

# Контент-анализ в социологических исследованиях

Курс «Современные методы социологических исследований», 2013

# Контент-анализ =

- стандартная методика исследования в области общественных наук, предметом анализа которой является содержание текстовых массивов и продуктов коммуникативной корреспонденции
- *количественный анализ текстов* и текстовых массивов с целью последующей содержательной интерпретации выявленных *числовых закономерностей*.
- формализованный метод изучения текстовой и графической информации, заключающийся в переводе изучаемой информации в *количественные* показатели и ее *статистической* обработке.
- Сущность метода контент-анализа состоит в фиксации определенных *единиц* содержания, которое изучается, а также в *квантификации* полученных данных.
- Объектом контент-анализа может быть содержание различных печатных изданий, радио- и телепередач, кинофильмов, рекламных сообщений, документов, публичных выступлений, материалов анкет.

## Г. Лассуэлл

- 1927 г. «Техника пропаганды в первой мировой войне»: социальные модели поведения, которыми манипулирует пропаганда воюющих стран (газеты, журналы, бюллетени, проповеди...):
  - *«мы» защищаемся, «враг» – коварный агрессор;*
  - *«враг» разрушил райское благоденствие и поэтому должен быть уничтожен;*
  - *«мы» победим, «враг» будет уничтожен*
    - Англия: доводы гуманности
    - Германия: великая германская культура, нуждающаяся в защите
  - «менты» («mention») – упоминания,
  - СЛОВО – «СИМВОЛ», «ЗНАК» - единица анализа

## Г. Лассуэлл (2-ая мировая война):

- насколько совпадают публикуемые материалы с тезисами немецкой пропаганды
- частота появления про- и антиамериканских, британских и немецких тем в информационном источнике
- определение политической линии той или иной газеты, в выявлении того, какой вклад — позитивный или негативный вносили американские издания в поддержание морального духа народа в связи агрессивными действиями фашистской Германии
- Верховный Суд США - решения о закрытии некоторых изданий (газета «Истинный американец»)

## Б. Берельсон: расовая дискриминация в США (1946 г. совместно с Сальтером)

- Объект анализа - короткие рассказы из 8 наиболее популярных американских изданий (1937-1943 гг.)
- Категории анализа: нац. принадлежность героя, роль-положение, вид изображения (с симп., без симп.) ....
- Подача героев разной национальности (в % ко всем упоминаниям каждой национальности)

Характеристики героев	«Истинные» американцы	Англосаксы и североевропейцы	Прочие
Главные герои	52	38	30
Роли, вызыв. симпатию	80	76	62
Высший соц.-эк. статус	39	24	16
Высшая моральная позиция	59	29	20

## Структура национальных групп (в % ко всему населению и к числу действ. лиц)

Национальные группы	Структура факт. населения	Действующие лица
Евреи	3,6	1,2
Афроамериканцы	9,8	1,9
Др. национальности	17,6	2,8
Англо-сакс. и норд. происхождения	8,8	3,3
Американцы	60,2	90,8

А.Шай: ведущие амер.газеты и журналы (1949-1950 гг.) – только иллюстрации – главная фигура иллюстрации (уровень профессиональной подготовки героя)

Статус	афроамериканцы		белые	
	в жизни	на фото	в жизни	на фото
Больше, чем квалифицированные рабочие и служащие	4	5	15	45
Квалифицированные	5	-	21	45
Неквалифицированные	91	95	54	8

Б. Берелсон «Контент-анализ в коммуникационных исследованиях» (нач. 50-х гг) - фундаментальное описание контент-анализа (виды, критерии, единицы к.-а.)

Л.Левенталь – подборка биографий в изданиях «Saturday Evening Post» и «Collirs» (рейтинг внимания к представителям разных кругов):

	1901-1914	1922-1930	1930-1934	1940-1941
Политики	46	28	31	25
Бизнесмены / предприниматели	28	18	14	20
Развлечения (шоу-бизнес)	26	54	55	55

М.Шпигельман, К.Тервиллингер, Ф.Феринг (1950): комиксы (тираж 17 млн. + 9 млн. – около 60 млн.читателей):

Высший слой – стремление к гуманным целям

Средний – к целям эгоистическим

Низший – к личным удовольствиям (наслаждениям)

Женщины – добрые дела, благополучная семейная жизнь, романтическая любовь

Мужчины – стремление к власти (любовь очень низко)

Соотношение между симп./несимп. героями 3:1 (М), 5:1 (Ж)

81% - юмор.комиксы, 56 % - город, 50 % - звери (антропоморфны), М в 2 раза чаще, чем Ж

# Требования к «документам» («текстам») для КОНТЕНТ-анализа

- исследуемое содержание (текст) должно позволять использовать однозначное правило для надежного фиксирования нужных характеристик (*принцип формализации*),
- интересующие исследователя элементы содержания должны встречаться с достаточной частотой (*принцип статистической значимости*).
- Е.g.: сообщения печати, радио, телевидения, массовой устной агитации и пропаганды, протоколы собраний, письма, приказы, распоряжения и т.д., данные свободных интервью и открытые вопросы анкет.



# Основные направления применения контент-анализа:

- выявление того, что существовало до текста и что тем или иным образом получило в нем отражение (текст как индикатор определенных сторон изучаемого объекта — окружающей действительности, автора или адресата);
- определение того, что существует только в тексте как таковом (различные характеристики формы — язык, структура, жанр сообщения, ритм и тон речи);
- выявление того, что будет существовать после текста, т.е. после его восприятия адресатом (оценка различных эффектов воздействия)

# Этапы проведения КОНТЕНТ-анализа

- *!!! Необходим материальный носитель информации*
- **Первый этап. Определение совокупности изучаемых источников** или сообщений с помощью набора заданных критериев, которым должно отвечать каждое сообщение: тип источника, тип сообщений, сопоставимый размер...
- **Второй этап. Формирование выборочной совокупности сообщений.** Выборка: репрезентативная или целевая (в зависимости от цели, доступности, массовости)
- *\*Обычно выборка сообщений средств массовой информации составляет 200-600 текстов.*

# Основные процедуры контент-анализа:

- 3. Выявление смысловых единиц контент-анализа, которыми могут быть:
  - а) понятия, выраженные в отдельных терминах;
  - б) темы, выраженные в целых смысловых абзацах, частях текстов, статьях, радиопередачах и т. п.;
  - в) имена, фамилии людей;
  - г) события, факты и т. п.;
  - д) смысл апелляций к потенциальному адресату.
- Единицы контент-анализа выделяются в зависимости от содержания, целей, задач и гипотез конкретного исследования.
- 4. Выделение единиц счета, которые могут совпадать либо не совпадать с единицами анализа. В 1-м случае процедура сводится к подсчету частоты упоминания выделенной смысловой единицы, во 2-м — исследователь на основе анализируемого материала и здравого смысла сам выдвигает единицы счета, которыми могут быть:
  - а) физическая протяженность текстов;
  - б) площадь текста, заполненная смысловыми единицами;
  - в) число строк (абзацев, знаков, колонок текста);
  - г) длительность трансляции по радио или ТВ;
  - д) метраж пленки при аудио- и видеозаписях,
  - е) количество рисунков с определенным содержанием, сюжетом и пр.
- 5. Процедура подсчета
- 6. **Интерпретация полученных результатов**

# Категории анализа = ключевые понятия (смысловые единицы)

- имеющиеся в тексте и соответствующие тем дефинициям и их эмпирическим индикаторам, которые зафиксированы в программе исследования, должны быть:
  - а) *уместными*, т.е. соответствовали решению исследовательских задач;
  - б) *исчерпывающими*, т.е. достаточно полно отражали смысл основных понятий исследования;
  - в) *взаимоисключающими* (одно и то же содержание не должно входить в различные категории в одинаковом объеме);
  - г) *надежными*, т.е. такими, которые не вызывали бы разногласий между исследователями по поводу того, что следует относить к той или иной категории в процессе анализа документа.

# Исследовательские инструменты

- Программа
- классификатор контент-анализа,
- протокол итогов анализа = бланк контент-анализа,
- регистрационная карточка или кодировальная матрица,
- инструкция исследователю, непосредственно занимающемуся регистрацией и кодировкой единиц счета
- каталог (список) проанализированных документов.

## *Классификатор контент-анализа =*

- общая таблица, в которую сведены все категории (и подкатегории) анализа и единицы анализа. Ее задача — предельно четко зафиксировать то, в каких единицах выражается каждая категория, используемая в исследовании.
- *Регистрационная карточка* представляет собой кодировальную матрицу, в которой отмечается количество единиц счета, характеризующее единицы анализа
- *Протокол (бланк) контент-анализа содержит:*
  - 1) сведения о документе (его авторе, времени издания, объеме и т.п.);
  - 2) итоги его анализа (количество случаев употребления в нем определенных единиц анализа и категорий анализа).

## Классификатор контент-анализа (на примере брачных объявлений) – фрагмент (пример)

Единицы анализа	Индикаторы (единицы счета)
Качества «жены»	«люблю готовить», «вкусно готовлю», «заботливая», «хороший друг», «умею создавать уют», «домашний труд не в тягость»...
Качества характера	«мягкая», «спокойная», «веселая», «жизнерадостна», «ровная в общении», «доброжелательная», «скромная», «добра»...
Увлечения в свободное время	«люблю ходить в походы», «люблю путешествовать», «обожаю театр», «ценю поэзию», «увлекаюсь игрой на гитаре»...
Хар-ки уровня мат. и жил. обеспечен.	«материально независимая», «в жилплощади не нуждаюсь», «рассчитываю на улучшение жилищных условий», «материально обеспечена» ...
Внешние данные	«выгляжу соответственно возрасту», «полная», «внешность приятная», «небольшого роста», «склонна к полноте», «обаятельная»...

# Пример регистрационной карточки

Категории	Подкатегории	Частота упоминания абсолют., раз	Частота упоминания относит., %
1 Категория (национальность)	01 подкатегория (кавказцы)	15	25
	02 подкатегория (русские)	7	14
	03 подкатегория (чеченцы)	25	61
Итого:		47	100

Наиболее употребительны такие единицы счета, как:  
объем внимания: время-пространство / (число строк, площадь в квадратных сантиметрах, минуты, время вещания и т.п),  
появление признаков в тексте;  
частота их появления (интенсивность)



# Пример регистрационной карточки

Признак	Текст					
	1	2	3		n	$\Sigma n$
A			+			
B	+	+				
C	+		+			
...						
n						
$\Sigma n$						

Простые способы измерения: удельный вес той или иной категории можно вычислить с помощью формулы:

$$K = \frac{\text{число единиц анализа, фиксирующих данную категорию}}{\text{общее число единиц анализа}}$$

# Матрица случайностей Ч.Осгуда

Реальное совпадение	Ожидаемое совпадение					
	<i>A</i>	<i>B</i>	<i>C</i>		<i>n</i>	$\Sigma n$
<i>A</i>	-	0,15	0,02			
<i>B</i>	0,05	-	0,06			
<i>C</i>	0,08	0,12	-			
...				-		
<i>n</i>					-	
$\Sigma n$						-

Выявляются меры случайности совпадения каждой классификационной единицы со всеми остальными.

Напр., единица *A* встречается в 30% анализируемых текстов ( $P = 0,3$ ), а единица *B* — в 50% текстов ( $P = 0,5$ ), тогда ожидаемая частота совместного появления этих единиц будет равна:  $P_{AB} = P_A \cdot P_B = 0,3 \cdot 0,5 = 0,15$ .

В действительности же признаки *A* и *B* совместно встретились лишь в 5% текстов  $AB = 0,05$ .

## Специальные процедуры подсчета применительно к контент-анализу: формула коэффициента Яниса ( $c$ )

- для вычисления соотношения положительных и отрицательных (относительно избранной позиции) оценок, суждений, аргументов:

$$c = \frac{f^2 - f \cdot n}{r \cdot t},$$

- где:  $f$  — число положительных оценок;  $n$  — число отрицательных оценок;  $r$  — объем содержания текста, имеющего прямое отношение к изучаемой проблеме;  $t$  — общий объем анализируемого текста.
- в случае, когда число положительных оценок меньше, чем отрицательных, коэффициент Яниса находится по формуле:

$$c = \frac{f \cdot n - f \cdot n^2}{r \cdot t}.$$

# Количественный контент-анализ (содержательный)

- основывается на исследовании слов, тем и сообщений, сосредоточивая внимание исследователя на содержании сообщения.
- Качественный контент-анализ (структурный): исследователя интересует не столько **что** говорится, сколько **как** говорится (способе презентации сообщения)

Для описания и сравнения особенностей речевого поведения рос. политиков - несколько параметров:

- Индекс многословия (лаконичности) - это средний объем ответа политика на один вопрос интервьюера
- Индекс речевого лидерства для текстов интервью получается простым суммированием величин двух второстепенных параметров:
  - индекса тематической новизны (постоянное введение собственных новых тем, а не простое следование темам беседы, которые предлагаются другими участниками) и индекса оценки вопросов (постоянное стремление оценивать высказывания партнеров и словесно выражать свои оценки)
- способность к коммуникативному сотрудничеству ⇒

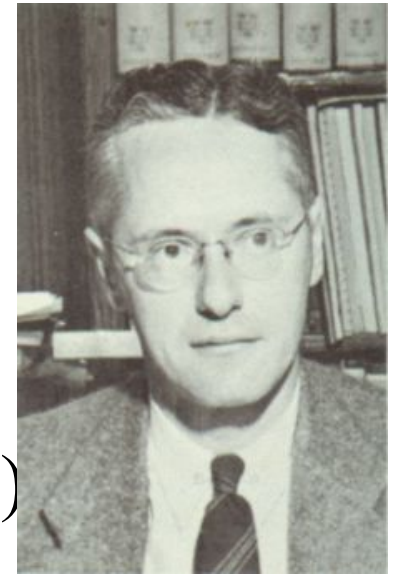
Способность к коммуникативному сотрудничеству (т.е. налаживанию и поддержанию продуктивного взаимодействия в процессе беседы: складывается из четырех составляющих :

- + переформулирование в ответе вопроса партнера (журналиста) с уточнением;
- + повторы в ответе слов и выражений, содержащихся в формулировке вопроса (вопрос: *"Почему? Разве вы сделали ему (Ельцину) что-то плохое?"* Ответ: *"Нет, ничего плохого я ему не делал и никогда не сделаю."*);
- + словесное выражение в ответе согласия с формулировкой или содержанием вопроса (вопрос: *«Насколько затронул экономику Москвы нынешний финансовый кризис?»* Ответ: *"Конечно, он ее затронул"*; или вопрос: *"И вы считаете, что нам удастся здесь найти общий язык с США?"* Ответ: *"Да, я считаю что мы сможем выйти на взаимоприемлемые формулировки"*).
- - антонимическое переформулирование вопроса (например, вопрос: *"Тем не менее на прошлых выборах, наверное, не столько Ельцин выиграл у Зюганова, сколько проиграл Зюганов как человек, олицетворяющий скомпрометированную коммунистическую идею?"* Ответ: *"Ну, во-первых, никто не победил. Все проиграли..."*. Или вопрос: *"Что же будет весной?"* Ответ: *"Ничего не будет, просто жрать будет нечего и топить нечем"* и т.п.)
- ⇒ наиболее дипломатичный и тактичный политик (Е.Примаков)

## 6 типов контент-анализа (Р.Мертон)

в зависимости от целей и объекта:

- Объем внимания (подсчет символов)
- Анализ отношения:
  - положительное, отрицательное, нейтральное (критерии? + общий контекст)
  - модальность сообщения: разрешение («разрешено», «допустимо»), запрет («нельзя допустить»), долженствование («обязан», «должен», «святой долг»), желательность («следовало бы», «хорошо было бы», «имело бы смысл»), побуждение к действию («надо освоить», «давайте подумаем вместе»)
  - характер сообщения (прямой/косвенный): «по мнению...»



## 6 типов контент-анализа (продолж.)

- Анализ по элементам (фотографии к статьям, песни в радиопрограмме, сцены в фильме...)
- Тематический анализ материалов (определение всего многообразия, явных/скрытых, выделение крупных классов)
- Структурный анализ (направлен на анализ взаимоотношения различных тем) – редко (пропаганд. компании)
- Ориентированный на изучение взаимоотношений различного рода материалов с целью достижения единой цели – редко (пропаганда)



# Контент-анализ советской прессы



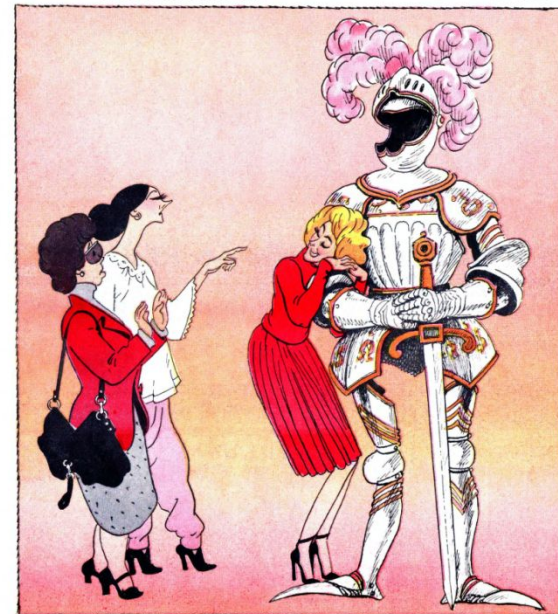
— Думаешь, легко мне эта дубленка досталась? Пришлось кое-кому горло перегрызть!

Рисунок Е. ШАБЕЛЬНИКА.



— Придут из ОБХСС, скажи, что все это золотая рыбка подарила...

Рисунок Г. ИОРША.



— Ну и что с того, что внутри пустой, зато какой блеск!

Рисунок А. КРЫЛОВА.

Думаешь, легко мне эта дубленка досталась? Пришлось кое-кому горло перегрызть

Придут из ОБХСС, скажи, что все это золотая рыбка подарила...

Ну и что с того, что внутри пустой, зато какой блеск!