2015 (Протвино, 22–25 сентября 2015 г.). 2015. С. 154-162.
5. Орехов Б.В. Моделирование терминологического тезауруса работ Р. Г. Назирова о мифологии и истории фольклорных сюжетов // Назировский архив. 2015. № 2 (8). С. 118–131 [Электронный ресурс]. URL: <http://nevmenandr.net/nazirov/journal20152.php> (10.07.2022)
6. Орехов Б.В. Что может рассказать теория графов о терминологической системе О.М. Фрейденберг? // Вестник РГГУ. Сер.: Литературоведение. Языкознание. Культурология. 2017. № 4 (25). С. 89–98 [Электронный ресурс]. URL: <https://history.rsu.ru/jour/article/view/292> (10.03.2022). DOI: 10.28995/2686-7249-2017-4-89-98

7. Орехов Б.В. В сети терминов М. М. Бахтина: теория графов о диалоге, карнавале и хронотопе // Диалог. Карнавал. Хронотоп. 2018. № 1. С. 105–115.
8. Пильщиков И. А., Сысоев М., Бойков В. Н., Захаров В. Е. Тезаурус как инструмент поэтологии // Моделирование и анализ информационных систем. 2010. Т. 17. № 1. С. 5–24.
9. Пильщиков И. А., Бойков В. Н. Семантическая модель "тезауруса по поэтологии" в составе информационно-аналитической системы // Интернет и современное общество. Сб. научных статей: Труды XVI Всероссийской объединенной конференции. Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики. 2013. С. 273–279.
10. Skorinkin D. A. Extracting character networks to explore literary plot dynamics // *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*. – 2017. – С. 257–270.
11. Орехов Б.В. Гуманитарная терминология как сеть: теория графов о закономерностях научного стиля // Критика и семиотика. 2016. № 2. С. 94–101.
12. Захаров В. Н. Идея этнопоэтики в современных исследованиях // Проблемы исторической поэтики. 2020. Т. 18. № 3. С. 7–19 [Электронный ресурс]. URL: https://poetica.pro/files/redaktor_pdf/1593805089.pdf (10.03.2022). DOI: 10.15393/j9.art.2020.8382
13. Борисова В.В. Евангельский текст в творчестве Ф. М. Достоевского: проблемы и перспективы изучения // Проблемы исторической поэтики. 2020. Т. 18. № 4. С. 186–208. [Электронный ресурс]. URL: https://poetica.pro/files/redaktor_pdf/1604078344.pdf
14. Есаулов И. А., Тарасов Б. Н., Сытина Ю. Н. Анализ, интерпретации и понимание в изучении наследия Достоевского. М.: Индрик, 2021. 336 с.
15. Segalovich I. A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine // *MLMTA*. – 2003. – С. 273–280.
16. Борисова В.В., Шаулов С.С. Терминологический тезаурус евангельского текста Ф. М. Достоевского // Неизвестный Достоевский. 2022. Т. 9. № 2. С. 117–136. [Электронный ресурс]. URL: https://unknown-dostoevsky.ru/files/redaktor_pdf/1657603083.pdf
17. Bastian M., Heymann S., Jacomy M. Gephi: An Open-Source Software for Exploring and Manipulating Networks // Abstract of the International AAAI Conference on Weblogs and Social Media. 2009.
18. Blondel V. D., Guillaume J., Lambiotte R., Lefebvre E. Fast unfolding of communities in large networks // *Journal of Statistical Mechanics: Theory and Experiment* (10), p. 1000. 2008.
19. Hu, Y. Efficient and high-quality force-directed graph drawing. // *Mathematica Journal*. 10. 37–71. 2006.
20. Захаров В. Н. Есть ли у нас литература? Концепты «литература» и «словесность» в русской критике // Проблемы исторической поэтики. 2016. Вып. 14. С. 7–15 [Электронный ресурс]. URL: https://poetica.pro/files/redaktor_pdf/1482396216.pdf
21. Моторин А. В. «Словесность» и «литература»: духовные основы русского и западноевропейского отношения к словесному творчеству // Проблемы исторической поэтики. 2008. Т. 8. С. 17–36 [Электронный ресурс]. URL: <https://poetica.pro/journal/article.php?id=2590> (10.03.2022).

Graph Visualization of the Terminological Thesaurus by F.M. Dostoevsky's Gospel Text (using the Example of Bigram Models)

V.V. Borisova¹

M. Akmullah Bashkir State Pedagogical University, Moscow State Linguistic University,
V. I. Dahl State Museum of the History of Russian Literature (Museum Center “Moscow
House of Dostoevsky”), Ufa, Russia

¹ ORCID: 0000-0002-9011-0160, vvb1604@gmail.com

Abstract

Methods and means of scientific visualization through graphs are actively used today in many areas, including the humanities. The effectiveness of their use is also evident in philological research. The article describes the graph model of the terminological thesaurus of the Gospel text by F.M. Dostoevsky, based on the clustering of digrams, taking into account their thematic belonging to different discourses. The identification and analysis of digrams as conceptual points of a terminological thesaurus is important for its network visualization, thanks to which it is clear in which conceptual conjugation key terms “work”, how they are distributed among certain clusters. In one cluster there can be terms from different discourses. This is an important indicator of how the terms are conjugated, how integral in this case is the meta-language of the studies of the Gospel text by F.M. Dostoevsky. Thus, the terminological thesaurus performs in this case a diagnostic and prognostic role.

Keywords: visualization of scientific data, graph model, digrams, terminological thesaurus, gospel text F.M. Dostoevsky.

References

1. Drozdova A.V. The specifics of visual research in modern humanitarian knowledge // *Yaroslavl Pedagogical Bulletin*. 2015. No. 3. S. 254-259.
2. Mochalova K.A., Efremov A.V. Specifics of data visualization in the field of digital humanities // *Modern Science: Theory and Practice*. Collection of articles based on the materials of the All-Russian scientific-practical conference. Yoshkar-Ola, 2019, pp. 100-110.
3. Kasyanov V. N., Kasyanova E. V. Information visualization based on graph models // *Scientific visualization*. 2014. V. 6, N 1. C. 31-50.
4. Kasyanov V.N. Visual Graph is a system for visualizing complexly structured information of a large volume based on graph models // *Proceedings of the Anniversary 25th International Scientific Conference GRAFICON'2015 (Protvino, September 22–25, 2015)*. 2015. S. 154-162.
5. Orekhov B.V. Modeling the terminological thesaurus of R. G. Nazirov's works on mythology and the history of folklore plots // *Nazirovsky archive*. 2015. No. 2 (8). P. 118–131 [Electronic resource]. URL: <http://nevmenandr.net/nazirov/journal20152.php> (07/10/2022)
6. Orekhov B.V. What can graph theory tell about O.M. Freidenberg? // *Vestnik RGGU. Ser.: Literary criticism. Linguistics. Culturology*. 2017. No. 4 (25). P. 89–98 [Electronic resource]. URL: <https://history.rsuh.ru/jour/article/view/292> (03/10/2022). DOI: 10.28995/2686-7249-2017-4-89-98
7. Orekhov B.V. In the network of terms of M. M. Bakhtin: graph theory about dialogue, carnival and chronotope // *Dialogue. Carnival. Chronotop*. 2018. No. 1. P. 105–115.
8. Pilshchikov I. A., Sysoev M., Boikov V. N., Zakharov V. E. Thesaurus as a tool of poetology // *Modeling and analysis of information systems*. 2010. V. 17. No. 1. S. 5-24.

9. Pilshchikov I. A., Boikov V. N. Semantic model of "thesaurus in poetology" as part of an information-analytical system // Internet and modern society. Sat. scientific articles: Proceedings of the XVI All-Russian United Conference. St. Petersburg National Research University of Information Technologies, Mechanics and Optics. 2013. S. 273-279.
10. Skorinkin D. A. Extracting character networks to explore literary plot dynamics // *Komp'yuternaja Lingvistika i Intellektual'nye Tehnologii*. - 2017. - S. 257-270.
11. Orekhov B.V. Humanitarian terminology as a network: graph theory about the patterns of scientific style // *Criticism and semiotics*. 2016. No. 2. P. 94–101.
12. Zakharov V.N. The idea of ethnopoetics in modern research // *Problems of historical poetics*. 2020. V. 18. No. 3. S. 7–19 [Electronic resource]. URL: https://poetica.pro/files/redaktor_pdf/1593805089.pdf (03/10/2022). DOI: 10.15393/j9.art.2020.8382
13. Borisova V.V. Gospel text in the works of F. M. Dostoevsky: problems and prospects of study // *Problems of historical poetics*. 2020. V. 18. No. 4. S. 186–208. [Electronic resource]. URL: https://poetica.pro/files/redaktor_pdf/1604078344.pdf
14. Esaulov I. A., Tarasov B. N., Sytina Yu. N. Analysis, interpretation and understanding in the study of Dostoevsky's heritage. M.: Indrik, 2021. 336 p.
15. Segalovich I. A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine // *MLMTA*. - 2003. - S. 273-280.
16. Borisova V.V., Shaulov S.S. Terminological thesaurus of the gospel text by F. M. Dostoevsky // *Unknown Dostoevsky*. 2022. V. 9. No. 2. S. 117–136. [Electronic resource]. URL: https://unknown-dostoevsky.ru/files/redaktor_pdf/1657603083.pdf
17. Bastian M., Heymann S., Jacomy M. Gephi: An Open-Source Software for Exploring and Manipulating Networks // *Abstract of the International AAAI Conference on Weblogs and Social Media*. 2009.
18. Blondel V. D., Guillaume J., Lambiotte R., Lefebvre E. Fast unfolding of communities in large networks // *Journal of Statistical Mechanics: Theory and Experiment* (10), p. 1000. 2008.
19. Hu, Y. Efficient and high-quality force-directed graph drawing. // *Mathematica Journal*. 10.37-71. 2006.
20. Zakharov VN Do we have literature? The concepts of "literature" and "literature" in Russian criticism // *Problems of historical poetics*. 2016. Issue. 14. S. 7–15 [Electronic resource]. URL: https://poetica.pro/files/redaktor_pdf/1482396216.pdf
21. Motorin A. V. "Literature" and "literature": the spiritual foundations of the Russian and Western European attitude to verbal creativity // *Problems of historical poetics*. 2008. V. 8. S. 17–36 [Electronic resource]. URL: <https://poetica.pro/journal/article.php?id=2590> (03/10/2022).