

*Посвящается Эндрю (отцу)
и Анджеле (жене)*

1

Бестселлерометр, или Почему анализ текста приводит к революции в книгоиздании

Одним прекрасным утром 2010 года литературный агент Стига Ларссона сидел и радовался жизни. 13 июня «Девушка, которая взрывала воздушные замки»¹ — третья книга в серии ранее неизвестного автора — дебютировала первой в списке бестселлеров в твердом переплете, составленном газетой New York Times. Спору нет, приятно увидеть такое за утренним кофе. «Воздушные замки» на первом месте, «Девушка с татуировкой дракона»² лидирует сразу в двух форматах среди книг в мягком переплете, и, в качестве приятного дополнения, «Девушка, которая играла с огнем»³ — на втором месте. И так — уже 49 недель в США и три года подряд в Европе. Есть чему радоваться.

Через месяц Amazon объявит, что Ларссон — первый в истории писатель, чьи творения разошлись в количестве миллиона экземпляров в формате для Kindle. За следующие два года совокупный тираж всех изданий Ларссона превысит 75 миллионов. Неплохо для никому не известного политического активиста из крохотной скандинавской страны, вдруг заделавшегося писателем. Особенно если учесть, что шведское издание вышло под довольноно

непривлекательным заглавием и что в книге содержатся откровенные описания насилия и пыток. «Мужчины, которые ненавидят женщин» — или «Девушка с татуировкой дракона», как назвали английский перевод, — стала сенсацией года в тридцати с лишним странах.

Журналисты не могли объяснить этот успех. Крупные газеты заказывали специалистам большие статьи, желая понять, что, черт побери, происходит в книжном мире. Почему именно эта книга? Откуда такой ажиотаж? В чем секрет? Кто мог бы предсказать этот поворот событий?

Ответы не впечатляли. Литературные критики чесали в затылке. Они указывали на недостатки структуры, стиля, сюжета и характеров героев. Они ужасались качеству переводов. Они жаловались на глупость читающей публики. Но читатели продолжали расхватывать книги практически с печатного станка — в Великобритании, США, Японии, Германии; мужчины, женщины, старые, молодые, черные, белые, любой ориентации. Почти в любой точке земного шара у любого нашлись бы знакомые, прямо сейчас читающие трилогию о девушке с татуировкой дракона.

В книжном мире такое бывает не слишком часто. Феномен вроде Ларссона появляется раз в год, а может, и того реже. После Ларссона подобный оглушительный успех снискала Э. Л. Джеймс с книгой «Пятьдесят оттенков серого»⁴ — причем, в отличие от Ларссона, она была жива и могла принять участие в

рекламном турне. Ларссон же умер, не дождавшись публикации. Его трилогия достигла такого огромного объема продаж вообще без участия автора. Непостижимо. Непредсказуемо. Просто аномалия какая-то.

Давайте посмотрим на цифры. В американском штате Делавэр есть компания Bowker — глобальный поставщик библиографической информации и единственный источник идентификационных номеров (ISBN) для книг, издающихся в США. По данным Bowker, каждый год в США выходит приблизительно 50–55 тысяч новых наименований художественной литературы. Учитывая, что все больше произведений публикуют сами авторы в электронном виде без получения ISBN, это число занижено. За год в список бестселлеров газеты New York Times попадает примерно 200–220 книг. Даже если использовать заниженную общую цифру, это меньше чем 0,5 % от всех опубликованных. Из этой половины процента лишь немногие задерживаются в списке на несколько недель («двузначные», как называют их профессионалы книжного дела). И только малая доля авторов умудряется продержаться в списке десять недель или больше, а из их числа, вероятно, лишь у трех или четырех одна книга достигает миллионных тиражей в США в течение года. Но почему именно эта, а не другая?

Считается, что беллетристу, желающему завоевать читательскую аудиторию, нужно владеть определенными навыками: он должен правильно

построить сюжет, убедительно обрисовать характеры и при этом грамотно писать. Очень успешные авторы пошли дальше: они зорко подмечают все состояния человеческого духа, умеют лихо закрутить сюжет не в ущерб правдоподобию и, наконец, редко, но метко пользуются точкой с запятой. Все это — признаки хорошего писателя, а хороший писатель, если он готов приложить достаточно усилий, в конце концов найдет своего читателя. Но мы говорим о всеобъемлющем успехе, когда сотни тысяч разом читают одно и то же — этот триллер, а не тот, этого кандидата на Пулитцеровскую премию, а не какого-либо другого. Если оставить в стороне вмешательство Опры Уинфри, это значит, что есть какой-то неучтенный фактор — будто на нашего автора просыпалась с неба щепотка звездной пыли, такой тонкой, что ее не распознать никакими приборами. Внезапный и необъяснимый успех таких книг, как трилогия о девушке с татуировкой дракона, «Пятьдесят оттенков серого», «Прислуга»⁵, «Исчезнувшая»⁶, «Код да Винчи»⁷, считается большой удачей автора, но при этом случайной, как выигрыш в лотерею.

Кстати сказать, слово «бестселлер» изначально было жаргонным словечком издательской отрасли. Оно относительно новое — впервые вошло в словари английского языка в конце XIX века, примерно тогда же, когда появились первые списки книг, лидирующих по количеству проданных экземпляров. Это слово по идее должно быть нейтральным, но за свою жизнь оно обросло определенными коннотациями, отчасти

вводящими в заблуждение. Литературный журнал *Bookman* начал публиковать ежемесячные отчеты по продажам книг в 1891 году в Лондоне и в 1895-м — в Нью-Йорке, после появления Закона о международном авторском праве (1891), осложнившего распространение дешевых пиратских копий британских романов в США. До этого вести статистику по продажам было невозможно. Списки выходили во всех крупных городах и обычно включали в себя шесть книг, которые лучше всего продавались за истекший месяц. С самого начала эти списки выполняли две функции, невиданные ранее в книжном мире: во-первых, они составлялись исключительно на основе количества проданных экземпляров, а во-вторых, служили рекомендацией упомянутых книг для читателей. Эти рекомендации были основаны не на мнениях горстки критиков или рекламе издателей, а на выборе точно таких же читателей, обычных людей. Покупка книги была и остается единственным способом за нее проголосовать. Таким образом, термин «бестселлер» не несет никакой информации о качестве и тематике книги и не является синонимом выражений «жанровая проза» или «популярное чтение». Некоторые деятели литературного истеблишмента действительно употребляют это слово как оскорбление: они считают, что коллективный вкус читательской аудитории равняется только на плохую литературу, но объективные данные свидетельствуют об обратном. В числе бестселлеров были как книги, удостоенные

Пулитцеровской премии, и великие американские романы, так и произведения, рассчитанные на массового читателя. Среди авторов можно найти не только Майкла Коннелли с Дебби Макомбер, но и Тони Моррисон с Маргарет Этвуд. Именно поэтому список бестселлеров — достойное внимания, интересное и динамичное культурное явление.

Ясно, что, написав такую книгу, автор внакладе не останется. Не останутся без прибыли и агенты, обнаружившие ее, и редакторы. И книготорговцы тоже — ведь они держатся на плаву и не уходят из бизнеса лишь благодаря немногим хорошо продающимся произведениям.

Конечно, мы сейчас говорим о материальной выгоде. Допустим, вы давно рассказывали друзьям, что у вас внутри сидит роман и просится наружу, — и вот вы в самом деле выплеснули его на бумагу и получили за это семи-или даже восьмизначный аванс. Такое бывает — с немногими авторами, но все же бывает. Можно окружать богемным флером фигуру бедного писателя, скрипящего карандашиком в блокнотике, но куда приятней думать о том, как твоя книга появится в каждой квартире — на полке или на тумбочке у кровати — и на iPad или Kindle у каждого пассажира в каждом вагоне метро! И все это — на многих языках, по всему миру!

Такой блеск, такое драматическое восхождение к славе характерны для немногих «книг года». В книжном мире они играют роль дорогих особняков, шикарных машин и бриллиантовых диадем. Если

автор попал в список и умудрился там зацепиться, его ждут почет, уважение, зависть и злые слова. Автора могут пригласить в жюри литературного конкурса или давать ему на рецензию другие книги. Возможно, у него купят права на экранизацию. О нем будут говорить.

Правда, было бы здорово, если бы такой успех можно было предсказать?

Белые лебеди

Смелая гипотеза авторов сего труда состоит в том, что романы попадают в списки бестселлеров New York Times отнюдь не случайным образом и что пути книжного рынка вовсе не так неисповедимы, как принято считать. Бестселлеры, независимо от жанра, похожи друг на друга по большому количеству параметров (впрочем, не бросающихся в глаза), и, изучая эти параметры, можно лучше понять, что мы читаем и почему. Более того, можно разработать алгоритмы для выявления носителей «ДНК бестселлера» среди новых, еще не опубликованных книг.

У профессионалов книжной отрасли бытует мнение, что успех полностью определяется громким именем, затратами на рекламу или широкомасштабными мероприятиями по продвижению книги. Конечно, все это имеет значение, но наши результаты доказывают: далеко не все определяется шумихой, поднятой вокруг книги. Этот вывод должен обрадовать писателей, которые трудятся, оттачивая свое мастерство. Результаты

нашей пятилетней работы дают основания предположить, что бестселлером книгу делают правильные слова, расположенные в правильном порядке. Это значит, что даже в самой интересной истории, связанной со списком бестселлеров New York Times, речь идет исключительно об авторской рукописи как таковой — только о черных чернилах на белой бумаге, и больше ни о чем.

Используя компьютерную систему, которая читает книги, распознает в них определенные характеристики и просеивает тысячи таких характеристик в тысячах текстов, мы открыли, что существуют удивительные сочетания параметров, свойственные книгам, которые с наибольшей вероятностью будут пользоваться успехом на рынке. Эти параметры могут многое поведать о нас, читателях, и о том, как мы читаем. На этих страницах мы расскажем, как и зачем мы построили такую модель и как она обнаружила, что в 80–90 % случаев очень легко различить бестселлер среди других книг в использованном нами корпусе текстов. Наши компьютеры выявили 80 % книг, побывавших в списке New York Times за последние 30 лет, и назвали их вероятными кандидатами в бестселлеры. Более того, наша система обрабатывала каждую книгу так, словно это новая, никому не известная рукопись, а затем не просто выносила один из двух вердиктов — «вероятно, бестселлер» или «скорее всего, нет», — но и вычисляла индекс, указывающий, каковы шансы книги обрести популярность. Эти индексы сами по себе чрезвычайно

интересны, и мы не только расскажем, как они устроены, но и объясним, что держит читателя как магнитом, заставляя переворачивать страницы одну за другой.

Рассмотрим индексы, вычисленные нашей системой. Она предсказала, что «Инферно»⁸ Дэна Брауна окажется бестселлером с вероятностью 95,7 %. «“Линкольн” для адвоката»⁹ Майкла Коннелли — с вероятностью 99,2 %. Обе побывали на первом месте в списке New York Times — то есть, пожалуй, на самой престижной позиции, какую может завоевать книга. Конечно, Браун и Коннелли — известнейшие авторы с прочным положением на литературном Олимпе. Но наша система не знает имен авторов и с той же уверенностью раздает оценки трудам неизвестных писателей. Например, «Вязание по пятницам»¹⁰, дебютный роман Кейт Джейкобс, получил оценку 98,9 %. «Счастливые девочки не умирают»¹¹, совсем другой по характеру дебютный роман Джессики Нолл¹², получил индекс успеха 99,9 % — исключительно на основе текста. Книги Джейкобс и Нолл попали в список бестселлеров и задержались там на много недель. «Марсианин»¹³ (еще до того, как Мэтт Деймон пообещал сыграть главную роль в экранизации) получил 93,4 %. Можно привести примеры из самых разных жанров: «Телефонный звонок с небес»¹⁴ (произведение из разряда «литература о духовности») Митча Элбома — 99,2 %; «Искусство поля»¹⁵, литературный дебют Чеда Харбаха, — 93,3 %;

«Обнаженная для тебя»¹⁶, эротический любовный роман Сильвии Дэй, — 91,2 %.

Эти индексы, измеряющие «бестселлерный потенциал» книги, приводят в восторг одних людей, возмущают других и у многих вызывают подозрение. Последних можно понять: эти индексы не укладываются в голове, они как будто ниспровергают существующий порядок вещей. Некоторым ветеранам книжной индустрии они кажутся абсурдными. Но они могут совершить переворот в книгоиздании, и уж точно благодаря им вы будете совсем по-другому воспринимать следующий бестселлер, который попадет вам в руки.

Здесь мы должны прояснить важный момент. Ни одна из книг, упоминающихся в данной работе, никогда никем не приобреталась на основе индексов, вычисленных нашей системой. Кроме того, за исключением данной публикации, мы никогда не доводили эти индексы до сведения какого-либо книжного агента или издательства. Следует также оговорить, что все полученные нами численные показатели специфичны для конкретного корпуса текстов, на котором проводилось исследование, — при составлении этого корпуса мы добивались максимального сходства с ассортиментом, который вы увидите, зайдя в любой крупный книжный магазин. Книжные агенты и издатели знают свое дело — они трудятся, наполняя полки магазинов книгами, и нам всегда есть что почитать. Иные деятели издательской нивы обладают чем-то вроде дара царя Мидаса —

обращать своим прикосновением рукописи в золото. Но не забывайте, что средняя доля бестселлеров среди опубликованных книг — меньше половины процента. Иными словами, чтобы наконец выиграть, приходится долго рисковать. Заметьте также, что год за годом в списках фигурируют одни и те же чрезвычайно успешные авторы, десятилетиями возглавляющие рейтинги. Стивену Кингу шестьдесят восемь лет. Джеймсу Паттерсону шестьдесят восемь лет. Даниэле Стил шестьдесят восемь лет. Конечно, поклонники рады каждой новой книге любимого автора, но это значит, что книжный мир еще не открыл следующее поколение писателей, которые так же стабильно будут поставлять бестселлеры на рынок следующие тридцать или сорок лет. Кроме того, хотя ежегодно тысячи рукописей отвергаются и тысячи публикуются, в 2014 году издателям так и не удалось найти бешено успешный бестселлер («Татуировка дракона», «Оттенки серого» и «Исчезнувшая» были хитами предшествующих годов). В 2012 году ни одно из опубликованных литературных произведений не впечатлило комитет по присуждению Пулитцеровской премии. Почему?

Если верить расхожему мнению, любой бестселлер — аномалия. Удачная флуктуация на рынке. Интересный мутант. Черный лебедь. Но если это правда, то, единожды найдя писателя, способного выдавать бестселлеры, надо вкладывать все деньги в него? Зачем рисковать миллионами, ставя их на никому не известного двадцатилетнего юношу вместо

Стивена Кинга? Ведь нет способа узнать, стоит ли новый автор всех тех денег, которые вы собираетесь в него вложить, и поможет ли он вам сорвать куш.

У жаргона книгоиздательской индустрии очень много общего с жаргоном игорного бизнеса, и это не случайно. На совещаниях по вопросам покупки рукописей часто употребляется выражение «поставить на» такого-то автора. Если два издательства наперебой торгуются за одну и ту же книгу, одно из них может «пойти ва-банк» и потратить на нее весь свой сезонный бюджет. Это чрезвычайно интересное занятие, и риск, на который идут издательства, безусловно, обоснован, но все же это — азартная игра, практически как в казино. Прежде чем рукопись «Гарри Поттера»¹⁷ приняла в издательстве Bloomsbury, Дж.К. Роулинг получила 12 отказов в других местах и совет «не бросать основную работу». Сегодня стоимость бренда «Гарри Поттер» оценивается в сумму около 15 миллиардов долларов. Джону Гришэму отказали не менее 16 разных издательств; с тех пор Гришэм создал более двенадцати бестселлеров года, опередивших по уровню продаж все остальные книги ¹⁸. Джеймс Паттерсон несколько раз пытался опубликовать свои рукописи и всюду встречал отказы; в 2010 году три его книги разошлись общим тиражом 3,5 миллиона. Кэтрин Стокетт получила отказы от шестидесяти агентов, прежде чем нашла одного, который согласился представлять ее «Прислугу»; впоследствии этот роман пробыл в списке бестселлеров New York

Times на протяжении ста недель. Без сомнения, множество подобных книг сейчас пылится в отвалах так называемого «самотёка» в конторах издателей, которыми усеяны Нью-Йорк и Лондон.

Если вы имеете хотя бы косвенное отношение к книжному миру, наверняка у вас есть знакомый или знакомый знакомого, который годами вставал в четыре часа утра, чтобы успеть написать хоть немножко до работы. У него был абсолютно беспрюжный сюжет, он чувствовал, что музы невидимо реют над его головой... А потом он разослал рукопись во все издательства на Манхэттене и не получил в ответ ничего, кроме стандартных отписок-отказов.

Эти знакомые знакомых вовсе не одиноки. Один издатель, прочитав рукопись «Шпиона, пришедшего с холода»¹⁹, заявил Джону ле Карре, что как писатель он безнадежен. «Повелитель мух»²⁰ Уильяма Голдинга получил 21 отказ. Ознакомившись с рукописью Джека Керуака «В дороге»²¹, ныне культовой книгой поколения, издатель заявил: «Я вообще не понял, о чем это». Урсуле ле Гуин сказали, что ее «невозможно читать». Впоследствии «совершенно нечитабельная» книга ле Гуин получила две крупные литературные награды. Даже «Скотный двор»²² Оруэлла сочли негодным к публикации — причем не кто иной, как Т. С. Элиот. Великий поэт решил, что эта выдающаяся политическая аллегория, актуальная для всех времен и народов, «недостаточно убедительна».

Публиковать или не публиковать? Вот в чем вопрос. Чтобы предсказать книге большой успех, нужно проанализировать душевные движения и чувствительные места сотен тысяч человек. Это непростая задача, и часто отказы выглядят вполне обоснованными. Например, мы опросили американских издателей, отвергнувших «Девушку с татуировкой дракона», и они сказали, что, по их мнению, американской публике неинтересны все эти детали шведской политики. Лисбет Саландер, на их взгляд, слишком агрессивна и непредсказуема для центрального женского персонажа. Они сочли, что широкие массы читателей отпугнет чудовищная сцена анального изнасилования и Лисбет-мстительница, вооруженная татуировочными иглами. Все эти соображения кажутся весьма разумными.

Поэтому неудивительно, что издатели в минуту откровенности заявляют: пророчить книге успех — все равно что определять направление ветра, послунявив палец, или смотреть в хрустальный шар, который, похоже, спрятан под столом у самых удачливых агентов и издателей. За исключением случаев, когда автор уже популярен — например Нора Робертс или Джеймс Паттерсон, — делать такие прогнозы — все равно что стрелять в цель с завязанными глазами. Иногда помогают обстоятельства: например, если автор — голливудская звезда, а книга посвящена ее сексуальным похождениям. Но бывало и так, что дело казалось верным, за большим авансом следовали огромные

тиражи — и их приходилось пускать под нож. Вкусы публики переменчивы.

Конечно, все агенты и издатели стараются понять, что движет коммерческим книжным рынком — в области чтения для масс, на которую ориентирована, к примеру, писательница Патрисия Корнуэлл, ветеран франшизы в литературе, или в более скромном, но все же удовлетворительном масштабе тиражей, характерных для успешных современных литераторов. Известен анекдот о том, как (ныне бывшего) генерального директора одного из крупнейших нью-йоркских издательств попросили предсказать заглавие книги, которая точно станет бестселлером. «Собака врача президента Линкольна», — ответил он. И впрямь, имя всеми уважаемого президента в сочетании со словами «врач» (наша одержимость собственным здоровьем) и «собака» (любимое домашнее животное американцев) — верный рецепт успеха.

Это, конечно, шутка, но, оказывается, под таким названием вышла даже не одна, а две книги! И обе провалились. Преподаватель литературы и писатель Джон Сазерленд, выпустивший два исследования на тему бестселлеров, завершил одно из них словами: «В конечном счете бестселлер отличается от всех остальных книг тем, что он лучше продается. И больше ничем». Далее он добавляет, что «искать какие-то значимые закономерности, тенденции, сходство [между книгами, завоевавшими популярность] бессмысленно, а результаты этого

поиска повергают в недоумение». Вердикт Сазерленда кажется обоснованным, справедливым и окончательным. Точнее, казался таковым до появления компьютеров, которые научились читать тексты и вычислять секретный рецепт, позволяющий попасть в список бестселлеров New York Times.

Из любви к книгам

Вернемся к писателям, которые получили множество отказов, но в конце концов прославились. Наша система предсказала Дж.К. Роулинг успех с вероятностью 95 %. Джону Гришэму — с вероятностью 94 %. Паттерсону — 99,9 %. История позволила нам убедиться в точности этих предсказаний. А вот с книгой Кэтрин Стокетт «Прислуга» наша система ошиблась. «Прислуга» оказалась в числе примерно 15 % книг, которые повергли компьютер в недоумение. Он дал ей лишь 50 % вероятности стать бестселлером. В следующих главах мы расскажем о сложностях, с которыми столкнулись, моделируя работу редактора с помощью компьютера. Пока достаточно будет сказать, что система проводит чрезвычайно глубокий анализ; прочитав книгу Стокетт, она заключила, что стиль в целом хорош и понравится американским читателям, что тематика тоже в целом хороша, но использование эмоционально нагруженного языка и в особенности глаголов отличает «Прислугу» от книг, с большой вероятностью попадающих в список бестселлеров. Речь идет о книге, которая привлекла внимание рецензентов тем, что белая писательница активно

использовала имитацию диалекта чернокожих персонажей. Мнения рецензентов о правомочности такого авторского решения разделились; и оказалось, что наша система в точности предсказала это разделение во всех источниках, от New York Times до сайта goodreads.com.

Вероятно, вы спросите: зачем пытаться с помощью компьютера смоделировать работу, которую и так выполняют хорошие редакторы? Возможно, будь наша система уже внедрена, творения Дж.К. Роулинг опубликовали бы раньше. Возможно, Джон Гришэм получил бы аванс побольше за свою первую книгу «Пора убивать»²³. Но в конце концов эти писатели добились славы. Редакторы были не уверены в успехе «Прислуги». Наша система — тоже. Какая тогда от нее польза?

Дело в том, что мы ищем компоненты успеха не только для того, чтобы получить от этого материальную выгоду. Конечно, очень интересно, что наша система выделила из общей массы Дж.К. Роулинг, Лиану Мориарти (вероятность успеха 99,6 %) или Джонатана Франзена (98,5 %). Мы считаем, что публичное обсуждение сотрудничества человека и компьютера важно, особенно в том, что касается творческой деятельности. Но поиски новых подающих надежды рукописей в пошатнувшейся книгоиздательской отрасли проводятся с целью не только удержать эту отрасль на плаву, но и поддержать ее разнообразие. Безусловно, нами движет желание выявить и объяснить скрытые

закономерности человеческой культуры. Но, если подходить к делу более практично, мы хотим помочь издательствам в поиске новых авторов, чтобы направить часть бюджета, отведенную Кингу, Паттерсону или Стил, на молодых писателей, которые придут им на смену. Мы хотим, чтобы писатели — как опытные, так и молодые — узнали больше о своем ремесле и получили информационную поддержку. Мы хотим, чтобы книги людей, не имеющих нужных контактов в Нью-Йорке, нашли своего читателя. Нашей системе все равно, публиковался ли этот автор раньше, есть ли у него диплом бакалавра искусств, мужчина это или женщина, латиноамериканского или китайского происхождения, красив автор или некрасив и сколько ему лет — двадцать пять или семьдесят. Поэтому наша работа еще и уравнивает шансы на успех. Может быть, в один прекрасный день знакомый вашего знакомого отправит рукопись в издательство, получит индекс 80 % и ему выдадут хороший аванс, так что он сможет наконец уволиться с работы, и ему больше не нужно будет вскакивать в четыре часа утра, чтобы выкроить время для писательских занятий.

Когда мы пишем о романах, фигурирующих в самом знаменитом и уважаемом списке — еженедельном списке бестселлеров *New York Times*, — мы тем самым кричим во все горло, призывая неравнодушных читателей — как профессионалов книжного мира, так и любителей — поговорить о книгах для масс ²⁴. Над бестселлерами чаще

потешаются, чем рассматривают как серьезные литературные произведения или хотя бы плоды писательского мастерства. Однако, игнорируя их, мы упускаем большую часть современной культуры и истории чтения. Помимо того что авторы бестселлеров приносят в экономику миллионы долларов, их заслуга — в том, что они заставляют нас *читать*. Они заставляют нас воображать, обсуждать, думать, чувствовать, сопереживать другим. Они помогают фантазировать, подглядывать за чужой жизнью, бежать от действительности. Романисты из списка *New York Times* дают пищу для разговоров и споров о литературе по всей стране — в барах, в поездах, дома за ужином. Именно на этих писателей смотрят, желая узнать, куда движется наша литература. Именно к ним обращаются, чтобы лучше понять жизнь. К ним прибегают, чтобы развить вкус, выработать мнения и научиться их выражать. Мы будем только рады, если сможем рассказать читателям что-то новое об их любимых писателях.

Вероятно, вы уже поняли: авторы, то есть мы, до того страстно преданы книгам и чтению, что отдали в общей сложности пятьдесят лет изучению и преподаванию нарратива и еще несколько — закупкам и продажам книг для крупнейших издателей страны. Мы выступали в роли наставников при написании романов и отстаивали свое право любить и ненавидеть разные (а иногда — одни и те же) произведения. Мы расхваливали издателям книги самых разных жанров, стараясь добиться их публикации. Мы (иногда втайне)

помогали своим лучшим студентам и своим друзьям, мечтающим о карьере литератора, писать письма их родителям, супругам и будущим издателям, объясняя, почему автор письма просто обязан был бросить устроенную жизнь, учебу на врача — и отправиться на поиски колдовского зелья, сулящего то экстаз, то депрессию, — на поиски жизни в словах и историях. Не будет преувеличением сказать, что мы «купились» на ту свободу, то высшее (высочайшее) образование, которое получаешь, читая и создавая книги. Мы в первую очередь читатели, а потом уже писатели. Учитывая наше преклонение перед книгами, нет ничего удивительного в том, что мы обратились за помощью к компьютеру.

Откуда мы взялись

Вероятно, больше всего открытию «бестселлерометра» (прозвище нашей системы) удивились мы сами. Честно говоря, наши исследования начались лишь потому, что нас «нутром» потянуло в эту сторону. Четыре года ежедневного совместного труда принесли результаты, каких ни один из нас не ожидал, несмотря на наш разнообразный опыт: у Джоди — в книгоиздании и современной литературе, а у Мэтта — в литературе и модной нынче дисциплине, именуемой «цифровые гуманитарные науки».

Все началось с того, что Джоди ушла с должности редактора-рецензента в издательстве Penguin Books, чтобы пойти в аспирантуру по английскому языку в Стэнфордском университете. От работы в книжной

индустрии у Джоди остался жгучий вопрос, на который она так и не получила ответа: отчего романы становятся бестселлерами? А также связанные с этим скрытые, еще более интересные вопросы: почему читатели читают? Какую роль выполняет чтение в современной культуре?

Когда Джоди только пришла работать в Penguin, ее определили в отдел продаж. Иногда в обеденный перерыв она заглядывала в ближайший книжный магазин, чтобы проверить, выполняет ли он свои обязательства по продвижению книг — действительно ли те размещены на бойких местах в соответствии с суммами, потраченными на маркетинг. Очень часто, и это отнюдь не секрет, издатели платят за то, чтобы их «хиты» выставлялись на виду. Руководство магазина берет деньги и обязуется расположить книгу, например, в первом ряду на первом столе или на полке так, чтобы обложка была полностью видна и смотрела прямо на покупателя. Считается, что такое стратегическое размещение способствует продажам. В то время по спискам бестселлеров шествовал победным маршем «Код да Винчи». Неделю за неделей Джоди видела большую синюю цифру «1» и убеждалась в том, что роман Дэна Брауна продолжает покорять мир.

Через несколько месяцев стало очевидно: сколько бы издатели ни платили за размещение клонов Дэна Брауна на видных местах, им не светило попасть в одну лигу с «Кодом да Винчи». Его феноменальный успех объяснялся не суммами, затраченными на

рекламу и маркетинг, а чем-то другим. Никакие расходы не гарантируют столь длительного воздействия на умы читателей всего мира и тем более не обеспечат тиража 80 миллионов. Такой успех не может объясняться только рекламной шумихой. Значит, есть что-то кроме маркетинга, связанное именно с этими словами на этих страницах.

Конечно, глупо было бы утверждать, что маркетинг и реклама совсем не влияют на продажи. Разумеется, влияют. Должна быть некая корреляция, объясняющая тот факт, что 80 % бестселлеров принадлежат пяти крупнейшим издательствам, обладающим значительными средствами на рекламу и маркетинг. Но было бы так же глупо утверждать, что любой доллар, затраченный на маркетинг, всегда дает одинаковую отдачу: слишком много мы видели и огромных сумм, потраченных безрезультатно, и публикаций силами автора, имевших бешеный успех. «Пятьдесят оттенков серого» сначала появилась в электронном виде, а также распространялась по модели «печать по требованию» в мягкой обложке — издательством, у которого вообще не было ни гроша на рекламу. Чтобы напечатать свою книгу «Хижина»²⁵, Уильям Пол Янг воспользовался кредитными карточками, а на маркетинг потратил всего 300 долларов — во столько обошлась ему разработка веб-сайта. На сегодня эта книга разошлась десятиллионным тиражом. Вот примеры других бестселлеров, непохожих на «Хижину», но также пришедших к успеху и похвале критиков через

нетрадиционные каналы распространения: экспериментальный онлайн-роман Марка Данилевского «Дом листьев»²⁶ и книга Криса Уэра «Джимми Корриган, самый умный ребенок на Земле»²⁷, первоначально опубликованная на средства автора, а ныне один из самых популярных «графических романов» новой волны. Примеров таких можно набрать много — достаточно, чтобы понять: на вопрос «почему одни книги читают миллионы, а другие в лучшем случае десятки людей?» ответить «маркетинг» — значит в лучшем случае ответить наугад.

Джоди пришла со своим исследовательским вопросом к Мэтту, который в то время был лектором в Стэнфордском университете и сооснователем Стэнфордской литературной мастерской. Объединив усилия, Джоди и Мэтт поняли, что существует другой ответ, не просто «маркетинг». В 2008 году Мэтт как раз закончил свою часть работы над исследованием авторского стиля священного текста «Книга Мормона» (The Book of Mormon) с помощью компьютера. Эта работа вызвала много споров. Результаты компьютерного анализа авторского стиля «Книги» дали возможность предположить, что ее действительно писал не один человек, то есть выдвигавшиеся ранее гипотезы, возможно, были верны. Исследование также подкрепило одну определенную теорию происхождения «Книги», которую церковь мормонов отвергла как ложную. Результаты анализа не давали сделать однозначное

заклучение, но отклики на статью, в том числе весьма интересное опровержение, сделанное мормонскими богословами из Университета имени Бригама Янга, показали, каким революционным орудием может быть компьютерный анализ текста.

Эта работа над определением авторства и «стилеметрией» убедила Мэтта в том, что компьютеры видят многое и помогают нам обнаружить в тексте то, чего мы обычно не замечаем. Дальнейшие исследования позволили Мэтту создать компьютерную программу, которая с точностью 82 % определяла пол автора всего лишь по его или ее манере использовать самые простые слова английского языка, такие как *the*²⁸ и *of*²⁹. Мэтт не первым заметил, что у авторов-мужчин и авторов-женщин — разные стилевые привычки, но в своей работе он сосредоточился исключительно на романах XIX века. Затем он обнаружил, что, опираясь лишь на использование слова *the*, компьютер может достаточно уверенно определить, английский это писатель или американский.

Реакцию Джоди можно примерно описать словами «Ну и что?». Конечно, очень интересно, что компьютер может отличить британца от американца, но это — искусственная задача, решение которой никому не интересно. Чтобы убедиться в могуществе компьютерного метода, Джоди хотела увидеть, как с помощью машин решаются настоящие литературоведческие задачи. Мэтт же, в свою очередь, был не в восторге от того, что Джоди заиклена на

современных бестселлерах. Он считал их одноразовыми поделками — прочел и забыл. Его надо было убедить, что в них скрыто золото, которое оправдает труды старателя.

Все это было несколько лет назад. С тех пор мы объединили свои силы и бросили их на доказательство следующей гипотезы: все бестселлеры характеризуются определенным набором скрытых сигналов, неким тайным кодом. Вместо того чтобы угадывать, какая книга окажется популярной, мы решили довериться тому, что читатели уже поняли, сами того не заметив. Список бестселлеров лишь на первый взгляд кажется случайной подборкой непохожих друг на друга книг — на самом деле это еженедельно обновляемый набор сигналов, на зов которых идет читатель, и этот набор постоянно уточняется посредством коллективного голосования. Учит ли чему-нибудь глас читающего народа? Смогут ли наши компьютеры выделить сигнал из этого шума? Есть ли у этих броских обложек — будь то высокоинтеллектуальные романы из университетской программы или чтиво для пляжа — что-нибудь общее?

Если ответ «да», то мы сможем узнать кое-что об анатомии успеха. Мы сможем даже опровергнуть распространенную в книжном мире теорию и создать алгоритм, позволяющий предсказать, станет ли та или иная книга бестселлером.

Итак, мы начали обучать компьютер чтению.

Машинное чтение

Конечно, компьютеры на самом деле не умеют читать, во всяком случае — в том смысле, в котором вы читаете эту страницу. Но можно также сказать, что они умеют читать в том смысле, в каком они умеют делать все остальное — они «считывают» текст, то есть принимают его в качестве входной информации. Затем они проводят разбор, то есть разбивают текст на единицы, которые мы, люди, считаем осмысленными: буквы, запятые, слова, предложения, главы и т. д. На этой стадии компьютер в значительной мере подражает человеку, и чем совершенней была обучающая программа, тем точнее подражание. Разница между читателем-человеком и читателем-машиной заключается в том, что человек понимает значение читаемого текста, а компьютер — нет. Ирония здесь, однако, в том, что компьютерный способ чтения позволяет нам взглянуть в текст пристальней любого самого опытного литературного критика. Все потому, что компьютеры — мастера различать детали в таком мелком масштабе, какой человеку и не снился.

Вспомним вопрос, с которого началось наше исследование. Можно ли предсказать заранее, что та или иная книга станет бестселлером? Чтобы предсказывать, нужно уметь обнаруживать закономерности. Если вы не умеете гадать на картах, то вся ваша надежда — на способность распознавать определенные характеристики текста. Обычно такими вещами занимаются литературные критики и ученые-литературоведы. Джозеф Кэмпбелл, великий

специалист по мифологии, всю жизнь читал мифы и легенды разных народов мира и научился видеть в них схожие мотивы. Он был мастером по распознаванию повторяющихся узоров. Но даже если человек всецело предан своему делу, объем текстов, которые он может переработать, конечен, и столь же ограниченно внимание, с которым он исследует каждую отдельную книгу. Следовательно, в обоих направлениях мы сталкиваемся с ограничениями масштаба и как будто вынуждены одним глазом смотреть в микроскоп, а другим — в телескоп.

Еще один ученый, обладающий завидным трудолюбием, — Кристофер Букер — потратил тридцать лет на чтение сотен книг, чтобы выдвинуть следующую теорию: все литературные произведения и вообще все истории соответствуют семи основным сюжетам. Допустим, он трудился сорок лет и прочел тысячу книг. Допустим также, что из прочитанного он запомнил гораздо больше, чем средний человек. Но кластер компьютеров, обученных надлежащим образом, может прочитать несколько тысяч романов и проанализировать все эти тексты по тысяче параметров примерно за один день. При этом компьютеры обладают невероятной способностью замечать то, что мы, люди, принимаем как должное или вообще игнорируем.

Вот лишь один пример. Как читатели — особенно читатели, обученные пристально вглядываться в текст, — мы отмечаем прилагательные, которые использует писатель. Но, скорее всего, не обратим внимания на

соотношение числа существительных и прилагательных, то есть на то, как часто автор использует прилагательное для описания существительного. А вот компьютер прекрасно приспособлен для того, чтобы замечать именно такие вещи — и они важны, поскольку определяют авторский стиль и манеру описаний. Компьютер может также вычислить это соотношение для данной книги и сравнить с тем же показателем у тысячи других книг. Если окажется, что у бестселлеров этот показатель в среднем выше (или ниже), значит, данный параметр имеет какое-то значение.

Предлагаю вам проделать небольшой эксперимент, когда будете в следующий раз выбирать книгу для чтения. Вместо того чтобы последовать рекомендации друга или выбрать книгу знакомого автора (или в любимом жанре), попробуйте взять недельный список бестселлеров NYT и прочитать их все по очереди. Можете это сделать вместе с книжным клубом, в котором состоите, или всей группой на курсе английской литературы. Если вы будете читать внимательно, то скоро уподобитесь нашим компьютерам и начнете замечать неожиданные параллели между литературными шедеврами и массовым чтивом, книгами «для мужчин» и «для женщин», Паттерсонами и пулитцеровскими лауреатами и т. п. Некоторые общие мотивы вас удивят. Например, вы зададитесь вопросом: почему так много книг, в которых главной героине двадцать восемь лет? Имеет ли это какое-то значение? Вы