

# Читат между строк:

Контент-анализ  
для точных выводов  
и сильных решений.

ИИ

Фёдор Северный

Фёдор Северный

**Читать между строк:  
Контент-анализ для точных  
выводов и сильных решений**

«Автор»

2026

## **Северный Ф.**

Читать между строк: Контент-анализ для точных выводов и сильных решений / Ф. Северный — «Автор», 2026

Десятки отзывов и жалоб — а руководству нужен ясный ответ: что происходит и какое решение принять? Частая тема ещё не доказывает причину, а цифра без контекста может ввести в заблуждение. «Читать между строк» показывает, как превратить тексты в проверяемые выводы: сформулировать исследовательский вопрос, собрать корпус материалов, обосновать выборку, определить единицы анализа и создать кодировочную книгу. Вы научитесь проводить пилотаж, контролировать кодирование и его надёжность, видеть закономерности в числах и смыслы в повторяющихся историях. Книга объясняет, как учитывать альтернативные объяснения, обозначать границы выводов и ответственно использовать автоматизацию и ИИ. Для исследователей, аналитиков, специалистов по коммуникациям и руководителей, которым нужны решения на основе данных, а не впечатлений. Вы получите систему, которая поможет сделать анализ прозрачным, обосновать выводы и превратить неопределённость в следующий проверяемый шаг.

## Содержание

|                                                             |    |
|-------------------------------------------------------------|----|
| Контент-анализ: что он может и чего не обещает              | 6  |
| От практической задачи к точному исследовательскому вопросу | 11 |
| Выбор источников и сбор корпуса материалов                  | 17 |
| Как сформировать выборку и не исказить картину              | 23 |
| Конец ознакомительного фрагмента.                           | 26 |

**Фёдор Северный**  
**Читать между строк: Контент-анализ**  
**для ТОЧНЫХ ВЫВОДОВ и СИЛЬНЫХ решений**  
**Читать между строк: Контент-анализ для точных выводов и сильных решений**  
*Фёдор Северный*

## Контент-анализ: что он может и чего не обещает

В 8:17 утра на экране переговорной уже светились жалобы. «Водитель даже не посмотрел в мою сторону». «На просьбу открыть заднюю дверь ответил так, будто я ему должен». «Опять это отношение к людям». Сообщения были короткими, разными по тону и похожими в главном: пассажиры всё чаще говорили о неуважении.

— Нам нужно понять, кто за это отвечает, — сказал директор. — Водители? Диспетчеры? Руководители маршрутов? Давайте не растягивать.

Исследовательница Мира не стала спорить с важностью вопроса. Она лишь уточнила:

— Вы хотите выяснить, что пассажиры пишут, или кто виноват?

В комнате стало тихо. Казалось, это один и тот же вопрос. Но именно в этом месте начинается работа контент-анализа — и заканчиваются обещания, которые ему не следует давать.

### Фонарь, а не детектор

Контент-анализ — это систематическое изучение содержания сообщений: текстов, изображений, аудио, публикаций, документов. Исследователь заранее определяет, какие материалы рассматривает, что именно в них отмечает и по каким правилам сравнивает. Метод помогает увидеть повторяющиеся темы, способы описания событий, различия между группами сообщений и изменения во времени.

Ключевое слово здесь — **систематическое**.

Если прочесть три особенно раздражённых комментария, можно составить впечатление о настроении пассажиров. Если собрать подходящие сообщения за определённый период, определить правила отбора, одинаково разметить материалы и проверить, как часто встречаются разные темы, получится исследование. Оно не становится безошибочным оттого, что в нём много таблиц. Но его ход можно объяснить и, в разумных пределах, повторить.

Контент-анализ — фонарь. Он освещает то, что попало в его луч, и помогает рассмотреть очертания. Он не показывает автоматически, что происходило за пределами освещённого участка, не читает мысли авторов и не назначает виновных.

Мира предложила команде начать не с поиска ответственного, а с более скромного вопроса: **какие именно переживания и ситуации пассажиры описывают, когда говорят о неуважении?**

Формулировка не уклонялась от проблемы. Напротив, она превращала смутное слово в предмет исследования.

### От слова к наблюдаемому признаку

«Неуважение» звучит ясно, пока не нужно решить, какие сообщения считать относящимися к этой теме. Один пассажир назовёт неуважением резкий ответ. Другой — отсутствие ответа. Третий использует это слово для любой неприятной ситуации: переполненного автобуса, сбой оплаты или задержки.

Чтобы анализировать тексты, нужно отличать понятие, которое волнует исследователя, от признаков, которые можно найти в материале. Например, команда может отмечать, упоминает ли автор:

- резкий или пренебрежительный тон;
- отсутствие ответа на обращение;
- отказ помочь или объяснить ситуацию;
- нарушение правил доступности;
- опасное или резкое поведение;
- задержку или сбой, описанные как пренебрежение к пассажирам;
- оценочное слово «неуважение» без конкретного эпизода.

Последний пункт особенно важен. Слово «неуважение» действительно присутствует в сообщении, но само по себе не раскрывает, какое событие за ним стоит. А иногда автор подробно описывает ситуацию, но вовсе не употребляет этого слова.

Исследовательница предложила команде разбирать каждый комментарий на три уровня.

**Что текст позволяет увидеть**

**Что можно осторожно предположить**

**Чего по этим данным утверждать нельзя**

«Попросила водителя открыть заднюю дверь. Он промолчал, двери закрылись». Упомянуты просьба, отсутствие ответа и закрытие дверей.

Пассажирка могла воспринимать молчание как игнорирование; сообщение может относиться к теме коммуникации или доступности.

Что водитель намеренно проигнорировал пассажирку; что он услышал просьбу; что так поступают водители этого маршрута в целом.

«На вопрос, почему автобус задерживается, водитель сказал: “Не знаю, отстаньте”». Приведена цитата с грубым выражением и задан вопрос о задержке.

Автор воспринимает ответ как резкий; случай можно отнести к описаниям тона общения.

Что цитата передана дословно; что причиной ответа была личная неприязнь; что именно водитель виноват в задержке.

«На остановке опять было унижительное отношение. Город совсем забыл о людях». Есть оценка «унижительное», но нет описания конкретного эпизода.

Автор может связывать накопленный опыт с более общей оценкой сервиса.

Какое именно событие произошло, кто участвовал и насколько часто подобные случаи происходят в действительности.

Эта таблица не заставляет исследователя отказаться от интерпретации. Она помогает не перепутать интерпретацию с наблюдением. В первой колонке — признаки, которые можно указать в самом сообщении. Во второй — объяснения, совместимые с текстом, но не подтверждённые им окончательно. В третьей — утверждения, для которых нужны другие данные.

Различие кажется очевидным, пока руководству не требуется ответ к обеду.

**Что мы считаем и с чем сравниваем**

Прежде чем считать сообщения, команда должна решить, какие именно сообщения попадут в исследование. Посты на официальной странице компании? Отзывы на городском портале? Упоминания перевозчика в открытых соцсетях? За неделю, за месяц или за год?

Нет нейтрального решения «взять всё». Платформы отличаются, люди пишут на них по разным поводам, а способы поиска могут пропускать часть публикаций. Жалоб на официальной странице может быть больше не потому, что там хуже сервис, а потому, что пассажиры знают, куда обратиться. Рост количества сообщений может отражать и реальный всплеск недовольства, и изменение модерации, и новый городской инфоповод, и то, что аккаунт компании стал популярнее.

Нужно определить и **единицу анализа**. Ею может быть отдельный пост, комментарий, фрагмент сообщения или описанный эпизод. Один длинный пост способен содержать три истории: о грубом ответе, задержке и неудобной пересадке. Если считать его одной жалобой, мы узнаем число публикаций, но не число упомянутых проблем. Если считать отдельные эпизоды, придётся установить правила, как их разделять.

Затем составляется схема кодирования — инструкция, по которой сообщения относят к темам. Например, что считать «отсутствием ответа»? Достаточно ли фразы «меня не слушали»? Нужно ли наличие описанного вопроса? Что делать с сарказмом? Как классифицировать сообщение, где пассажир одновременно жалуется на грубость и на опоздание?

Чем точнее правила, тем меньше исследование зависит от интуиции одного человека. Полезно попросить двух аналитиков независимо разметить часть корпуса, сравнить результаты и обсудить разногласия. Если они по-разному понимают категорию, проблема может быть не в невнимательности кодировщиков, а в неясном определении. Согласие между ними не доказывает, что найдено единственно верное толкование. Оно показывает, насколько последовательно применяются правила.

Для качественного анализа разногласие тоже бывает полезно: оно показывает, где текст многозначен, где не хватает контекста и какие категории слишком грубы. Цифры здесь не заменяют мышление — они дисциплинируют его.

### **Почему яркие комментарии — ещё не картина**

После совещания директор попросил показать «самые показательные» сообщения. Мира согласилась, но добавила: эти примеры помогут услышать конкретные голоса, а не измерить распространённость проблемы.

Несколько комментариев — это иллюстрации, а не автоматически репрезентативная картина. Яркие истории запоминаются лучше спокойных. Грубая цитата кажется убедительнее, чем десятки коротких сообщений «автобус приехал вовремя». Если выбирать примеры по эмоциональной силе, можно незаметно построить анализ вокруг того, что сильнее задело исследователя.

Контент-анализ способен ответить, например, **какие темы чаще встречаются в собранных публикациях или как меняется доля сообщений о тоне общения внутри этого корпуса**. Но даже точный подсчёт не означает, что мы знаем, как чувствуют себя все пассажиры. Люди, которые пишут публично, могут отличаться от тех, кто молчит, звонит в диспетчерскую или вообще не пользуется соцсетями.

Важно и то, что именно стоит в знаменателе. Если в одном месяце найдено 80 сообщений о грубости, а в другом — 100, это ещё не доказывает ухудшение. За второй месяц могло появиться вдвое больше всех сообщений. Тогда доля публикаций о грубом общении, возможно, даже снизилась. Имеет значение не только число упоминаний, но и объём корпуса, правила его формирования и сравнимость периодов.

Контент-анализ также отличается от опроса. Опрос помогает задавать людям вопросы напрямую: сталкивались ли они с определённой ситуацией, как её оценивают, что предпочли бы изменить. Но ответы зависят от формулировок, памяти, готовности участвовать и состава выборки. Публичные тексты, напротив, часто возникают без исследовательского вопроса, но исследователь не может уточнить у автора, что тот имел в виду. Эти методы не конкурируют: они освещают разные стороны проблемы.

А анализ причин поведения — ещё одна задача. Даже если публикации показывают рост жалоб на резкий тон, сами по себе они не объяснят, почему он возник. Изменились ли условия труда? Перераспределили ли маршруты? Появилось ли обучение сотрудников? Стало ли больше пассажиров? Чтобы проверять причинные объяснения, нужны дополнительные данные, подходящий дизайн исследования и, как правило, несколько источников свидетельств.

Текст может подсказать, **что стоит проверить дальше**. Он не всегда может показать, **что именно вызвало наблюдаемое явление**.

### **Как команда получила полезный ответ**

Мира предложила ограничить первый этап исследования публичными сообщениями за последние три месяца, где упоминались городской перевозчик и поездка. Команда описала

правила поиска и исключения материалов: например, дубликаты не считались новыми сообщениями, а объявления без рассказа о пассажирском опыте не включались.

Затем аналитики размечали не «виновников», а наблюдаемые темы: тон общения, отсутствие ответа, доступность, задержки, переполненность, технические сбои и общие оценки без конкретного эпизода. В одном сообщении можно было отметить несколько тем. Для каждой категории команда записала примеры и спорные случаи.

Результат оказался менее эффективным, чем список обвиняемых, зато полезнее для решений. В собранном корпусе значительную часть сообщений, называвших сервис неуважительным, составляли описания ситуаций, где пассажир не получал ответа или объяснения. Часть жалоб относилась к доступности транспорта, часть — к задержкам, а некоторые содержали общее недовольство без конкретного эпизода.

Это не означало, что сотрудники намеренно проявляли неуважение. Но стало понятно, что первоначальное слово скрывало несколько разных проблем. Для одних сообщений следовало изучить порядок коммуникации с пассажирами, для других — доступность и информирование о сбоях. Теперь компания могла сопоставить публикации с обращениями в диспетчерскую и данными о выполнении рейсов, а затем проверить, совпадают ли публичные жалобы с зарегистрированными случаями.

Контент-анализ не вынес приговор. Он помог точнее поставить следующие вопросы.

Так выглядит хорошая работа метода: не эффектное разоблачение, а более ясная карта местности. Иногда карта показывает, что дорога ведёт не туда, куда предполагало руководство. Это не недостаток исследования. Это и есть одна из его задач.

### **Практический маршрут: от тревоги к анализу**

Если перед вами уже есть массив текстов и хочется немедленно «понять, что происходит», пройдите несколько шагов.

**1. Сформулируйте наблюдаемую задачу.** Вместо «почему компания не уважает пассажиров?» спросите: «Какие ситуации пассажиры связывают с неуважительным отношением в публичных сообщениях за выбранный период?»

**2. Определите корпус.** Укажите источники, временные рамки, критерии включения и ограничения поиска. Не называйте собранные публикации «мнением пассажиров вообще», если это только открытые сообщения на двух площадках.

**3. Выберите единицу анализа.** Решите, что именно будете считать: публикации, отдельные эпизоды или упоминания тем. Объясните, как поступите с длинными и многотемными текстами.

**4. Опишите категории простыми правилами.** Для каждой категории запишите определение, подходящий пример и случай, который в неё не входит. Если правило нельзя применить к реальному сообщению, его нужно уточнить.

**5. Проверьте схему на небольшой части материалов.** Разметьте сообщения, обсудите расхождения, исправьте неясные категории и только затем переходите к основному массиву.

**6. Представьте результат с границами.** Сообщайте, что обнаружено в конкретном корпусе, как собраны данные и чего они не позволяют установить. Отдельно обозначьте гипотезы, которые требуют проверки другими методами.

Для небольшой практики выберите двадцать открытых комментариев о сервисе, удалите имена и другие личные данные и попробуйте выделить в них описанные события. Не начинайте с подсчёта слова «плохо». Спросите себя: какой эпизод описан, что сказано прямо, что я достраиваю по привычке? Затем попросите другого человека применить ваши правила к тем же материалам. Разница между разметками — не повод объявить кого-то неправым, а подсказка, где инструкция требует ясности. Даже учебный разбор стоит проводить бережно: пуб-

личность сообщения не отменяет обязанности не раскрывать лишние персональные сведения и не вырывать цитаты из контекста.

### **Перед тем как включить фонарь**

Контент-анализ особенно полезен, когда нужно упорядочить большой массив материалов, сравнить периоды или группы, обнаружить повторяющиеся способы описания проблемы, проследить, какие темы соседствуют в сообщениях. Он помогает перейти от «кажется, жалоб много» к проверяемому описанию: что именно встречается, в каких данных и по каким правилам это было замечено.

Но метод не превращает текст в прозрачное окно в человеческий разум. Он не устанавливает намерения автора по одной фразе, не представляет автоматически тех, кто молчит, и не доказывает причинность одним подсчётом. Чем серьёзнее вывод, тем важнее контекст и дополнительные свидетельства.

Перед началом анализа задайте себе три вопроса:

- **На какой вопрос отвечает метод?** Что я хочу описать в содержании сообщений: темы, формулировки, оценки, различия или изменения?

- **Какие данные для этого нужны?** Где, за какой период и по каким правилам я соберу материалы? Что станет единицей анализа?

- **Где проходит граница вывода?** Что можно будет показать непосредственно по текстам, что останется осторожной интерпретацией и какие утверждения потребуют опроса, наблюдений или других данных?

В начале совещания руководству нужен был виновный. К его концу у команды появился вопрос, на который действительно можно было ответить: какие повторяющиеся ситуации пассажиры описывают, когда говорят о неуважении, и как проверить эти наблюдения по другим источникам? Это не самый быстрый путь к уверенности. Зато он защищает от быстрой ошибки.

Любое исследование начинается не с таблицы и не с алгоритма, а с точного понимания того, что именно мы пытаемся узнать. Первый рабочий вопрос звучит так: **какие закономерности в публичных обсуждениях можно проверить систематически?** Ответ на него контент-анализ дает не по отдельной яркой цитате, а по совокупности материалов, изученных по прозрачным правилам. Он помогает увидеть повторяющиеся темы, формулировки и способы описания опыта — но сам по себе не доказывает, что наблюдаемое характерно для всех пассажиров, объясняет причины их мнений или устанавливает, что произошло на самом деле. Сила вывода зависит от того, какие материалы вошли в исследование, как определены категории и чем проверены результаты.

Поэтому задача исследователя — не обещать больше, чем позволяют данные, а ясно обозначить границу между тем, что обнаружено в текстах, и тем, что еще предстоит выяснить. Чтобы провести эту границу, нужно сначала превратить практическую проблему в вопрос, на который можно ответить с помощью конкретного материала и понятной процедуры. Именно с этого шага начинается надежное исследование.

## От практической задачи к точному исследовательскому вопросу

В понедельник в 10:17 на совещании уже спорили громче, чем работал проектор.

Новая система оплаты проезда запустилась три недели назад. В приложении можно было пополнить баланс, оплатить поездку и посмотреть историю списаний. Но в отзывах и обращениях пользователей всё чаще встречались слова «непонятно», «списали дважды», «никому не дозвониться». Директор по продукту считал, что дело в неудобном интерфейсе. Руководитель поддержки — что пассажиры не доверяют компании. Маркетинг винил неудачное объяснение запуска, а операционный директор напоминал, что «людям всегда что-нибудь не нравится».

На экране появилась презентация с заголовком: «Почему пользователи негативно воспринимают новую систему оплаты?»

— Давайте просто поймём, что им не нравится, — предложил кто-то.

Исследовательница Марина, до этого молча делавшая заметки, подняла глаза:

— А какое решение мы примем, когда поймём?

В комнате стало тише. Это был не эффектный вопрос. Но именно он остановил спор — и впервые за утро направил его в полезную сторону.

### **Спор о причинах и вопрос без границ**

Фраза «почему людям не нравится?» кажется естественной. Она звучит как начало исследования, но на деле часто скрывает сразу несколько нерешённых задач. Кто именно эти «люди»? Что считать «не нравится»? Где искать ответ — в отзывах, переписке со службой поддержки, интервью или статистике отказов? И главное: какое решение будет зависеть от результата?

Марина попросила команду не отвечать на эти вопросы вслух, а сначала отредактировать исходную формулировку. На проекторе появилось первое уточнение:

### **«Понять всё про отношение пользователей к новой оплате».**

— Всё? — переспросил аналитик. — Включая тарифную политику, дизайн кнопки и детские воспоминания о первом автобусе?

Посмеялись, но фраза была опасна не только своей широтой. «Понять всё» нельзя проверить на завершённость: даже после года работы всегда найдётся ещё один аспект, ещё одна группа, ещё один канал. Такой запрос приглашает собирать материалы без ясного конца. В результате команда может получить много данных и всё ещё не знать, какой вывод считать достаточным.

Марина вывела следующую версию:

### **«Насколько неудачно компания внедрила оплату?»**

Тут проблема была другой. «Неудачно» — оценка, а не наблюдаемый признак. В неё заранее помещён вердикт, хотя исследователям ещё предстоит выяснить, что именно происходило и как это описывают пользователи. Такая формулировка незаметно подталкивает искать подтверждения провалу: каждую жалобу записывать в доказательство, а нейтральные или положительные сообщения считать исключениями.

Третья версия появилась почти мгновенно:

### **«Как неудобный интерфейс вызывает негативные отзывы?»**

— А если люди жалуются не на интерфейс? — спросила Марина.

Фраза уже назначила виновника и даже нарисовала причинную связь: неудобство якобы вызывает негатив. Исследование в таком виде превращается в допрос с заранее известным подозреваемым. Команда будет замечать подтверждения собственной гипотезы и хуже видеть

альтернативы — например, сбои платежей, непонятные списания, недостаток информации или трудности с обращением в поддержку.

У всех трёх формулировок разные недостатки, но общий симптом один: вопрос пытается сразу перескочить от тревоги к объяснению. Чтобы сделать его исследуемым, нужно пройти несколько ступеней — от решения, которое предстоит принять, до признаков, которые можно обнаружить в материалах.

### **Редакторская лестница вопроса**

Марина нарисовала на доске пять ступеней. Не для красоты: каждая не давала перескочить через важное уточнение.

#### **1. Решение, которое нужно принять**

Сначала команда сформулировала не тему исследования, а практическую ситуацию.

Варианты решения были вполне конкретны:

- переработать отдельные шаги в приложении;
- проверить и улучшить коммуникацию о списаниях и возвратах;
- изменить маршрутизацию обращений в поддержку;
- отложить крупные изменения, пока не выяснится, где именно затруднение.

**Важно:** исследователь не обязан сам выбрать один из этих вариантов по одним текстам. Но нужно понимать, какие решения результаты должны поддержать. Иначе легко провести аккуратный анализ, который окажется любопытным, но никак не повлияет на работу.

Управленческая проблема звучит примерно так: «**Мы не знаем, на что направить ограниченный бюджет и внимание команды**». Это реальная проблема, но ещё не исследовательский вопрос. Ответ «нужно срочно переделать приложение» был бы рекомендацией. А задача анализа — собрать основания, которые помогут сравнить варианты.

#### **2. Информационный пробел**

Следующая ступень — назвать, чего именно не хватает для решения.

Команда уже знала, что отзывов с негативной оценкой стало больше. Но не знала, какие ситуации и темы за ними стоят: непонятный интерфейс, технические ошибки оплаты, вопросы доверия, трудности с возвратом или что-то другое. Кроме того, руководители смешивали два разных вопроса:

1. Что люди пишут о системе?
2. Почему они на самом деле так относятся к компании?

На первый можно искать ответ в текстах. Второй значительно шире: текст может перекладывать объяснение самого автора, но не открывает прямой доступ к его внутренним мотивам. Человек может написать «вам нельзя доверять», однако по одной этой фразе нельзя достоверно установить, что именно сформировало его отношение и насколько оно устойчиво.

Контент-анализ позволяет выявить закономерности в сообщениях: какие темы повторяются, как они формулируются, в каких контекстах встречаются. Но он не превращает текст в окно в сознание автора. Это важное ограничение, а не недостаток метода.

#### **3. Объект исследования**

Теперь нужно определить, какие именно материалы будут рассматриваться. Не «отношение пассажиров вообще», а конкретный объект, очерченный границами.

Марина спросила:

— Какие источники у нас уже есть?

Оказалось, что доступны отзывы в магазине приложений, обращения в поддержку и публичные комментарии в социальных сетях. Но в пределах проекта команда могла надёжно выгрузить только отзывы и обращения. Период — первые три недели после запуска. Аудитория — люди, которые оставили текст о новой системе, а не все пассажиры города.

Это уточнение сразу уменьшило риск чрезмерных выводов. Если анализировать только тексты пользователей, которые решили написать отзыв или обратиться в поддержку, нельзя

автоматически утверждать, что они представляют всех пассажиров. Молчание остальных не означает ни удовлетворённости, ни безразличия: причины не писать могут быть самыми разными.

Итак, объектом стали **отзывы в магазине приложения и обращения в поддержку, содержащие упоминания новой системы оплаты, за первые три недели после запуска**. У такого определения есть края. Например, в него не входят разговоры, которые люди не записали, и проблемы, о которых они сообщили только устно.

Границы не обедняют исследование. Они показывают, на какую территорию распространяются выводы.

#### **4. Исследовательский вопрос**

Только теперь команда попробовала сформулировать вопрос, который можно проверить по выбранным материалам:

**«Какие проблемы и темы, связанные с новой системой оплаты, чаще всего описывают пользователи в отзывах и обращениях в поддержку в первые три недели после запуска?»**

Эта формулировка уже не обещает раскрыть все причины недовольства и не назначает заранее главный источник проблем. Она спрашивает о том, что можно наблюдать в текстах: темы и описания пользователей.

Но одного вопроса оказалось мало. Если исследование должно помочь выбрать дальнейшие шаги, полезно разбить его на под-вопросы:

- Какие технические затруднения описываются?
- Какие затруднения связаны с пониманием интерфейса или последовательности действий?
- Как часто упоминаются списания, возвраты и баланс?
- Какие обращения связаны с доверием, безопасностью или ощущением несправедливости?
- Какие сообщения содержат конкретное описание ситуации, а какие — только общую оценку?
- Есть ли заметные различия между отзывами и обращениями в поддержку?

Под-вопросы помогают понять, что именно искать, но не должны превращаться в скрытые рекомендации. «Какую кнопку следует передвинуть?» — уже не вопрос о содержании материалов, если в них нет сопоставимых описаний вариантов дизайна. А «что чаще упоминают пользователи?» — вопрос, на который тексты могут дать ответ.

#### **5. Наблюдаемые признаки**

Даже хорошая формулировка остаётся слишком абстрактной, пока исследователь не решит, что будет считать свидетельством. Слово «негатив» не размечает себя само. Нужно определить признаки, которые можно обнаружить в тексте.

Команда составила предварительный список:

##### **Признак**

##### **Что искать в сообщении**

##### **Чего не следует автоматически заключать**

Проблема с действием

Автор пишет, что не смог оплатить, пополнить баланс или найти нужную функцию

Что интерфейс неудобен для всех пользователей

Ошибка или сбой

Описаны неудачная транзакция, задержка, повторное списание

Что причина обязательно в программном обеспечении

Непонимание правил

Есть вопрос о тарифе, списании, возврате или порядке оплаты

Что пользователь «не прочитал инструкцию»

Доверие и безопасность

Упоминаются опасения за данные, деньги или честность компании

Что у автора устойчивое недоверие ко всей компании

Эмоциональная оценка

Встречаются слова раздражения, благодарности, тревоги или одобрения

Что эмоциональный тон равен объективной оценке качества

Конкретность описания

Есть последовательность действий, время, сумма или результат

Что подробно описанный случай типичен для всех

Последний столбец не менее важен первого. Например, «дважды списали деньги» — это описание пользовательского опыта. Оно ещё не доказывает, что списание действительно было двойным: для этого нужны данные о транзакциях. «Не могу найти кнопку» — свидетельство затруднения конкретного автора, но не готовый диагноз интерфейса.

В исследовании полезно отделять **то, что явно сказано**, от **того, что исследователь интерпретирует**. Если человек пишет: «Оплатил, но приложение всё равно показывает неоплаченный проезд», это явное описание расхождения. Если аналитик кодирует его как «недоверие к компании», он делает дополнительный шаг интерпретации. Такой шаг может быть оправдан, но его нужно определить заранее и не выдавать за прямые слова автора.

### **Как вопрос меняет качество вывода**

Представим, что в 200 сообщениях кодировщики обнаружили 60 упоминаний трудностей с оплатой, 45 вопросов о возвратах, 30 жалоб на поиск нужной функции и 25 высказываний о недоверии. Эти числа могут показать распределение тем **в изученной выборке**. Но они не отвечают автоматически на вопрос, какая проблема важнее для всех пассажиров.

Для этого нужны дополнительные основания: например, данные о числе пользователей, столкнувшихся с каждым сбоем, интервью, наблюдение за выполнением задачи или репрезентативный опрос. Частота упоминаний — полезный сигнал, но не равна масштабу проблемы. Одно и то же событие разные люди могут описать с разной вероятностью; а некоторые вообще не оставят сообщения.

Так вопрос удерживает вывод в границах источника. Если вопрос звучит «Какие темы встречаются в обращениях?», можно описывать обращения. Если команда хочет установить, какой элемент интерфейса причинно вызывает отказ от оплаты, одного корпуса текстов может быть недостаточно. Возможно, потребуется эксперимент, аналитика пользовательских действий или тестирование интерфейса.

Это не означает, что контент-анализ бесполезен. Напротив, он помогает понять, какие гипотезы стоит проверять дальше и где находятся повторяющиеся узкие места. Просто метод должен отвечать на свой вопрос, а не на любой вопрос, который хочется задать руководству.

### **Шаблон формулировки**

Для следующего проекта можно использовать простой шаблон:

**«Какие [темы, оценки, описания или способы аргументации] встречаются в [конкретном типе материалов] у [определённой аудитории или авторов] в [указанный период] в связи с [явлением или событием]?»**

Например:

**«Какие причины затруднений при оплате упоминают пользователи в обращениях в поддержку в первый месяц после запуска приложения?»**

Если интересуют различия, добавьте основание для сравнения:

**«Чем отличаются темы отзывов в магазине приложения от тем обращений в поддержку за тот же период?»**

Шаблон — не заклинание. Он не гарантирует хорошего проекта, но заставляет назвать ключевые элементы: что изучаем, в каких материалах, у кого и когда.

### **Короткая проверка перед началом**

Перед тем как собирать корпус и выбирать инструменты, проверьте вопрос по трём критериям.

**Измеримость.** Можно ли описать признаки, по которым два подготовленных исследователя распознают нужное явление? Если вопрос содержит слова «понять», «успешный», «плохой» или «лояльный», уточните, какие именно проявления будут считаться свидетельством.

**Выполнимость.** Уложится ли проект в доступное время и объём данных? Можно ли обработать весь корпус или потребуется обоснованная выборка? Не обещает ли вопрос изучить аудиторию, для которой нет материалов?

**Соответствие источникам.** Действительно ли выбранные материалы позволяют ответить на вопрос? Отзывы могут показать, что люди пишут о своём опыте. Они обычно не позволяют точно измерить долю всех пользователей с такой проблемой или установить техническую причину сбоя.

Если хотя бы на один пункт ответ «нет», менять нужно не только формулировку. Иногда стоит сузить вопрос, найти другой источник или честно разделить проект на несколько исследований.

### **Пятиминутная редактура**

Возьмите одну расплывчатую задачу из своей работы — например, «понять, почему клиенты уходят» или «выяснить отношение аудитории к новому продукту». Затем письменно ответьте на пять вопросов:

1. Какое решение предстоит принять?
2. Какой информации для него не хватает?
3. Какие конкретно материалы доступны?
4. На какой вопрос можно ответить именно по этим материалам?
5. Какие наблюдаемые признаки подтвердят, что найден ответ?

Теперь проверьте, не спряталась ли в формулировке подсказка к желаемому результату. Слова «почему они недовольны», «насколько плохим оказался запуск» и «как доказать эффективность кампании» часто сообщают больше о тревогах заказчика, чем о задаче исследования.

На том совещании команда в итоге не решила, что нужно немедленно менять интерфейс, и не объявила, что причина — недоверие. Она договорилась начать с анализа отзывов и обращений за первые три недели, отдельно отмечая описанные затруднения, вопросы о списаниях и высказывания о доверии. Это был не финальный ответ, а точный план, который мог привести к ответу.

В этом и состоит дисциплина исследовательского вопроса: он не обещает узнать всё и не торопится дать совет. Он связывает практическое решение с информационным пробелом, выбранными материалами и наблюдаемыми признаками. Хорошее исследование начинается не с того, какой инструмент открыть, а с того, на какой вопрос этот инструмент способен ответить.

А следующий шаг — понять, где взять материалы, которые действительно позволят его задать и проверить.

## Выбор источников и сбор корпуса материалов

На совещании открыли четыре окна: сайт транспортного ведомства, страницу городского сообщества в соцсети, ленту местного новостного агрегатора и выгрузку обращений в поддержку. В одном окне — аккуратно выверенные заявления о реформе, в другом — сотни возмущённых комментариев, в третьем — заголовки про «транспортный коллапс», в четвёртом — вопросы о том, где теперь пересаживаться. Руководитель посмотрел на экраны и сказал: «Отлично, у нас есть всё».

Вот тут исследование и начинается.

Потому что «всё» — это не название корпуса. Это обещание, которое трудно выполнить и легко дать. Материалы из разных источников не складываются в единую картину сами собой. Они освещают разные стороны события, создаются разными людьми и попадают к исследователю по разным правилам. Выбор источников заранее определяет, что именно получится заметить, а что останется за пределами поля зрения.

### Сначала — вопрос, потом — платформы

Представим, что команда изучает транспортную реформу, запущенную 1 сентября. У неё мало времени, а материалов столько, что можно потратить месяц только на их сортировку. Команда хочет понять, как обсуждалась реформа в первый месяц. Это уже лучше, чем «изучить отношение горожан»: последнее звучит масштабно, но собранные материалы могут и не позволить говорить об отношении горожан в целом.

Более честный вопрос мог бы звучать так:

Какие темы, аргументы и способы описания транспортной реформы встречаются в доступных публичных материалах трёх выбранных источников с 1 по 30 сентября?

В этой формулировке есть границы. Мы изучаем доступные публичные материалы, а не все разговоры о реформе. Мы можем описывать, что в них встречается, но не заключать, что так думают все жители. И мы должны заранее объяснить, почему выбрали именно эти источники.

На столе у команды — четыре потенциальных источника. В исследование войдут три: официальный сайт транспортного ведомства, публичные публикации и комментарии в социальной сети, материалы нескольких местных новостных изданий. Обращения в поддержку пока останутся за рамками: они могут рассказать о практических затруднениях, но это частные данные с отдельными требованиями к доступу, обезличиванию и хранению. Кроме того, обращение в службу поддержки — не то же самое, что публичное высказывание. Если смешать его с комментариями под постами, различия между каналами легко принять за различия в настроениях.

### Источник

#### Что помогает увидеть

#### Доступность, полнота и стабильность

#### Важное ограничение

Сайт транспортного ведомства

Официальные объяснения, изменения маршрутов, заявленные цели

Публикации обычно доступны, но архив может быть неполным; страницы могут обновляться

Показывает позицию ведомства, а не общественное мнение

Публичная страница в соцсети и комментарии

Темы, возникающие в видимом публичном обсуждении

Просмотр может зависеть от интерфейса, сортировки, ограничений платформы и модерации; посты и комментарии могут исчезнуть

Комментаторы не представляют всех пользователей платформы, а видимая лента не равна полному потоку сообщений

Местные новостные издания

Редакционные рамки, выбранные темы и цитируемые стороны

Поиск и архивы доступны неравномерно; статьи могут редактироваться, а один текст — публиковаться на нескольких сайтах

Новости отражают редакционный отбор, а не весь ход обсуждения

Обращения в поддержку

Вопросы и трудности, с которыми люди сталкиваются при использовании транспорта

Доступ часто требует разрешения; нужна работа с чувствительными данными и идентификаторами

Это отдельный канал обратной связи с собственной аудиторией и правилами

Таблица не назначает источникам оценки вроде «хороший» или «плохой». Она помогает сопоставить их с вопросом. Официальный сайт полезен, если мы хотим изучить, как реформу объясняет ведомство. Комментарии могут показать, какие темы появляются в видимом публичном разговоре. Новости — какие аспекты события становятся редакционными сюжетами. Ни один источник не заменяет другой, но и ни один не даёт права говорить за всех.

Это важно ещё и потому, что платформы устроены по-разному. В соцсети пользователь может видеть не все публикации, а только те, что система показывает ему в выбранной выдаче. Комментарии зависят от того, кто решил написать, кто увидел пост, что прошло модерацию и что сохранилось к моменту сбора. Если в корпусе много недовольных комментариев, это говорит о заметности недовольства в собранных материалах, но само по себе не показывает, сколько жителей недовольны реформой.

Не смешивайте и простые количества. Если на сайте ведомства десять пресс-релизов, а под публикациями — тысяча комментариев, из этого нельзя заключить, что «голоса жителей в сто раз громче». В одном случае единица — публикация, в другом — комментарий; условия производства этих материалов несопоставимы. Для анализа полезнее сохранять сведения о каждом источнике отдельно, а не сводить их в один общий счётчик.

### **Корпус — не весь интернет**

Корпус — это собрание материалов, которое исследователь намерен изучить в заданных границах. Но полезно различать предполагаемый корпус и фактически собранный. Первый задаёт правило: какие материалы должны попасть в исследование, если они доступны и соответствуют критериям. Второй состоит из записей, которые команда действительно смогла найти, получить и проверить.

Выборка — это часть материалов, отобранная из более широкого массива для анализа. Например, команда может собрать все доступные подходящие публикации за месяц, а затем отобрать из них 300 для подробной ручной разметки. Если же исследователи включили только посты с большим числом реакций, это уже не просто техническое сокращение: оно меняет то, что может быть обнаружено. Вопрос о том, как строить такую выборку и не исказить картину, потребует отдельного разговора. Пока задача — чётко определить границы материала, из которого она будет формироваться.

Начинать удобно с правил включения и исключения. В нашем случае подходящий материал:

- опубликован с 1 по 30 сентября включительно;
- размещён на официальном сайте транспортного ведомства, на выбранной публичной странице в соцсети или на сайте заранее определённого местного издания;
- прямо касается реформы — например, упоминает её название, затронутые маршруты или изменения в расписании;
- опубликован на русском языке;
- доступен для исследовательской работы на момент сбора.

Из корпуса исключаются материалы, в которых транспорт упоминается только мимоходом, сообщения вне заданных источников и публикации до или после выбранного периода. Такое правило кажется очевидным, пока не встречается спорный текст: новость о пробках из-за ремонта на улице, комментарий о маршруте, изменённом ещё до реформы, или пост о реформе без ключевых слов из поискового запроса. Критерий нужен именно для таких случаев: чтобы команда не решала каждый раз по настроению.

### **Период, язык и другие границы**

Период исследования — не декоративная строка в отчёте. Если реформу запустили 1 сентября, первый месяц после запуска можно задать как календарные даты с 1 по 30 сентября. Но нужно указать, по какому часовому поясу определяются даты публикации. Для городского исследования логично использовать местное время; при этом платформа может хранить временные отметки в другом формате. Важно не только установить правило, но и записать его.

Нужно решить, что считать датой материала: первоначальную публикацию или последнюю редакцию. У новостной статьи могли исправить заголовок на следующий день; пост ведомства могли обновить, не создавая новой публикации. Практичное решение — сохранять и дату первоначальной публикации, и дату сбора, а при наличии фиксировать сведения об обновлении. Так исследователь не будет выдавать версию страницы, увиденную сегодня, за неизменный текст, опубликованный месяц назад.

Языковая граница тоже влияет на содержание корпуса. Если включаются только русскоязычные материалы, в отчёте следует так и написать. Это не недостаток, если решение соответствует вопросу и ресурсам команды. Недостатком было бы умолчание, создающее впечатление, будто изучены все языки, на которых обсуждали реформу.

В цифровых материалах важно заранее определить, что делать с повторениями. Один и тот же пресс-релиз может появиться на сайте ведомства и в новостной ленте. Новостные сайты могут перепечатывать один и тот же текст. Пользователь может разместить чужую запись без изменений или добавить к ней собственный комментарий.

Не стоит удалять такие материалы без следа. Сначала нужно понять, что именно вы изучаете: уникальные тексты, распространение публикаций или отдельные акты размещения. Для анализа тем команда может пометить точные копии и считать их одним уникальным текстом, сохранив все ссылки и указав, где копии обнаружены. Для анализа распространения, напротив, важно оставить каждое размещение как отдельную запись. Если к перепосту добавлен собственный текст, его можно хранить отдельно от исходного сообщения: в нём может содержаться самостоятельная мысль.

### **Сбор — это маршрут, а не охота за сокровищами**

До начала сбора полезно зафиксировать не только источники и даты, но и способ поиска. Если пресс-релизы собирают через архив сайта, нужно указать путь к архиву и ключевые запросы. Если статьи ищут по списку изданий, нужно заранее составить список. Если команда использует поиск внутри соцсети, следует записать запросы, настройки выдачи и ограничения интерфейса.

Это не гарантирует, что будут найдены все публикации. Поиск может пропустить материал, платформа может показать только часть комментариев, архив сайта может быть неполным. Задача протокола — не выдать ограниченный доступ за полный охват, а дать другому

человеку возможность понять, что именно делали исследователи и где могли возникнуть пропуски.

Практичный маршрут сбора выглядит так.

**1. Определите, что именно будет единицей материала.** Для сайта это может быть пресс-релиз, для новостного источника — статья, для соцсети — пост и отдельно комментариев. Если пост и комментарии попадут в один файл без различия, анализировать их будет труднее.

**2. Составьте список источников и способов поиска.** Зафиксируйте адреса официальных страниц и сайтов, перечень изданий, поисковые слова и возможные варианты названия реформы. Сохраняйте и сами запросы: знание о том, как найден материал, часто не менее важно, чем ссылка на него.

**3. Собирайте сведения вместе с текстом.** Записывайте название источника, дату публикации, дату сбора, ссылку, тип материала, язык, автора или аккаунт, связанный пост — если он есть, и решение о включении. Для исключённых записей сохраняйте причину: например, «не относится к реформе» или «опубликовано вне периода».

**4. Проверяйте записи и повторения.** Убедитесь, что ссылка открывается, дата не перепутана с датой обновления, а копии отмечены. Если материал включён или исключён по пограничному случаю, запишите краткое объяснение.

**5. Зафиксируйте состояние корпуса.** Платформы меняются: публикации редактируют, комментарии удаляют, ссылки перестают работать. Укажите дату, на которую материалы были собраны и проверены. Если разрешено и уместно, храните исследовательскую копию страницы или загрузки вместе со ссылкой. Вопросы авторского права, приватности и условий доступа стоит решать до сбора, а не после того, как папка уже заполнена.

Для публичных материалов также действует важное правило: публичная доступность не означает, что личными данными можно распоряжаться без ограничений. Для анализа часто не нужны имена, фотографии профиля и другие идентификаторы. Собирайте только необходимые сведения, ограничивайте доступ к рабочим файлам и отдельно продумайте, как будете хранить и приводить цитаты из комментариев.

Удалённые записи требуют такой же прозрачности. Не нужно пытаться любой ценой восстанавливать недоступный материал. Если команда видела запись во время сбора и сохранила её в допустимом формате, это можно отразить в протоколе. Если позже ссылка перестала открываться, отметить изменение лучше, чем молча удалить след из данных. Если же запись исчезла до того, как исследователи её нашли, команда не может считать, что включила её в корпус.

#### **Паспорт корпуса: кратко, но достаточно**

Когда сбор завершён, составьте паспорт корпуса. Он не заменяет данные и не доказывает, что собрано всё. Его задача — дать читателю возможность понять границы исследования и оценить, подходят ли они для сделанных выводов.

#### **Образец паспорта корпуса**

##### **Поле**

##### **Описание**

Исследовательский вопрос

Какие темы и аргументы встречаются в доступных публичных материалах о транспортной реформе в течение первого месяца после запуска?

Источники

Сайт транспортного ведомства; одна заранее выбранная публичная страница в соцсети; сайты трёх местных новостных изданий

Период

С 1 по 30 сентября 2024 года включительно, по местному времени города

Критерии включения

Русскоязычные материалы из заданных источников, прямо относящиеся к реформе и опубликованные в указанный период

Критерии исключения

Случайные упоминания транспорта; материалы вне периода и заданных источников; обращения в поддержку

Единицы материала

Пресс-релиз; новостная статья; публикация в соцсети; публичный комментарий, связанный с подходящей публикацией

Способ сбора

Архив сайта ведомства; поиск на страницах и в архивах выбранных изданий по заранее составленному списку терминов; просмотр публичных публикаций и доступных комментариев на странице соцсети

Дубликаты и перепосты

Точные копии помечаются; уникальные тексты различаются с повторными размещениями; авторские добавления к перепостам сохраняются отдельно

Метаданные

Источник, дата публикации, дата сбора, ссылка, тип материала, язык, аккаунт или издание, основание включения

Известные ограничения

Поиск платформы может быть неполным; часть комментариев могла быть удалена или не отображаться; новостные публикации могли обновляться; источники не представляют население города

Дата фиксации

7 октября 2024 года

Это пример, а не готовый универсальный шаблон. Для другого вопроса паспорт будет иным: в нём могут появиться дополнительные языки, источники, временные границы или правила работы с ответами на комментарии. Главное, чтобы ключевые решения не оставались в памяти одного участника команды.

**Упражнение: проверьте свои границы**

Перед тем как открывать поисковую строку, уделите десять минут четырём вопросам:

1. Как звучит исследовательский вопрос, если убрать обещание говорить за всех?
2. Что каждый выбранный источник позволяет увидеть — и чего он не показывает?
3. Какие три признака автоматически включают материал в корпус, а какие ведут к исключению?

**4.** Что вы будете делать, если запись удалят, скопируют на другом сайте или опубликуют на другом языке?

Если ответы вызывают спор, это не потеря времени. Значит, вы обнаружили решение, которое иначе незаметно принял бы интерфейс, поисковый алгоритм или самый настойчивый участник команды.

Корпус не обязан быть огромным, чтобы быть полезным. Он должен быть соразмерен вопросу, собран по понятным правилам и описан без притворства, будто цифровые платформы дают полный доступ ко всему. Источники определяют, какую часть явления увидит исследователь; паспорт корпуса показывает читателю, где проходит граница этого взгляда. А когда граница обозначена, становится возможен следующий шаг: решить, какие материалы из уже определённого корпуса изучать подробно, чтобы выбор не исказил картину.

## Как сформировать выборку и не исказить картину

«Все пассажиры возмущены!» — так гласил заголовок популярной публикации о повышении стоимости проезда. Под ним — тысяча постов: «Опять содрали деньги», «Город совсем обнаглел», «Не дождётесь, чтобы я ещё раз сел в этот автобус». Одни сообщения собрали тысячи лайков, другие разошлись по городским каналам, третьи перепечатывали с возмущёнными комментариями.

Вывод казался очевидным: пассажиры единодушны. Но команда исследователей решила проверить, откуда взялась эта тысяча. Оказалось, что публикация собрала посты с наибольшим откликом на нескольких платформах за короткий период после объявления о повышении. В подборке почти не было сообщений с небольших районных страниц, постов, опубликованных позднее, и коротких нейтральных реплик. Зато один и тот же текст встречался десятки раз — в перепечатках и цитатах.

Тысяча сообщений — солидная цифра. Только тысяча чего именно? Тысяча публикаций, отобранных алгоритмами за громкость, — не тысяча случайно выбранных пассажиров. А гневный пост, каким бы выразительным он ни был, не становится голосом города лишь потому, что его много раз переслали.

### **Громкость — не частота**

В цифровой среде заметность легко принять за распространённость. Платформы показывают пользователю не все сообщения, а отобранные по множеству сигналов: реакции, подписки, интересы, актуальность, предполагаемая вовлечённость. Популярный пост получает больше внимания, а внимание помогает ему стать ещё популярнее. Так возникает усиление: сообщение оказывается на виду не только потому, что отражает распространённое мнение, но и потому, что хорошо распространяется.

Есть и другой механизм — самоотбор авторов. Пост обычно пишет не каждый, кто столкнулся с изменением, а тот, кому есть что сказать и кто готов сказать это публично. Одни пассажиры выражают недовольство, другие молча пересчитывают расходы, третьи не пользуются соцсетями, четвёртые вообще не заметили перемен. Молчание в корпусе публикаций нельзя автоматически прочесть как согласие, равнодушие или отсутствие проблемы.

Наконец, люди, лайки и тексты — разные единицы подсчёта. Один человек может написать несколько постов. Один текст могут скопировать десятки аккаунтов. Один пост может получить много реакций, но это не значит, что с ним согласились все поставившие лайк: кто-то поддержал мысль, кто-то отреагировал на спор, кто-то нажал на кнопку машинально.

Поэтому, увидев цифру «тысяча сообщений», полезно спросить не только «сколько?», но и «тысяча отобрана как?». Если в выборку попали наиболее эмоциональные публикации, по ним можно изучать, как устроена эмоциональная критика. Нельзя без дополнительных данных утверждать, что именно так думают все пассажиры.

### **Сначала — кому и чему мы хотим сделать вывод**

Выборка — это не просто уменьшенная версия корпуса, которую удобнее обработать. Это правило отбора, определяющее, какие материалы получают шанс попасть в анализ, а какие останутся за кадром. А значит, именно правило отбора во многом задаёт границы выводов.

Прежде чем выбирать посты, нужно разделить три понятия.

**Целевая совокупность** — явление или группа материалов, о которых исследователь хочет что-то узнать. Например: публичные публикации о городском транспорте за первый месяц после изменения тарифов.

**Рамка отбора** — то, что исследователь действительно может обнаружить и собрать: посты в конкретных открытых группах, публикации по заданным запросам, записи из определённого архива.

**Выборка** — та часть рамки, которая по установленной процедуре попадёт в анализ.

Эти три множества редко полностью совпадают. Если исследователь собирает открытые посты по нескольким хештегам, он изучает не все разговоры о транспорте и тем более не все мнения пассажиров. Он изучает доступные публичные публикации, обнаруженные по заданным правилам. Это не недостаток само по себе. Это граница, которую следует назвать.

Полезный вопрос: **на какие объекты в итоге распространяется мой вывод?** Если единицами отбора были посты, корректно говорить о распределении тем или тональности среди отобранных постов. Если хочется сделать вывод о жителях города, понадобится исследовательский дизайн, который позволяет судить о жителях, например репрезентативный опрос. Контент-анализ публикаций может показать, что люди пишут публично, но не измеряет автоматически, что думает население.

**Четыре стратегии: не лучше и хуже, а для разных задач**

Иногда можно изучить все доступные материалы. Иногда нужно отобрать часть случайным образом. В другом случае лучше взять каждый десятый документ, а для качественного анализа — целенаправленно выбрать публикации, в которых особенно подробно раскрывается интересующее явление. Универсального способа нет: хороший выбор зависит от вопроса, объёма корпуса, доступности материалов и допустимой точности.

**цель**  
**процедура**  
**плюсы**  
**риски**  
**границы обобщения**

Изучить весь небольшой и чётко ограниченный корпус

Сплошной сбор: включить все доступные материалы за заданный период и из заданных источников, подходящие под критерии

Не нужно решать, какие единицы оставить; можно описать весь ограниченный корпус

«Все» легко перепутать со «всем явлением»: часть материалов может быть недоступна, удалена или не найдена

Выводы относятся к определённому корпусу, источникам и периоду, а не автоматически ко всем жителям или разговорам

Оценить распределение характеристик в большой рамке

Вероятностная выборка: случайно отобрать единицы так, чтобы каждая имела известную вероятность попасть в выборку; при необходимости разделить рамку на слои, например по платформам или неделям

Позволяет оценивать характеристики корпуса с учётом случайной ошибки, если рамка и процедура устроены корректно

Слабая рамка исключит целые группы; неответы и пропуски создают смещение; нужны ресурсы на построение списка и контроль отбора

Распространяется на охваченную рамку, а не на любые публикации и не на всё население

Получить управляемую выборку из упорядоченного списка

Систематический отбор: выбрать случайную стартовую позицию, затем брать каждую  $k$ -ю единицу

Прост в реализации; помогает распределить отбор по списку

Периодичность в порядке списка может совпасть с шагом отбора и исказить состав

Выводы ограничены исходным списком; метод не исправляет пробелы в его покрытии

Подробно изучить разные способы выражения темы или редкие случаи

Целевой отбор: включить материалы по заранее заданным признакам — например, разные аргументы, позиции, ситуации или типы авторских групп

Помогает увидеть разнообразие, механизмы и исключительные случаи; полезен для качественного анализа

Отбор зависит от исследовательского решения; нельзя по нему оценивать долю мнений в совокупности

Позволяет говорить о выбранных случаях и обнаруженных способах выражения темы, но не о распространённости этих способов

**Сплошной сбор** хорош, когда корпус посилен и чётко ограничен: например, все опубликованные пресс-релизы транспортного ведомства за год. Но «мы собрали всё, что нашли» и «мы собрали всё существующее» — разные заявления. Поиск мог не охватить закрытые группы, удалённые материалы или публикации без выбранных слов.

**Вероятностная выборка** уместна, если исследователю важно оценить, как часто характеристика встречается в определённой совокупности материалов. Принцип прост: отбор должен происходить по случайной процедуре, а вероятность включения каждой единицы — быть известной. При этом случайность отбора не спасает от плохой рамки. Если список содержит только крупные городские сообщества, даже образцовый жребий не добавит в него жителей, которые публикуются на районных страницах.

Чтобы не потерять важные части корпуса, можно применить стратификацию: разделить список на осмысленные группы — например, платформы, недели или районы — и отбирать случайно внутри каждой. Это помогает не получить выборку, в которой случайно оказалась почти одна платформа или один период. Но стратификация не превращает публикации в представителей населения: она улучшает охват заданной рамки, а не меняет объект исследования.

**Систематический отбор** удобен, когда материалы уже выстроены в список и нужно выбрать, скажем, каждый двадцатый пост. Стартовую позицию следует определить случайно. Иначе исследователь может неосознанно взять начало списка, первые страницы выдачи или наиболее удобные записи.

Метод требует проверить порядок списка. Представим, что в архиве посты отсортированы по дням, а каждый десятый отбор попадает на понедельник. Если понедельничные публикации заметно отличаются от публикаций других дней, выборка будет иметь регулярный перекося. Систематический отбор особенно опасен, если в структуре списка есть периодический ритм, совпадающий с шагом выборки.

## **Конец ознакомительного фрагмента.**

Текст предоставлен ООО «Литрес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на Литрес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.