

ЕВГЕНИЙ ОБУХОВ



## ПОИСК ИНТЕРТЕКСТУАЛЬНЫХ СВЯЗЕЙ С ПОМОЩЬЮ LLM

Новые возможности

30 ноября 2022 года компания OpenAI выпустила приложение ChatGPT. Ежедневно выходят тысячи приложений, но выход ChatGPT стал своего рода рубежом, после которого искусственные нейронные сети (или более точно «Большие языковые модели» — large language model или LLM) вошли в жизнь миллионов людей. И эту жизнь изменили. Возможно, для специалистов этот день и не был чем-то особенным. Он готовился долго и его ждали. Но де факто после 30 ноября наступила другая эпоха: LLM GPT-3.5 — заговорила с человеком на языке человека.

Статья филолога и математика Евгения Обухова посвящена одному конкретному применению LLM. Литература представляет собой связный контекст, где все объекты (литературные произведения) перекликаются друг другом. И LLM способна эти связи увидеть. Возможно, ей понадобится дообучение, как в эксперименте Обухова. Найденные моделью интертекстуальные связи интересны по-разному, но это именно пробная модель, которая работала на сравнительно небольшом материале и даже на таком объеме она многое заметила.

На мой взгляд, нейросети способны радикально изменить не только поиск интертекстов, но многие другие области литературоведения. До сих пор серьезные филологи даже самые разносторонние сосредотачиваются на определенной теме. Человек просто не может охватить все поле литературы. Поэтому мы вынуждены работать с довольно ограниченным набором «шедевров» (или «каноном»), которые сохранились в «культурном архиве». Время просеивает материал и нам остается только то, что остается.

Такой подход сегодня выглядит недостаточным, но не потому, что мы с помощью LLM найдем нетленные шедевры в прошлом или на каких-то Богом забытых цифровых ресурсах. Такие открытия возможны, но маловероятны. Но мы можем с помощью нейросетевых приложений увидеть связи и выделить характеристики, которые без таких приложений не видны: нейросеть может найти то, о чем мы ее вообще не спрашивали, потому что не знали, что об этом можно и нужно спрашивать.

Используя нейросетевые инструменты, мы можем по-новому подойти к такой, например, проблеме, как литературная эволюция. Тынянов предположил, что на рождение шедевра основное влияние оказывают «второстепенные жанры». Но проблема в том, что «второстепенная литература» — огромна, и почему на гения повлияло то, а не другое произведение, мы не знаем, и ищем практически всегда среди вещей, которые по каким-то причинам остались в «оперативной памяти», а до огромных культурных пластов дотянуться не можем.

---

Обухов Евгений Яковлевич родился в 1989 году в Одессе. Окончил мехмат МГУ им. М. В. Ломоносова. Кандидат филологических наук. Публиковался (в соавторстве с Сергеем Горбушиным) в журнале «Вопросы литературы». Автор (в соавторстве с Сергеем Горбушиным) книг «Удивить сторожа. Перечитывая Хармса» (М., 2012), «Перечень зверей. Перечитывая Хармса» (М., 2017). Постоянный автор «Нового мира». Живет в Вене.

*Статья Евгения Обухова предлагает попробовать поискать на тех путях, где еще никто не ходил, но по которым нас может провести LLM. Не факт, что там лежат одни сокровища, но и сокровища там точно есть.*

*Владимир Губайловский*

**Н**ачнем с мотивирующего примера. В «Сонете» Даниила Хармса есть запоминающиеся строки: «Каково же было их и мое удивление, когда они вдруг обнаружили, что тоже не могут вспомнить порядок счета. 1, 2, 3, 4, 5 и 6 помнят, а дальше забыли». А вот фрагмент из знаменитого романа Ярослава Гашека «Похождения бравого солдата Швейка»: «Велели этому санитару ухаживать за помешанным, который целый божий день ничего не делал, а только сидел в углу и считал: „Раз, два, три, четыре, пять, шесть“, и опять: „Раз, два, три, четыре, пять, шесть“. Это был какой-то профессор. Санитар чуть не лопнул от злости, видя, что сумасшедший не может перескочить через шестерку». Интересующиеся литературоведением прекрасно знают, какое большое внимание в нем уделяется интертекстуальным связям. Хармсоведение — не исключение, что только не упоминается в связи с этим автором. Но конкретно этого сопоставления со Швейком мне не удалось найти ни в монографиях по Хармсу, ни в главной статье по «Сонету», ни с помощью поисковиков, как общих, так и для научной литературы. И это с учетом двух важных обстоятельств: речь идет об очень известных в русскоязычном пространстве текстах (желающий получить представление о прозе Хармса непременно прочтет сборник «Случаи», где «Сонет» практически в самом начале, а роман Гашека входит в негласный список для обязательного прочтения) и об очень точной и вместе с тем нетривиальной связи, на мой взгляд намного более существенной и продуктивной, чем многие другие сопоставления.

Возможно, мне просто не удалось найти соответствующее упоминание (отдельная серьезная тема — современный поиск информации, который в реальности сильно хуже, чем мог бы быть с учетом имеющихся технологий). Но меня здесь интересует не конкретный пример, а суть феномена. Если интертекстуальным связям мы придаем такое большое значение, то странно не пользоваться для их поиска имеющимися возможностями. До появления мощных компьютеров исследователь мог найти связь только благодаря своим навыкам и эрудиции, а также чтению статей и разговорам с коллегами. Но уже достаточно давно у нас есть инструменты для поиска некоторых интертекстуальных связей с помощью методов так называемой обработки естественного языка (NLP). Проблема в том, что, несмотря на постоянный прогресс, у этих методов все еще очень существенные ограничения и ими достаточно трудно пользоваться. Однако появление широкодоступных, удивительно развитых больших языковых моделей (LLM, известный частный случай — ChatGPT, набравший первый миллион пользователей всего за 5 дней) открыло новые революционные возможности для такого поиска. Дело в том, что (если максимально упрощать) прошлые методы так или иначе были алгоритмическими — и придумывать эти алгоритмы должен человек, а большие языковые модели сами по себе генерируют разговор на (том или ином) языке, подражая огромному числу текстов, на которых они обучались, то есть до известной степени могут генерировать речь как люди, примерно по тем же законам (но пока без намерений и самосознания). Четко и окончательно определить, что такое «нетривиальная интертекстуальная связь» — задача, кажется, на сегодняшний день неразрешимая, но примерно объяснить, что это, можно. И для языковой модели — этого достаточно! Уже сейчас она способна и находить высокоуровневые связи, которые или вовсе не получилось бы найти другими вычислительными методами, или было бы необходимо соответствующим образом специально настроить алгоритм (именно для данного вида связи, а

таких локальных видов очень много), и весьма эффективно проводить отбор, не выдавая (в большом количестве) связи, которые нам не нужны. Об этом я и хочу здесь рассказать.

Прогресс в области языковых моделей сейчас идет чрезвычайно быстро. То, с чем ведущая модель не справляется сегодня, уже через месяц может не представлять никаких проблем для новой версии. Именно поэтому здесь будет мало технических подробностей и цитат ответов модели. Мы будем обращать внимание только на те связи, которые модель найти смогла. Новые версии будут становиться только лучше, поэтому эти результаты не устареют. Вместе с тем мы принципиально не будем здесь задумываться о связях, которые модель найти не сумела. Как уже говорилось, прогресс настолько стремительный, что неспециалистам думать о том, чего модель не может, довольно бессмысленно. Возможности, которые есть уже сейчас, велики настолько, что поспеть бы хотя бы за ними...

Для поиска связей я составил корпус из 98 коротких текстов (с ограничением сверху примерно в 10 тысяч знаков с пробелами, полный список приведен в сноске<sup>1</sup>). Корпус получился и случайным, и не совсем. Он не совсем случайный в том смысле, что там намеренно собраны известнейшие русскоязычные авторы разных периодов, у которых часто (хотя не всегда) выбирались очень известные рассказы. Корпус во многом случайный (и именно это было целью!), поскольку при составлении я намеренно старался не думать, как эти тексты связаны между собой, большинство рассказов я либо почти полностью забыл, либо не читал вовсе, а выбор менее известных авторов, как и их текстов, тоже в очень большой степени произволен.

<sup>1</sup> «Студент», «Тоска», «Хамелеон», «Толстый и тонкий», «Смерть чиновника» Антона Чехова, «Алеша Горшок» Льва Толстого, «Сиерра-Морена» Николая Карамзина, «Премудрый пискарь» Михаила Салтыкова-Щедрина, «Старуха», «Капля жизни», «Воробей» Ивана Тургенева, «Три сестры. Видение Минваны» Василия Жуковского, «Мальчик у Христа на елке», «Два самоубийства» Федора Достоевского, «Волчий прием» Алексея Константиновича Толстого, «Пакин и Ракукин», «Молодой человек, удививший сторожа», «Вываливающиеся старухи», «Сундук» Даниила Хармса, «Косцы», «Холодная осень» Ивана Бунина, «Цветок на земле», «Неизвестный цветок» Андрея Платонова, «Белые ночи», «Сильнее смерти», «Безумие», «Счастье», «Сказка» Александра Куприна, «Праздник с сифилисом» Михаила Булгакова, «Очень коротенький роман», «Лягушка-путешественница» Всеволода Гаршина, «Чародей Голод» Лидии Чарской, «Мой первый гусь» Исаака Бабеля, «Аристократка», «Нервные люди», «Баня» Михаила Зощенко, «Счастье» Максима Горького, «Дитя и безумец» Валерия Брюсова, «Тухлый дьявол», «Тайна сия велика есть», «Русалка», «Яшмовая тетрадь» Алексея Толстого, «Дракон», «Австралиец» Евгения Замятина, «Измена» Аркадия Аверченко, «Настенька» Леонида Пантелеева, «Орлянка» Бориса Житкова, «Зеленая лампа» Александра Грина, «Великан» Леонида Андреева, «Бобыль и дружок» Сергея Есенина, «Матрос» Леонида Добычина, «Английская бритва» Константина Паустовского, «Мирная война» Саши Черного, «Слово» Владимира Набокова, «Петька Краснов рассказывает», «Жил человек...» Василия Шукшина, «Жилища» Василия Гроссмана, «Сука Тамара», «Ночью» Варлама Шаламова, «Демоническая женщина» Тэффи, «Путешествие» Юрия Трифонова, «Слово о Родине» Михаила Шолохова, «Подарок» Юрия Визбора, «Щенок» Юрия Бондарева, «Эстафета» Василя Быкова, «Вася Конаков» Виктора Некрасова, «Мы-то не умрем» Александра Солженицына, «Старый да малый» Василия Белова, «Тихая птица» Виктора Астафьева, «Мама куда-то ушла» Валентина Распутина, «Золотые руки» Федора Абрамова, «Привычка наводить порядок», «Палач», «Светофор» Фазиля Искандера, «Ариэль», «Блюз для Натэллы», «Эмигранты» Сергея Довлатова, «Любовь» Владимира Сорокина, «ГКЧП как тетраграмматон» Виктора Пелевина, «Девочка-зверь» Эдуарда Лимонова, «В бане», «Вечная женственность», «Новые нравы» Юрия Мамлеева, «Капля», «Жизнь радостна» Михаила Елизарова, «Дочка» Захара Прилепина, «Бедные родственники», «Дед-шептун» Людмилы Улицкой, «На липовой ноге» Татьяны Толстой, «Продавец добра» Юрия Буйды, «Восемь минут», «Ночь» Дмитрия Глуховского, «Вороний глаз» Олега Владимировича Григорьева, «Я жду» Анны Старобинец, «По-человечески» Линор Горалик, «Красота спасет мир», «Отец Мисаил» Майи Кучерской, «Гойка и хайка» Марка Фрейдкина.

Приведу, на мой взгляд, самые интересные связи, которая смогла найти модель<sup>2</sup>. Ограничусь лишь краткими замечаниями — цели (даже попытаться) провести литературоведческие анализы сейчас нет, цель — ознакомить с грандиозными техническими возможностями (а специалисты в соответствующих областях могут осмыслить потенциал этих связей намного лучше, чем я). Оценивая качество находок модели, важно помнить, что ей не были известны ни автор, ни название текста<sup>3</sup>.

Достаточно нетривиальная связь между главной героиней рассказа «Бедные родственники» Улицкой и «Демонической женщиной» Тэффи. Модель обращает внимание на следующие цитаты: «Ася превращала ее почтенную одежду в лохмотья сумасшедшего. Пришитые Асей черные ленточки в некоторых местах отстали и образовали петли и бантики, и все вместе это напоминало остроумный маскарадный костюм нотной тетради» («Бедные родственники»). А вот из рассказа Тэффи: «Демоническая женщина отличается от женщины обыкновенной прежде всего манерой одеваться. <...> Носит она также и обыкновенные предметы дамского туалета, только не на том месте, где им быть полагается. Так, например, пояс демоническая женщина позволит себе надеть только на голову, серьгу на лоб или на шею, кольцо на большой палец, часы на ногу». Модель также делает некоторые замечания более общего характера, позволяющие еще сильнее и интереснее связать эти образы. Здесь я их не привожу, поскольку скоро модель будет формулировать такие вещи намного лучше, но подчеркиваю, что, на мой взгляд, ей удалось найти весьма перспективную пару для интертекстуального изучения. Эта же Ася сопоставлена с героем еще одного рассказа: «Продавец добра» Буйды. Модель в полном соответствии с текстами отмечает, что в обоих рассказах сумасшедшие

<sup>2</sup> Я использовал o1-preview от OpenAI. Каждый запрос (на русском языке) начинался с задания (для лучшего восприятия текст, изначально предназначавшийся машине, здесь сокращен и отредактирован): «Найди, пожалуйста, в текстах ниже „нетривиальные“ интертекстуальные связи. „Нетривиальные“ означают, что они специфические, небанальные, поразительные, как будто бы один переписывал у другого, они должны выглядеть уникальными. Это должны быть действительно нетривиальные связи, их в этих текстах может и не быть <...> От тебя не требуются интерпретации, пояснения. Ты должен только очень четко, с деталями (совпадающие или очень похожие слова, образы, ситуации) показывать, в чем состоит связь. Если есть одинаковые, похожие слова, приводи соответствующие цитаты. Если связь более общая, пожалуйста, очень четко определи, в чем она состоит. Повторяю, эта связь должна быть настолько близкой, нетривиальной, небанальной, что таких связей между текстами разных авторов обычно не бывает. Если такую же связь можно провести и с многими другими текстами, на которых ты проходил обучение, о которых ты знаешь, то эту связь мы считаем тривиальной и приводить ее здесь не надо. Для каждой связи укажи также „уровень преднамеренности“ (от 0 до 100). Если такая связь могла у авторов получиться случайно, то значение должно быть около 0 (если в литературе такие ходы встречаются довольно часто). Если же детали настолько хорошо соответствуют друг другу, и они сами по себе достаточно специфические, настолько, что почти наверняка один автор, когда это пишет, держит в голове текст другого автора или даже ссылается, то тогда значение должно быть ближе к 100. Ничего не выдумывай и перепроверяй. Запрещается использовать какие-либо ссылки или упоминания авторов, их известные произведения, цитаты и философские идеи, если они не присутствуют прямо в тексте. Связи должны быть выведены строго из содержания данных текстов. Если нет ни одной действительно нетривиальной, неслучайной связи, напиши, пожалуйста, что нет ни одной». Далее следовали 9 текстов. Было четко разделено, где заканчивается один и начинается следующий, однако для большей чистоты эксперимента не были указаны ни название текста, ни его автор. Всего таких запросов к модели (по 9 текстов в каждом) требовалось 143 (чтобы любая пара из 99 текстов была представлена хотя бы один раз, см. теорию систем Штейнера). (Девяносто девятый текст с искусственно сгенерированными связями был контрольный.)

<sup>3</sup> Во многих случаях мы не можем отделить, где модель «рассуждает» сама, а где фактически повторяет написанное в текстах, на которых она обучалась. Однако по моим ощущениям, основанных на почти ежедневном двухлетнем опыте, в этом исследовании модель преимущественно использовала полученные ею «умения», а не «знания».

персонажи в итоге занимаются тем, что приносят людям добро (модель четко проговаривает, что герой Буйды раздает людям «добро» в буквальном смысле). На мой взгляд, это пример блестящей работы модели, которая связала не абстрактно, а аспекты, претендующие на выражение главной мысли обоих рассказов. Интересно сопоставление «Демонической женщины» с «Аристократкой» Зошенко: «— Селедка? Да, да, дайте мне селедки, я хочу есть селедку, я хочу, я хочу. Это лук? Да, да, дайте мне луку, дайте мне много всего, всего, селедки, луку, я хочу есть, я хочу пошлости, скорее... больше... больше, смотрите все... я ем селедку!» («Демоническая женщина»); «Съела она с кремом, цоп другое. <...> И берет третье. <...> И берет четвертое» («Аристократка»). Еще пример блестяще найденной связи: «Старик, увидав девочку, остановился. Катя решилась спросить его. — Скажите, пожалуйста, как пройти в Вифлеем?» («Дитя и безумец», Брюсов); «Молодой человек почистил перчатками свои брюки и деликатным голосом спросил: — Скажите, дедушка, как тут пройти на небо?» («Молодой человек, удививший сторожа», Хармс).

Финальное предложение текста «Капля жизни» Тургенева: «Мальчик недаром проглотил эту каплю — он стал знать все, что только доступно человеческому пониманию, он проник в тайны человеческого организма, и не только излечил своих родителей, — стал могуществен, богат, и слава о нем далеко прошла по свету». Похожее происходит с героем «Зеленой лампы» Грина: некогда бездомный осваивает науки и становится врачом, хирургом. Модель также резонно сопоставляет «Каплю жизни» с только что упоминавшимся рассказом Брюсова «Дитя и безумец»: в центре обоих текстов дети, которые (дейтельно) стремятся спасти своих родителей.

Модель отметила существенные связи между рассказами «Тайна сия велика есть» Алексея Толстого и «Сильнее смерти» Куприна. В обоих текстах умирает (в одном случае «под вечер», в другом «ночью») возлюбленная героя (бывшая — в случае Куприна), и ее дух является ночью герою (возможно, это видение, возможно, нет): «В темную комнату влетела белая и воздушная Наденька, быстро легла к Забелину на постель, охватила холодными руками и стала терзать грудь, чтобы уж, не долго муча, вырвать сердце» («Тайна сия велика есть»); «И вдруг, открыв глаза и точно очнувшись от минутной полудремоты, он увидел ее перед собою, сидящую на диване <...> Она печально покачала головой и вмиг рассеялась, расплылась, как предутренний туман» («Сильнее смерти»). Этот рассказ Алексея Толстого модель сопоставляет и с рассказом Бунина «Холодная осень»: и там и там присутствует тема встречи возлюбленных после смерти. Добавлю, что фактически это главная тема в обоих текстах, вот их финалы: «И Забелин, ослепленный, пронизанный светом, в котором утонуло все вокруг, прильнул к ее лицу и сказал: — Мы в Раю!» («Тайна сия велика есть»); «И я верю, горячо верю: где-то там он ждет меня — с той же любовью и молодостью, как в тот вечер. „Ты поживи, порадуйся на свете, потом приходи ко мне...“ Я пожила, порадовалась, теперь уже скоро приду» («Холодная осень»).

Сопоставляя рассказы «Косцы» Бунина и «Подарок» Визбора модель очень точно фиксирует, что эти тексты объединяет темы ностальгии, невозвратимости прошлого и важность песни. «Это было давно, это было бесконечно давно, потому что та жизнь, которой все мы жили в то время, не вернется уже вовеки. <...> никогда не забыть нам этого предвечернего часа и никогда не понять, а главное, не высказать вполне, в чем такая дивная прелесть их песни. <...> В чем еще было очарование этой песни, ее неизбывная радость при всей ее будто бы безнадежности» («Косцы»); «...во всей стране не было в этот вечер студентов счастливее нас. <...> теперь встречаемся редко — хорошо, если соберемся раз в год. И от нашей поездки в Удмуртию остались воспоминания да песня» («Подарок»). Это всего лишь несколько цитат, замечу, что модели часто удается достаточно тонко уловить и сопоставить общие тональности текстов в целом. Также напомним, что модель не знала, что автор одного из текстов — выдающийся бард.



Интересную нетривиальную связь модель находит между рассказом «Я жду» Анны Старобинец и рассказом «Дочка» Захара Прилепина. В обоих текстах имеет место неожиданное отношение главного героя к кому-то/чему-то как к дочери. В «Дочке» едва ли не весь текст выстроен вокруг суждения рассказчика о том, что к жене надо относиться как к дочери. В «Я жду» главный герой начинает относиться к субстанции, которая вышла из протухшего супа, фактически как... к дочери. При этом слова «дочь» в этом рассказе нет, от этого находка модели еще более впечатляющая. Рассказ «Я жду» тесно связан с «Русалкой» Алексея Толстого. В обоих рассказах герои у себя дома заботятся о необычном женском существе (у Толстого это русалка), идут ради нее на жертвы, полностью изолируются от остальных: «Я очень много для нее делал. Многим пожертвовал. <...> Я позвонил матери и всем своим знакомым и попросил их больше никогда ко мне не приходить. Позвонил на работу и сказал, что увольняюсь. А потом перерезал телефонный провод» («Я жду»); «Продам овцу <...> Замкнулся дед, никого в избу не пускал, ходил мрачный. <...> продал дед лошадь, принес из городе сундучок камушков» («Русалка»). В итоге герой «Я жду» окончательно сходит с ума и попадает в сумасшедший дом, а деда русалка убивает.

Сильная связь найдена между рассказами «Безумие» Куприна и «Вороний глаз» Олега Владимировича Григорьева. Главные события текстов — приход к герою во время помутненного сознания демонической женщины в белом: «В комнату мою входит женщина в белой длинной одежде <...> Лицо ее страшно бледно, длинные черные ресницы опущены <...> но губы необычайно яркие и чувственные. Странная женщина медленно подходит ко мне, ложится со мною рядом и обнимает меня...» («Безумие»); «В фокусе его зрения постепенно проявилось длинное белое платье. Потом длинные же распущенные по плечам белокурые волосы, большие глаза на бледном лице и тонкие руки, сжатые на груди... <...> смотрела ласково, гладила его легкой рукой» («Вороний глаз»). После этих встреч герои разбиты и обессилены. Не менее сильная связь у «Безумия» с рассказом Глуховского «Ночь». И там и там ночь источает силы героя, словно вампир: «с каждым поцелуем она пьет мою жизнь <...> оргические ночи понемногу истощают мою жизнь... <...> это лицо с пунцовыми губами вампира» («Безумие»); «Ночь. Ночь убила мой мир. Она хочет убить и меня <...> Разгадавших вампирическую природу ночи и сущность ее тайнств ждет суровая кара» («Ночь»). Герои неизбежно приближаются к своему концу: «Чем это кончится? Я ослабел, грудь моя ноет <...> Может быть, я скоро умру или сойду с ума?» («Безумие»); «Но силы покидают меня <...> идущие за мной <...> становятся все смелее <...> близится миг их торжества» («Ночь»). Интересно, что модель здесь достаточно успешно выступает и как литературовед, отмечая, что хотя во втором тексте «прямого упоминания женщины нет, ночь персонифицирована и может ассоциироваться с женским образом», что еще больше усиливает связь между текстами (важно напомнить, что модель не знала названия текста — тем удивительнее!).

С одной стороны, вполне очевидную, с другой, весьма глубокую связь обнаружила модель между «Мальчик у Христа на елке» Достоевского и «Мама куда-то ушла» Распутина. Оба текста начинаются с того, как просыпается маленький мальчик, который обращается к маме, но не получает ответа: «мальчик, но еще очень маленький, лет шести или даже менее. Этот мальчик проснулся утром в сыром и холодном подвале. <...> несколько раз с утра подходил к нарам, где <...> лежала больная мать его. <...> раз в десятый уже подходил разбудить свою маму. Жутко стало ему наконец в темноте» («Мальчик у Христа на елке»); «— Мама, я проснулся! Никто ему не ответил. <...> — <...> я проснулся, а тебя нету. <...> Его лицо напряглось в ожидании ответа, он поворачивал его во все стороны, но ответ не пришел, и мальчишка заплакал» («Мама куда-то ушла»).

Про тексты «Мирная война» Саши Черного и «Дракон» Замятина модель отмечает, что в них одна из основных характеристик солдата (и там и там он главный герой, у Замятина назван драконом) дается через спасение им птиц.

Теперь, на мой взгляд, чуть менее сильные/интересные связи, однако все еще достойные упоминания. В рассказах «Вороний глаз» Григорьева и «Русалка» Алексея Толстого главный герой по имени Семен попадает под дурманящее влияние демонического женского существа (мистическая соседка и русалка), сюда же примыкает еще один рассказ Алексея Толстого «Тухлый дьявол» (о котором ниже). В рассказах «Мирная война» Саши Черного и «Красота спасет мир» Майи Кучерской у главных героев вновь фактически одно имя (солдат Лукашка и отец Лука), они оба спасают других. Этот рассказ Кучерской модель сопоставила и со «Студентом» Чехова, отмечая «духовное преображение через взаимодействие с религиозным лицом» (в случае «Студента» следовало бы сделать множество оговорок), а одну из слушательниц студента звали Лукерья (параллель с именем отца Луки у Кучерской). В «Капле жизни» Тургенева и «Капле» Елизарова ключевая роль у этой самой капли, которая падает со свода/потолка (только у Тургенева она приводит к целебному сверхзнанию, у Елизарова к смерти) — снова напомним (здесь это вновь очень важно!), что модели не были известны названия текстов. В центрах рассказов «Цветок на земле» Платонова и «Дед-шептун» Улицкой — ребенок со своим дедом/прадедом; модель делает довольно точные литературоведческие замечания: «в обоих текстах старики передают детям понимание сущности жизни, оба персонажа физически почти слепы, но духовно прозорливы». Рассказ «Настенька» Пантелеева модель сопоставляет с «Аристократкой» Зошенко: и там и там в центре эксцентричное поведение женского персонажа в ресторане/буфете. В текстах «Очень коротенький роман» Гаршина и «Счастье» Горького модель довольно точно отмечает такие сходства: герои вспоминают моменты, когда они могли сделать нравящихся им женщин своими женами, но выбрали другие действия, в результате которых у этих женщин другие мужья.

Сделаю достаточно очевидное, но все же уместное замечание. Есть достаточно ограниченный круг самых распространенных событий, которые могут произойти в художественном тексте. Таково, например, убийство какого-либо персонажа. Еще меньше возрастных и гендерных категорий, поэтому, в частности, женский персонаж старше шестидесяти распространен чрезвычайно. Если у персонажа нет доступа к профессиональному оружию, то использовать для убийства топор — среди самых очевидных вариантов. Таким образом, «убийство старушки топором» — просто контаминация трех вероятных элементов (примерно как «гибель старика от волшебного существа»), но из-за огромного влияния «Преступления и наказания» подобная интертекстуальная связь просто обречена быть очень сильной. Связи из абзаца выше — хорошие примеры того, как существенность связи сильно зависит от значимости (в нашем контексте) соответствующих текстов: если эти тексты нам важны, то эти связи мгновенно становятся очень интересными.

Еще несколько достойных упоминания связей. В текстах «Мальчик у Христа на елке» Достоевского и «Сказка» Куприна очень важную роль играет неготовность людей помочь тем, кто в помощи остро нуждается (и это вопрос жизни и смерти), на фоне холодной зимней ночи. В текстах «Старуха» Тургенева и «Вечная женственность» Мамлеева ключевой персонаж — «хищная» беззубая старуха, символ неотвратимой смерти. И в рассказе «Жилища» Гроссмана, и в крохотке «Мы-то не умрем» Солженицына подчеркивается быстрое забвение умерших, отсутствие должного почтения и памяти. Тексты «Сиерра-Морена» Карамзина и «Любовь» Сорокина обыгрывают сюжет трагической любви, ведущей к смертям (самоубийства, убийство), под аккомпанемент идеализации. «Вот, милые мои, а вы говорите — Беатриче, Беатриче» (финал «Любви»). Несмотря на то, что между рассказами «В бане» Мамлеева и

«Жизнь радостна» Елизарова в общем-то точных совпадений нет, модели все же удалось установить (глубокую) связь между этими родственными авторами, процитируем: «Совпадение имен персонажей (Катя), сходные ситуации, связанные с порочным деторождением, смертью и гротескными образами, делают связь между текстами очень специфичной».

Машинный поиск интертекстуальных связей среди текстов одного автора намного менее актуален: изучая конкретного автора, специалист и так концентрируется на всем корпусе его текстов. Однако наша цель сейчас, напомним, демонстрация возможностей модели — и в этом контексте удачно найденные ею связи между текстами одного автора тоже очень интересны. Да, соответствующие инварианты автора могут быть тривиальными для специалистов, но раз уж модель способна их найти (не зная про тексты ни имени автора, ни названия), то аналогичная работа при сопоставлении текстов разных авторов может увенчаться настоящей ценной находкой.

Сначала приведу, на мой взгляд, самые сильные найденные связи среди текстов одного автора. В рассказах Платонова «Цветок на земле» и «Неизвестный цветок» модель четко выделяет главную тему: цветок-труженик, борющийся за жизнь со смертью в неблагоприятных условиях (напомним, что названия рассказов модель не знала и в первом рассказе цветок фигурировал не так уж часто). Модель приводит отличные цитаты: «Цветок этот — самый святой труженик, он из смерти работает жизнь» («Цветок на земле»); «Он трудился день и ночь, чтобы жить и не умереть» («Неизвестный цветок») и т. д. В обоих рассказах дети с помощью этого цветка постигают что-то очень важное про жизнь. Про тексты «Пакин и Ракукин» и «Молодой человек, удививший сторожа» Хармса модель говорит, что их связывает абсурдность, сюрреализм и специфическая потусторонняя параллель: маленькая злобная душа Ракукина вылезает из тела; молодой человек, спрашивавший, как пройти на небо, исчезает... Также модель сопоставляет Пакина и сторожа.

Остальные найденные связи между текстами одного автора менее сильные, однако это не отменяет хорошей работы модели. В центре рассказов Чехова «Хамелеон» и «Смерть чиновника» страх перед генералом: Жигаловым в первом случае, Бризжаловым во втором. А чеховские «Хамелеон» и «Толстый и тонкий» парные потому, что и там и там подчеркнут контраст поведения персонажа в зависимости от того, о каком чине идет речь. Модель отмечает тесную связь текстов Достоевского «Мальчик у Христа на елке» и «Два самоубийства». И там и там есть смерти молодых людей из-за бедности; в одном тексте мальчик перед смертью видит «Христову елку», в другом говорится о самоубийстве девушки с образом в руках; модель указывает на стилистическую близость, темы кротости и смирения, схожий характер саморефлексии автора. Модель обратила внимание на связь рассказов «Нервные люди» и «Баня» Зошенко: общий стиль (в частности, отмечено использование в обоих текстах специфического слова «теперича»), а также конфликт из-за мелких бытовых предметов (ежика для чистки примуса на кухне в коммуналке и шайки в общественной бане). Модель говорит о большой схожести — в общем и в частности — рассказов Алексея Толстого «Тухлый дьявол» и «Русалка»: и там и там герой сталкивается с дьявольским существом (говорящий козел и русалка, то есть говорящая полурыба), что в финале приводит к катастрофическим последствиям. Модель указывает на сильную связь между рассказами «В бане» и «Вечная женственность» Мамлеева: от общего плана (атмосфера аморальности, саморазрушения, тотальность дьявольского) до конкретных деталей (и там и там очень ярко упоминается Беатриче в контексте «извращенной женственности» — словосочетание модели; и там и там в конце рассказа персонаж поедает плоть — кошки, человека). Не меньше связей модель видит и между «В бане» и другим рассказом Мамлеева «Новые нравы»: в первом персонаж поедает голову мокрой кошки, во втором персонаж, которого позже сравнят с «потусторонним котом», кладет отрубленную голову гостя на стол, оба текста наполнены «(зверо)философией». Сравнивая тексты Хармса, модель (почти)



всегда отмечает абсурдность, но также находит интересные достаточно точные параллели. В текстах «Пакин и Ракукин» и «Вываливающиеся старухи» модель сопоставляет реакции на смерть: «— Так, — сказал Пакин, — подох Ракукин. Пакин перекрестился и на цыпочках вышел из комнаты»; «Когда вывалилась шестая старуха, мне надоело смотреть на них, и я пошел на Мальцевский рынок, где, говорят, одному слепому подарили вязаную шаль». Между текстами Хармса «Пакин и Ракукин» и «Сундук» модель видит такие связи: и там и там есть акцент на шее находящегося между жизнью и смертью героя («шея синяя, просто гадость»; «человек с тонкой шеей»), в обоих текстах герой мистическим образом выходит за пределы замкнутого пространства, где находился (душа Ракукина проходит сквозь стены и выходит из комнаты; герой вдруг видит себя вне исчезнувшего сундука, в котором лежал), и там и там есть персонификация смерти («ангел смерти»; герой устраивает поединок между жизнью и смертью, знаменитый финал: «Значит, жизнь победила смерть неизвестным для меня способом»).

Подчеркну: модели удастся не только найти и сформулировать четкие параллели, но и каким-то образом «почувствовать» глубокие сходства, даже если она иногда (пока) не способна сформулировать, в чем они состоят.

Модель также нашла достаточно много точечных/слабых связей, но я их приведу только в сноске<sup>4</sup>.

<sup>4</sup> В текстах «Ариэль» Довлатова и «На липовой ноге» Татьяны Толстой подсвечено специфическое использование английских слов в русской речи. Этот же текст Толстой сопоставлен с «ГКЧП как тетраграмматон» Пелевина — и там и там речь идет о воздействии слов на реальность; а также с «Премудрым пискарем» Салтыкова-Щедрина — и там и там особая роль у слова «уха». В рассказах «Жизнь радостна» Елизарова и «Ночью» Шаламова персонажи равнодушно относятся к чужим смертям, а если могут, используют их. В текстах «Сказка» Куприна и «Вываливающиеся старухи» Хармса модель отмечает игнорирование персонажами трагических событий. Но вместе с тем в «Сказке» есть и некоторая тревога, оцепенение игнорирующих перед чужой трагедией, нечто похожее модель видит и в тексте «Австралиец» Замятина. В отношении «Лягушки-путешественницы» Гаршина и «Смерти чиновника» Чехова модель, сама того не понимая, делает весьма комичное, но не лишнее резона замечание: к печальному/трагическому исходу главных героев приводит неумение в нужные моменты держать язык за зубами. Не менее забавно замечание модели об этой же сказке Гаршина и «Мирной войне» Саши Черного: в обоих случаях имеет место остроумное использование простого предмета для нетривиального решения трудной задачи (прутик для полета с птицами и перетягивание каната солдатами вместо войны). Между рассказами «Тоска» Чехова и «Шенок» Бондарева выделена такая связь: извозчик/шофер тяжело переживает смерть сына/щенка, оказываясь в психологической изоляции. Этот же рассказ Чехова сопоставлен с рассказом «Бобыль и Дружок» Есенина: и там и там герой не нужен, неинтересен другим людям и фактически единственным его другом оказывается животное. В рассказах «Холодная осень» Бунина и «Сильнее смерти» Куприна разлучившаяся/расставшаяся пара собирается встретиться/встречается мистически после/во время(?) смерти. В текстах «Очень коротенький роман» Гаршина и «Новые нравы» Мамлеева модель выделяет потерю героем ноги. Про тексты «Сказка» Куприна и «Дитя и безумец» Брюсова модель пишет, что и там и там ребенок ночью улавливает зов о помощи или духовный сигнал, хочет откликнуться, но взрослые подавляют от страха или безразличия этот импульс. И в «Зеленой лампе» Грина, и в «Счастье» Куприна богат и царь помещают своих жертв в замкнутое пространство, но они там возвышаются силой разума. В текстах «Сиерра-Морена» Карамзина и «Блюз для Натэллы» Довлатова фигурирует соперничество за прекрасную девушку и тема смерти от оружия. В текстах «Сундук» Хармса и «Восемь минут» Глуховского модель отмечает борьбу жизни и смерти в замкнутом пространстве, недостаток воздуха. Модель проводит параллели между русалкой и демонической женщиной из одноименных рассказов Алексея Толстого и Тэффи. «Русалка» сопоставлена и с рассказом Елизарова «Капля»: и там и там мужчина погибает из-за женского персонажа (русалка, жена), смерть связана с водой. Модель также сопоставляет «Каплю» с «Пакиным и Ракукиным» Хармса — хотя и нельзя сказать, что между ними можно провести по-настоящему четкие параллели, но в целом этот рассказ Елизарова действительно имеет (формально) невероятно много общего с прозой Хармса.

Теперь можно подвести некоторые итоги обработки нашего корпуса из примерно 100 сравнительно коротких текстов. Модель нашла 15 нетривиальных, довольно интересных связей. Еще примерно столько же чуть менее интересных. Около 10 связей между текстами одного автора (это число сильно зависит от того, у скольких авторов в корпусе было несколько текстов и в каком количестве), и примерно 20 точечных/слабых связей. Здесь надо отметить, что с ростом числа текстов в корпусе число пар растет по так называемому квадратичному закону. То есть если у нас в корпусе из примерно 100 текстов было 15-40 интересных нам связей, то в корпусе, в котором в 10 раз больше текстов (примерно 1000), мы можем ожидать уже не 150-400, а целых 1500-4000 связей.

Должен признать, что ни одна из найденных связей не показалась мне столь же нетривиальной и интересной, как связь, приведенная в самом начале — о неспособных считать после 6. А ведь и эту связь трудно отнести к самым интересным в литературоведении. Но здесь надо помнить, что мы исследовали достаточно небольшой по объему корпус (примерно на 20% короче четвертого тома «Войны и мира») во многом случайных текстов и странно было бы ожидать обнаружить в нем что-то феерическое.

Целью было показать огромные возможности для (машинного!) поиска связей. Ясно, что делать это можно совершенно по-разному. Можно искать связи не между случайными текстами, а между текстами, где мы бы ожидали найти что-то нас интересующее. Можно сравнивать тексты, написанные на разных языках, и в этом тоже огромный потенциал. Даже филологи, как правило, владеют всего несколькими языками, а модель может без особой потери в качестве (если речь о распространенных языках) искать связи между оригиналами, а не переводами.

Более того, языковые модели настраиваются. Можно как специальным образом обучить модель, так и дать существующей модели лучше сформулированное задание. Если, например, филолог решит, что при перечислении связей выше совершенно неверно расставлены акценты, что интересны и неинтересны совсем другие связи, это — в большой степени — почти наверняка можно «объяснить» модели. Ее можно натренировать находить именно то, что (вам) нужно. И с течением времени модели наверняка будут становиться и существенно лучше, и дешевле. В этом — определенный вызов филологическому сообществу, сможет ли оно сформулировать, что такое «нетривиальная интертекстуальная связь», захочет ли использовать современные вычислительные возможности. Конечно, «интертекстуальность» не грааль, но с чего-то же надо начинать... До сих пор технологии в литературоведении используются очень слабо.

Важно сказать об одном из ограничений. Работа больших языковых моделей сравнительно дорогостоящая. Каким образом и как быстро будет падать цена их использования в точности предсказать невозможно. Однако уже сейчас, если задаться целью и быть готовым задействовать определенный ресурс, можно обработать достаточно большие объемы текстов.

Приведу несколько примеров потенциальных исследований. Собрать топ-500/топ-1000/топ-5000 главных литературных текстов и найти все связи между ними. Это требует больших вычислительных затрат, зато исследование может быть проведено единожды (а далее будет возможность активно использовать полученные результаты). Можно проводить и намного менее затратные исследования. Взять несколько (возможно, длинных) произведений и искать связи только между фрагментами текстов разных авторов. Или же взять несколько интересующих нас текстов и найти все связи этих текстов с каким-то выбранным нами корпусом (скажем, найти связи «Старухи» и «Случаев» Хармса со всеми текстами Кафки, Брехта, Ионеско, Беккета).

В перспективе можно задуматься о какой-нибудь общей базе данных, на основе которой будет работать интерактивная карта связей. Для данного текста будут выдаваться все известные системе связи. Описания связей могут быть

улучшены авторизованными пользователями (существует множество примеров, когда горизонтальное цифровое сообщество поднимает уровень проекта очень высоко). Разумно также добавить разнообразные оценки связей пользователями, это даст возможность в случае большого числа связей сразу видеть самые существенные.

Хочется надеяться, что для серьезного анализа текстов у нас появятся мощные и удобные инструменты уже в обозримом будущем. Но (пока) за нас это никто не делает.

